RNA-Polymerase und Nus-Faktoren – Interaktionen als Schlüssel zur Regulation der bakteriellen Transkription

Dissertation

Zur Erlangung des Doktorgrades
der Fakultät Biologie, Chemie und Geowissenschaften
an der Universität Bayreuth

Vorgelegt von
M. Sc. Biochemie
Johanna Drögemüller

Bayreuth 2015

Die vorliegende Arbeit wurde in der Zeit von Oktober 2011 bis April 2015 in Bayreuth am Lehrstuhl Biopolymere unter Betreuung von Herrn Professor Dr. Paul Rösch 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: 21.04.2015

Zulassung durch die Promotionskommission: 29.04.2015

Wissenschaftliches Kolloquium: 22.07.2015

Amtierender Dekan: Prof. Dr. Rhett Kempe

Prüfungsausschuss:

Prof. Dr. Paul Rösch	(Erstgutachter)
Prof. Dr. Clemens Steegborn	(Zweitgutachter)
Prof. Dr. Klaus Ersfeld	(Vorsitz)
Prof. Dr. Rainer Schobert	

Inhaltsverzeichnis

Zusammenfassung

Die Transkription ist der erste Schritt der Genexpression und damit ein zentraler Prozess in Zellen. Katalysiert wird sie durch das Enzym RNA-Polymerase (RNAP), das in Bakterien aus den Untereinheiten α_2, β, β‘ und ω besteht. Die Transkription ist durch eine Vielzahl von Transkriptionsfaktoren hoch reguliert, unter denen die N-*utilisation substance* (Nus) Faktoren eine wichtige Rolle spielen. Diese wurden bisher hauptsächlich im Modellorganismus *Escherichia coli* charakterisiert und die Unterschiede zu anderen Organismen sind kaum aufgeklärt. Auch über die strukturellen Grundlagen ihrer Interaktion mit der RNAP ist wenig bekannt, was jedoch entscheidend für das Verständnis der Regulation der Transkription auf atomarer Ebene ist.

In dieser Arbeit wurde zunächst der Transkriptionsfaktor NusG des hyperthermophilen Bakteriums *Thermotoga maritima* (tmNusG) untersucht. Auf Grundlage struktureller Analysen, die eine stabile, für NusG-Proteine einzigartige, Interaktion der flexibel verbundenen N- und C-terminalen Domänen (NTD und CTD) zeigten, wurde die Dynamik der Interaktion charakterisiert. Da die Wechselwirkungen mit seinen Interaktionspartnern verhindert werden, ist tmNusG autoinhibiert. Durch gezielten Aminosäureaustausch war es möglich, die Autoinhibition aufzuheben. Im Gegensatz zu NusG aus *E. coli* und vielen anderen Bakterien verfügt tmNusG über eine zusätzliche, in die NTD insertierte Domäne, DII. Mittels Fluoreszenz-Anisotropie und NMR-Spektroskopie konnte für DII sowohl eine sequenzunabhängige Nukleinsäurebindung mit Präferenz für doppelsträngige DNA als auch eine Bindung an die RNAP nachgewiesen werden. Daher ist DII möglicherweise in die Rekrutierung des autoinhibierten tmNusGs zur RNAP involviert und könnte die Bindung von tmNusG an den Elongationskomplex durch RNAP- und DNA-Bindung stabilisieren.

Obwohl die RNAP aus *E. coli* strukturell und mechanistisch bereits intensiv analysiert wurde, existieren kaum Informationen über Intra- und Interdomänendynamiken oder transiente Interaktionen, die vorzugsweise mit NMR-Spektroskopie zugänglich sind. Zunächst wurde ein neues, effizientes Protokoll für die Assemblierung der aktiven RNAP aus den einzeln exprimierten Untereinheiten und deren Reinigung entwickelt. Dies erlaubt die Markierung einzelner Untereinheiten mit spezifischen NMR-Sonden. Mit dem deuterierten Enzym mit $^{1}H,^{13}C$-markierten Methylgruppen von Ile, Leu und Val-Resten und dem deuterierten Enzym mit methylgruppenmarkierter β‘-Untereinheit konnten anschließend [$^{1}H,^{13}C$]-Korrelationsspektren gemessen und die Bindung von NusG-NTD beobachtet werden. Die isolierten Untereinheiten der RNAP konnten ebenfalls löslich und funktionell gereinigt werden. Es wurde eine Methode entwickelt, um die mit einem Transkriptionsfaktor interagierende RNAP-Untereinheit zu bestimmen. Nach Validierung mit NusG-NTD und zwei NusA-Domänen wurde mit diesem Ansatz die β-Untereinheit als Bindungspartner für NusE ermittelt. Zusätzlich wurde eine Methode entwickelt, um die Bindestelle der RNAP auf einem Transkriptionsfaktor mit NMR-Spektroskopie zu identifizieren. Damit war es möglich, die bereits bekannten Bindestellen der RNAP auf NusG-

NTD zu bestätigen und diejenigen auf NusE und NusA-NTD zu ermitteln. Die RNAP-Bindestelle auf NusE überlappt mit der Interaktionsfläche für NusG-CTD, wodurch eine kompetitive Bindung entsteht, die möglicherweise in der Antitermination der Transkription eine Rolle spielt. Mit der RNAP-Bindestelle von NusA-NTD konnte ein detailliertes Modell zur RNAP-Bindung erstellt werden. Dieser Ansatz kann allgemein auf Systeme übertragen werden, in denen ein supramolekularer Komplex mit einem kleineren Bindungspartner (< 30 kDa) interagiert.

Summary

Transcription is the first step in gene expression and thus a central process in cells. It is catalyzed by the enzyme RNA polymerase (RNAP) that consists of the subunits α_2, β, β‘ and ω in bacteria. RNAP is highly regulated by a multitude of transcription factors among which the N-*utilisation substance* (Nus) factors play an important role. These Nus factors have been primarily analyzed in the model organism *Escherichia coli* and only little is known about differences to other organisms. Furthermore, the structural basis of their interaction with the RNAP is only poorly understood, which, however, is crucial for the complete understanding of transcription regulation in atomic detail.

First, the transcription factor NusG from the hyperthermophilic bacterium *Thermotoga maritima* (tmNusG) was explored in this work. On the basis of structural studies that revealed a strong domain interaction between its flexibly connected N- and C-terminal domain (NTD and CTD), a feature unique among NusG proteins, the dynamic of this interaction was analyzed. As the interaction with its binding partners is prevented, tmNusG is autoinhibited. By directed amino acid exchange, it was possible to suppress the autoinhibition. In contrast to NusG from *E. coli* and many other bacteria, tmNusG contains the additional domain DII, which is integrated into the NTD. Fluorescence anisotropy and NMR spectroscopy were used to demonstrate that DII binds nucleic acids with a preference for double stranded DNA and furthermore revealed its interactions with RNAP. Hence it was hypothesized that DII could be involved in the recruitment of the autoinhibited tmNusG to the RNAP. It could thus support the stabilization of the tmNusG:elongation complex interaction by binding to the RNAP and the DNA.

Although the structure and mechanism of *E. coli* RNAP has been analyzed extensively, hardly anything is known about inter- and intradomain dynamics and transient interactions - information preferably accessible by NMR spectroscopy. Thus a new, efficient protocol for the assembly of core RNAP from its separately expressed subunits and the subsequent purification of the active enzyme was developed. The process allows the labeling of a specific subunit within the complete RNAP with specific NMR probes. [^{1}H,^{13}C] correlation spectra of deuterated RNAP in which the methyl groups of Ile, Leu and Val residues were ^{1}H,^{13}C labeled and RNAP with methyl group labeled β‘ subunit could be measured. Furthermore, the binding of NusG-NTD to the isolated, methyl group labeled β‘ subunit was observed. Additionally, the isolated RNAP subunits were purified solubly and functionally and a method was developed to identify the RNAP subunit to which a certain transcription factor binds. Having validated this approach with NusG-NTD and two NusA domains, the β subunit was determined to interact with NusE. We further established a method to identify the RNAP binding surface on a transcription factor by NMR spectroscopy. In this way, we confirmed the known RNAP binding sites on NusG-NTD and determined those on NusA-NTD and NusE. The RNAP binding site on NusE overlaps with its interaction surface for NusG-CTD and the resulting competitive binding might play a

role in transcription antitermination. The identification of the RNAP binding surface of NusA-NTD allowed the proposal of a detailed model of how NusA-NTD interacts with RNAP. This approach can be transferred to any system in which a supramolecular complex contacts a smaller binding partner (< 30 kDa).

1 Einleitung

1.1 Transkription

Die Transkription, das Umschreiben der Desoxyribonukleinsäure (DNA) in Ribonukleinsäure (RNA), ist der erste Schritt der Genexpression in allen drei Domänen des Lebens, den Bakterien, Archaeen und Eukaryonten. Die dabei synthetisierte *messenger* RNA (mRNA) dient als Grundlage der Ribosomen zur Proteinherstellung bei der anschließenden Translation. Die RNA kann auch das direkte Genprodukt darstellen, wie z.B. bei ribosomaler RNA (rRNA) und Transfer-RNA (tRNA). Katalysiert wird die Transkription durch das Enzym RNA-Polymerase (RNAP).

1.1.1 RNA-Polymerase

Während in Eukaryonten unterschiedliche RNAPs für die Herstellung von rRNA (RNAPI), mRNA (RNAPII) und tRNA (RNAPIII) zuständig sind (Cramer *et al.*, 2008), gibt es in Archaeen und Bakterien jeweils nur eine RNAP, die jede Art von RNA herstellt. Diese unterschiedlichen RNAPs sind sich sehr ähnlich, haben ein vergleichbares katalytisches Zentrum, arbeiten mit dem gleichen enzymatischen Mechanismus und weisen gleiche Wechselwirkungen mit Nukleinsäuren auf (Werner und Grohmann, 2011). Auch der Grundaufbau aller dieser zellulären RNAPs mit mehreren Untereinheiten ist konserviert, wobei die RNAPs von Archaeen und Eukaryonten deutlich komplexer aufgebaut sind als bakterielle RNAPs (Werner, 2007). Daher wird davon ausgegangen, dass der letzte universelle gemeinsame Vorfahre (*last universal common ancestor,* LUCA) der Bakterien, Archaeen und Eukaryonten eine RNAP mit einfachem Aufbau ähnlich der von Bakterien besaß (Werner und Grohmann, 2011).

Die Untereinheiten der RNAPs lassen sich in drei funktionale Gruppen unterteilen: Katalyse, Assemblierung und zusätzliche Funktionen (Werner, 2007). Dabei weisen die zentralen Untereinheiten der RNAPs aus Archaeen und Eukaryonten eine hohe Homologie zu denen der RNAP aus Bakterien auf, die aus den Untereinheiten α_2, β, β‘ und ω besteht (Burgess, 1969). Das katalytische Zentrum bilden die beiden größten Untereinheiten, β und β‘ (Abb. 1-1). Diese bilden eine krebsscherenartige Struktur, die den zu transkribierenden DNA-Doppelstrang binden, umschließen und aufschmelzen kann (Vassylyev *et al.*, 2007a; Kontur *et al.*, 2010). β und β‘ katalysieren mit Hilfe von zwei Magnesiumionen die RNA-Synthese aus Nukleosidtriphosphaten (NTPs), die durch den sekundären Kanal zum aktiven Zentrum gelangen (Zhang *et al.*, 1999; Korzheva *et al.*, 2000). Nach jedem NTP-Einbau wird die DNA durch β und β‘ weitergeschoben. Die synthetisierte RNA wird durch einen separaten Austrittskanal entlassen und der DNA-Doppelstrang nach erfolgter Transkription wieder geschlossen (Yarnell und Roberts, 1999; Park und Roberts, 2006).

Für die Assemblierung der bakteriellen RNAP ist vor Allem die α-Untereinheit (Abb. 1-1) von Bedeutung. Die Dimerisierung von α über die N-terminalen Domänen (NTDs) bildet den ersten Schritt für den Zusammenbau des Enzyms (Ito *et al.*, 1975). Anschließend folgt die Assoziation von β, dann

von dem Komplex aus β‘ und ω (Ito *et al.*, 1975; Ghosh *et al.*, 2001). Eine weitere wichtige Rolle spielt α in der Regulation der Transkription, da vor allem die flexiblen C-terminalen Domänen (CTDs) das Ziel vieler regulatorischer Proteine sind (Jeon *et al.*, 1997; Schweimer *et al.*, 2011).

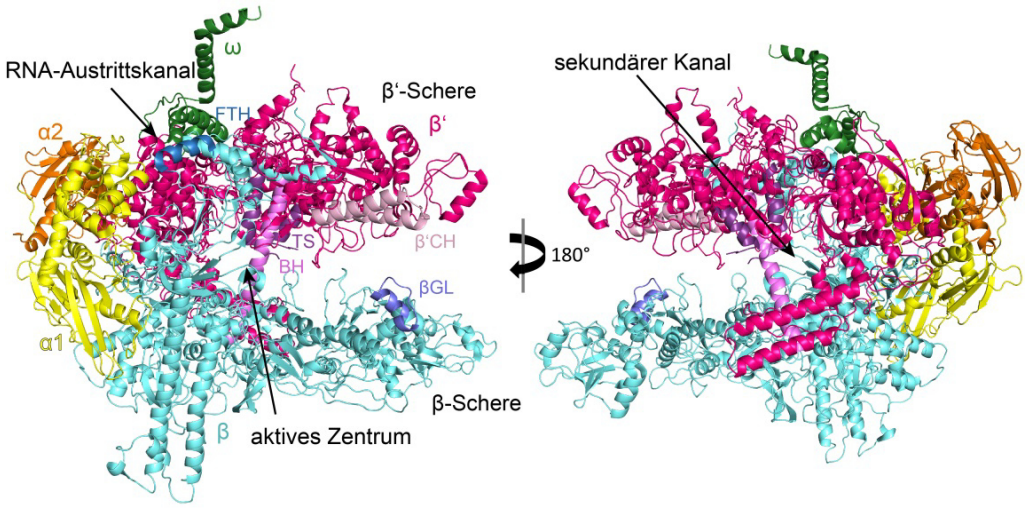


Abbildung 1-1: *Cartoon*-Darstellung der Kristallstruktur der RNAP aus *Escherischia coli* (*Protein Data Bank* (PDB)-Code: 4KMU). Die RNAP hat einen krebsscherenartigen Aufbau, ein Teil der Schere wird von der β-Untereinheit (cyan) gebildet, der andere von der β‘-Untereinheit (pink). Diese umschließen die DNA. Die α-Untereinheiten sind gelb und orange, die ω-Untereinheit grün dargestellt. Wichtige Elemente (β‘-Klappenhelix (β‘-CH), hellrosa; β-Schrankenschleife (βGL), hellblau; flexible Helix der β-Klappendomäne (FTH), blau; Brückenhelix (BH), rosa; Triggerschleife (TS): unterer Teil der beiden lila Helices, Elektronendichte der TS fehlt;) sind farbig hervorgehoben. Die TS faltet sich bei der Nukleotidaddition zu den Triggerhelices (TH).Das aktive Zentrum, RNA-Austrittskanal und sekundärer Kanal sind gekennzeichnet.

Als Untereinheit mit zusätzlicher Funktion existiert bei der bakteriellen RNAP die ω-Untereinheit (Abb. 1-1). ω ist weder essentiell für die Lebensfähigkeit der Bakterien (Gentry *et al.*, 1991) noch für die Funktion der RNAP (Zalenskaya *et al.*, 1990; Tang *et al.*, 1995) und hat eine eher strukturelle als funktionale Aufgabe. ω bindet die N- und C-Termini von β‘, unterstützt die korrekte Faltung von dieser Untereinheit und verhindert ihre Aggregation. Außerdem erleichtert ω die Assoziation von β‘ mit dem $\alpha_2\beta$-Komplex beim Zusammenbau der RNAP (Ghosh *et al.*, 2001). Weiterhin ist die Bindung des σ-Faktors an die RNAP zur Initiation in Abwesenheit der ω-Untereinheit erschwert (Mukherjee *et al.*, 1999). In Cyanobakterien beeinflusst die Deletion von ω die Rekrutierung des primären σ-Faktors ebenfalls negativ. Dadurch ändern sich die Expressionslevel vieler Gene, vermutlich durch Rekrutierung alternativer σ-Faktoren, was bedeutet, dass ω Einfluss auf die Genexpression hat (Gunnelius *et al.*, 2014). Auch die zu ω homologen Proteine RpoK aus Archaeen und RPB6 aus Eukaryonten können mit der Regulation der Transkription in Zusammenhang gebracht werden, da sie mit Transkriptionsfaktoren interagieren (Magill *et al.*, 2001; Ishiguro *et al.*, 2000). Bisher konnten keine derartigen Wechselwirkungen für ω in der bakteriellen Transkription gefunden werden. Es ist aber möglich, dass ω auch in die Regulation der Transkription involviert ist (Mathew und Chatterji, 2006).

Der Kern der bakteriellen RNAP (*core*-RNAP, bestehend aus α_2, β, β‘ und ω) kann ohne weitere Faktoren die RNA-Synthese katalysieren (Mukherjee *et al.*, 1999; Burgess *et al.*, 1969). In *Escherichia coli* (*E. coli*; Abb. 1-1) besitzt das Enzym eine Molekülmasse von etwa 400 kDa (Gasteiger *et al.*,

2003). Der Transkriptionszyklus, den dieses Protein immer wieder unterläuft, besteht aus drei Phase: Initiation, Elongation und Termination (Abb. 1-2, Mooney *et al.*, 1998).

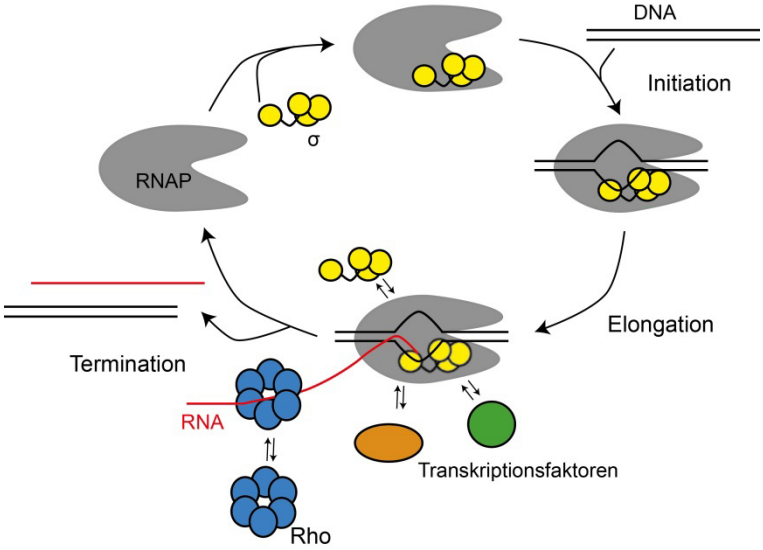


Abbildung 1-2: Schematische Darstellung der bakteriellen Transkription. Der σ-Faktor (gelb) bindet an die *core*-RNAP (grau), wodurch das holo-Enzym gebildet wird. Dies kann nun spezifisch an die Promotorregion der DNA binden, diese entwinden und die Initiation einleiten. Nach Synthese eines etwa 12 nt langen Transkripts geht der Initiationskomplex in die Elongationsphase über. Diese wird durch eine Vielzahl von Transkriptionsfaktoren (grün und orange) reguliert. Am Ende des offenen Leserasters kommt es zur Termination, entweder eingeleitet durch den Terminationsfaktor Rho (blau) oder durch intrinsische Termination. Dabei werden DNA und RNA freigesetzt und die RNAP kann einen neuen Transkriptionszyklus starten (nach Mooney *et al.*, 2009a).

1.1.2 Initiation

Um die Transkription am Promotor zu starten, ist die Bindung eines σ-Faktors an das core-Enzym der RNAP notwendig, wodurch das holo-Enzym gebildet wird (Burgess *et al.*, 1969). Der σ-Faktor erkennt spezifisch die Promotorregion auf der DNA, die von der RNAP aufgeschmolzen werden kann (Abb. 1-2, Sakata-Sogawa und Shimamoto, 2004; Murakami *et al.*, 2002). In Bakterien werden die σ-Faktoren in zwei unterschiedliche Klassen eingeteilt (Paget und Helmann, 2003). Die Faktoren der Klasse 1 (σ^{70}-Familie) enthalten den primären σ-Faktor (in *E. coli* σ^{70}) und sind für die Transkription der allgemeinen organisatorischen (*housekeeping*) Gene verantwortlich (Borukhov und Severinov, 2002). Die Transkriptionsinitiation mit diesen σ-Faktoren verläuft spontan und Energie-unabhängig (Borukhov und Severinov, 2002; Davis *et al.*, 2007). Durch σ-Faktoren der Klasse 2 (σ^{54}-Familie) dagegen wird die RNAP in einer Konformation gehalten, die die DNA-Doppelhelix nicht aufschmelzen kann (Buck *et al.*, 2000; Wigneshweraraj *et al.*, 2008). Hierfür wird, wie auch in Eukaryonten, ein Adenosintriphosphat (ATP)-abhängiges Effektorprotein benötigt, das die strukturellen Einschränkungen des σ^{54}-Faktors aufhebt (Wigneshweraraj *et al.*, 2008; Kim *et al.*, 2000). Ein Austausch des reversibel gebundenen σ-Faktors gegen einen alternativen σ-Faktor ist ein wichtiger Teil der Regulation der Transkription, da unterschiedliche Promotoren erkannt werden und dadurch verschiedene Gene exprimiert werden können, die z.B. unter Stressbedingungen benötigt werden (Gruber und Gross, 2003).

σ^{70}-Faktoren bestehen aus vier konservierten Regionen (Regionen 1-4, Lonetto *et al.*, 1992). Die Bindung des σ-Faktors an die RNAP wird hauptsächlich durch Region 2 vermittelt, die an die Klammern der β'-Untereinheit (β' *clamp helices*, β'CH; Abb. 1-1) bindet. Region 4 bindet an die flexible Helix der Klappendomäne von β (β *flap tip helix*, FTH; Abb. 1-1) am RNA-Austrittskanal Weitere schwache Wechselwirkungen mit der RNAP werden durch die anderen Regionen von σ vermittelt (Vassylyev *et al.*, 2002). Die Promotorbindung erfolgt ebenfalls hauptsächlich durch die Regionen 2 und 4, die zwei Promotorelemente mit konservierter hexamerer Sequenz erkennen. Region 2 bindet von der Transkriptionsstartstelle ausgehend an das -10-Element und Region 4 an das -35-Element des Promotors (Murakami *et al.*, 2002). Durch diese Bindung werden strukturelle Änderungen in der RNAP hervorgerufen, die zu einer stärkeren Bindung der doppelsträngigen (ds) DNA und deren Aufschmelzen führen (Murakami *et al.*, 2002; Zhang *et al.*, 2012). Der codierende DNA-Strang wird im aktiven Zentrum platziert und die RNAP beginnt mit der RNA-Synthese. Dabei werden zunächst nur kurze RNA-Abbruchfragmente synthetisiert, die etwa 2-15 Nukleotide (nt) lang sind und freigesetzt werden. Dieser Vorgang wird als unvollständige Initiation bezeichnet (Goldman *et al.*, 2009), seine Funktion ist jedoch nicht vollständig geklärt. Möglicherweise können diese kurzen RNA-Stränge als Primer an anderen Promotoren fungieren oder dienen als *antisense*-Effektoren für spezifische RNAs (Goldman *et al.*, 2009). Dabei können die kurzen RNA-Stränge intrinsische Termination (s. Kapitel 1.1.4) verhindern, indem sie während der Transkription an die mRNA binden und die Bildung einer Haarnadelschleife verhindern, wodurch die Transkription über den Terminator hinaus fortgesetzt wird (Lee *et al.*, 2010). Weiterhin wurde für eukaryontische RNAPII vorgeschlagen, dass unvollständige Initiation zur Regulation am Promotor dient, sodass der Übergang von Initiation zur Elongation nur erfolgen kann, wenn für die Elongation nötige Transkriptionsfaktoren vorhanden sind, die die unvollständige Initiation unterdrücken (*promoter proofreading*, Liu *et al.*, 2011).

Sobald das Transkript etwa 12 nt umfasst, ist es lang genug, um den RNA-Austrittskanal auszufüllen und kann die Bindung der σ-Region 4 an der FTH destabilisieren (Nickels *et al.*, 2005). Dies ist der erste Schritt in der Dissoziation von σ, wodurch das Verlassen der Promotorregion der DNA eingeleitet wird (Daube und von Hippel, 1999). Damit geht die Transkription von der Initiationsphase in die Elongationsphase über. Der σ-Faktor kann jedoch auch manchmal während der Elongation gebunden bleiben und erst später durch kontinuierliche Abnahme der Affinität zum Transkriptions-Elongations-Komplex (TEC) zufällig dissoziieren (Mukhopadhyay *et al.*, 2001). In diesem Fall spielt σ möglicherweise eine Rolle bei der Regulation der Elongation (Ring *et al.*, 1996; Ko *et al.*, 1998).

1.1.3 Elongation

Während der Elongationsphase (Abb. 1-2) wird stabil ein neuer RNA-Strang synthetisiert. Der Zyklus des NTP-Einbaus besteht dabei aus vier Schritten: der Bindung des NTP-Substrats, der Katalyse der

NTP-Addition, der Freisetzung des Pyrophosphats (PP_i) und der Translokation der RNAP auf dem DNA-Templat (Tomar und Artsimovitch, 2013). Dafür sind vor allem zwei Elemente der β‘-Untereinheit wichtig: die Brücken-Helix (BH) und die Trigger-Helices (TH, Abb. 1-3, Vassylyev *et al.*, 2002; Wang *et al.*, 2006).

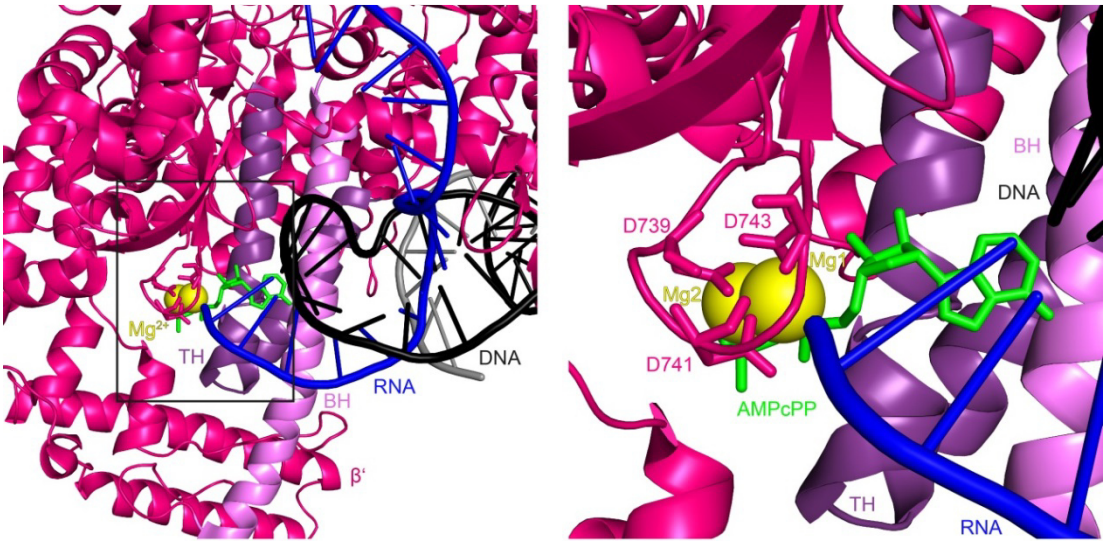


Abbildung 1-3: Aktives Zentrum der RNAP. *Cartoon*-Darstellung der β‘-Untereinheit (pink) der RNAP aus *Thermus thermophilus* (PDB-Code: 2O5J) mit DNA (codierender Strang: schwarz, Gegenstrang: grau), naszierender RNA (blau) und dem nicht spaltbaren Nukleotidanalogon AMPcPP (grün). Die katalytischen Magnesiumionen sind in gelb dargestellt, ihre koordinierenden Aminosäuren aus β‘ als Stäbchen. Die für die Katalyse wichtigen Elemente BH (rosa) und Trigger-Helices (TH, lila) sind hervorgehoben. Der Übersichtlichkeit halber wurde die β-Untereinheit weggelassen. Die rechte Abbildung vergrößert die Koordination der Magnesiumionen im Kasten links.

Das durch den sekundären Kanal eintretende NTP bindet an die freie Base der DNA-Matrize und bildet entsprechende Wasserstoffbrückenbindungen aus (Zhang *et al.*, 1999; Korzheva *et al.*, 2000). Die Trigger-Schleife (TS, Abb. 1-1) faltet sich daraufhin zu zwei Helices (TH, Abb. 1-3) und verengt den Eintrittskanal der NTPs, wodurch die NTP-Dissoziation verhindert wird (Vassylyev *et al.*, 2007b). Regulatoren, die durch den sekundären Kanal eintreten und das aktive Zentrum beeinflussen, können trotzdem weiterhin wirken (Nickels und Hochschild, 2004). Die Umfaltung der TS wird vor Allem durch die Ausbildung eines 3-Helix-Bündels mit der BH stabilisiert (Vassylyev *et al.*, 2007b). Aminosäuren aus TH und BH binden das NTP und positionieren es für den Einbau in den RNA-Strang. Dabei wird durch Bindung anderer Aminosäuren der β‘-Untereinheit an das 2‘O und 3‘O der Ribose sichergestellt, dass ein NTP und kein Desoxyribonukleosidtriphosphat (dNTP) gebunden wird (Vassylyev *et al.*, 2007b). Weiterhin kontaktieren Aminosäuren aus TH und BH die Basen des NTP-Substrats und der Matrize und scheinen so den Einbau der richtigen Base zu kontrollieren und Purin-Purin oder Pyrimidin-Pyrimidin Basenpaarungen auszuschließen (Wang *et al.*, 2006). Der Zustand mit gefalteten TH wird als „geschlossener, aktiver Zustand“ bezeichnet, im Gegensatz zum „offenen, inaktiven Zustand“ vor der Umfaltung der TS (Vassylyev *et al.*, 2007b).

Die Reaktion des NTP-Einbaus erfolgt an den beiden katalytischen Magnesiumionen (Abb. 1-3, Steitz, 1998). Mg1, das direkt am 3‘-Ende der naszierenden RNA sitzt und vom 3’OH der Ribose koordiniert

ist, wird von der katalytischen Aspartattriade (D739, D741, D742 in *T. thermophilus*) von β‘ fest gebunden. Eine weitere Koordinationsstelle von Mg1 wird vom α-Phosphat des Substrat-NTPs besetzt. Mg2, das nur lose von einem Aspartatrest der katalytischen Triade von β‘ gebundene wird, wird von allen drei Phosphatresten des NTPs koordiniert (Abb. 1-3). Ein weiterer Mg2-Ligand ist ein von β koordiniertes Wassermolekül. Die jeweils freie Koordinationsstelle der beiden Magnesiumionen wird vermutlich von einem einzigen Wassermolekül besetzt, das ebenfalls von der β-Untereinheit positioniert wird und möglicherweise katalytische Funktion besitzt (Vassylyev *et al.*, 2007b).

Zum Einbau des NTPs greift dann das freie 3’OH der naszierenden RNA in einer S_N2-Reaktion nukleophil am α-Phosphat des NTPs an (Yee *et al.*, 1979). Mg1 polarisiert dabei die 3‘OH-Bindung und begünstigt dadurch den nukleophilen Angriff auf das Phosphat des NTPs (Steitz, 1998). Weiterhin stabilisieren beide Magnesiumionen den trigonal-bipyramidalen Übergangszustand und Mg2 unterstützt die Dissoziation des PP_i (Tomar und Artsimovitch, 2013; Steitz, 1998).

Nach der Reaktion wird das PP_i freigesetzt und die TH entfalten sich wieder zur TS. Gemeinsam mit der Umfaltung der TS bewegen sich die Scheren der β- und β‘-Untereinheiten, die die dsDNA krebsscherenartig umschließen. Bei Ausbildung der TH öffnet sich die β-Schere und bildet weniger Kontakte zur stromabwärts gelegenen dsDNA aus (Vassylyev *et al.*, 2007b). Dadurch wird die Mobilität der DNA erhöht und die Translokation der RNAP begünstigt. Die Entfaltung der TH bei Freisetzung des PP_is geht mit einem Schließen der Scheren einher, was die RNAP in einem post-translozierten Zustand hält (Vassylyev *et al.*, 2007b). Somit folgt auf die Freisetzung des PP_i unmittelbar die Translokation der RNAP (Yin und Steitz, 2004). Der geschlossene, aktive Zustand der RNAP mit gefalteten TH und geöffneten Krebsscheren um die DNA wird als hochenergetisch angesehen, der bei der Freisetzung des PP_i durch Entfaltung der TH zur TS und Schließen der Krebsscheren um die DNA relaxiert (Vassylyev *et al.*, 2007b).

Dieser NTP-Einbauzyklus wird während der Transkription permanent wiederholt, bis ein Terminationssignal erreicht wird. Dabei kann es immer wieder zum Anhalten der Transkription kommen. Diese Pausen haben regulatorische Funktion bei der Genexpression (Landick, 2006) oder sollen sicherstellen, dass die Transkription mit der Translation korreliert wird (Winkler und Yanofsky, 1981; Yakhnin *et al.*, 2006). Die allgemeine Transkriptionsgeschwindigkeit der RNAP beträgt *in vivo* etwa 40-90 nt/s (Gotta *et al.*, 1991; Condon *et al.*, 1993). Sie hängt jedoch stark von den Wachstumsbedingungen der Bakterien, dem Codongebrauch und natürlich der Regulation durch eine Vielzahl von Transkriptionsfaktoren ab (Artsimovitch und Landick, 2000).

1.1.4 Termination

Am Ende des offenen Leserasters kommt es zur Termination der Transkription (Abb. 1-2). Dafür gibt es bei Bakterien zwei verschiedene Mechanismen: intrinsische und Rho-abhängige Termination. Bei

der intrinsischen Termination wird auf dem RNA-Transkript aufgrund einer GC-reichen, palindromischen Sequenz eine Haarnadelschleife ausgebildet (Abb. 1-4A), auf die eine uridinreiche Sequenz folgt (Platt, 1981; Brendel *et al.*, 1986). Dadurch wird die Bindung innerhalb des DNA-RNA-Hybrids geschwächt und es kommt zur Freisetzung der RNA und somit zur Termination (Nudler und Gottesman, 2002).

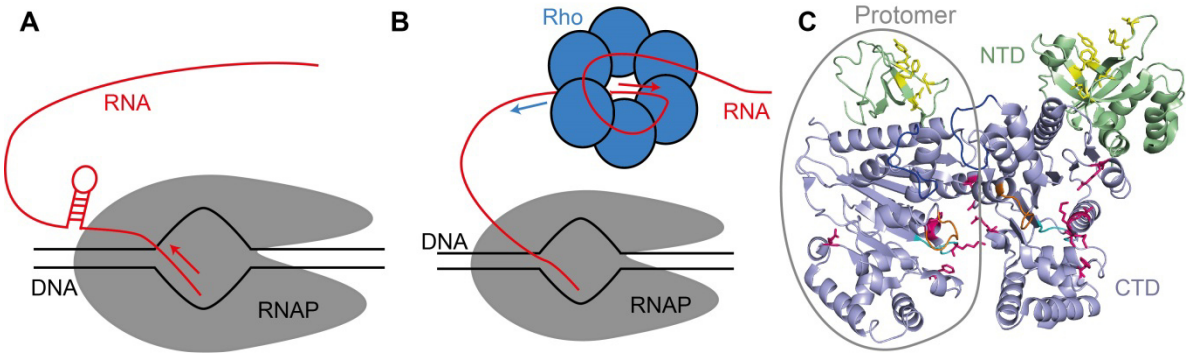


Abbildung 1-4: Terminationsmechanismen der Transkription. (A) Intrinsische Termination, **(B)** Rho-abhängige Termination. RNAP: grau, DNA: schwarz, RNA: rot, Rho: blau. Pfeile geben die Bewegungsrichtungen einzelner Elemente an. **(C)** Kristallstruktur von zwei Protomeren von Rho aus *T. maritima* (PDB-Code: 3L0O). Aminosäuren der primären RNA-Bindestelle in der NTD (hellgrün) sind in gelb hervorgehoben. Die Schleifen der sekundären RNA-Bindestelle in der CTD (hellblau) sind blau, orange und cyan. Aminosäuren der ATP-Bindestelle sind in magenta hervorgehoben.

Bei der Rho-abhängigen Termination wird der Terminationsfaktor Rho benötigt. Rho ist ein homohexameres, ringförmiges Protein, das essentiell für das Überleben der Bakterien ist (Das *et al.*, 1976). Es bindet an die naszierende RNA und gelangt aufgrund seiner ATP-abhängigen 5‘-3‘-Helikaseaktivität zur RNAP (Abb. 1-4, Brennan *et al.*, 1987). Dort ist Rho in der Lage, die Termination zu beenden. Die 70-80 nt lange Bindestelle für Rho auf der RNA wird als Rho-*utilisation site* (*rut*) bezeichnet und kann in seiner Sequenz und Position auf der RNA stark variieren. Charakteristisch für die *rut*-Sequenz sind hauptsächlich die fehlende Sekundärstruktur der RNA und ein hoher Anteil an Cytidinen (Morgan *et al.*, 1985; Zhu und von Hippel, 1998). Sie wird von der primären RNA-Bindestelle von Rho gebunden, die in den N-terminalen Domänen der sechs Untereinheiten lokalisiert ist (Abb. 1-4C, Briercheck *et al.*, 1998; Bogden *et al.*, 1999). Die C-terminalen Domänen sind ringförmig assoziiert und bilden einen Kanal in der Mitte, der die sekundäre RNA-Bindestelle beinhaltet (Miwa *et al.*, 1995; Thomsen und Berger, 2009). Durch eine Ringöffnung gelangt die RNA in diesen Kanal und wird von den RNA-bindenden Schleifen der einzelnen Untereinheiten gebunden (Abb. 1-4C, Skordalakes und Berger, 2003). Der Ring schließt sich wieder und die C-terminalen Domänen hydrolysieren ATP zur Translokation von Rho auf der RNA in Richtung RNAP (Thomsen und Berger, 2009; Kim und Patel, 2001). Die sechs ATP-Bindestellen sind zwischen den einzelnen Protomeren des Rho-Hexamers lokalisiert (Abb. 1-4C) und befinden sich in unterschiedlichen Zuständen der ATP-Bindung bzw. -Hydrolyse. Die RNA-Bindeschleifen der sekundären RNA-Bindestelle sind treppenartig aufsteigend positioniert und erfahren durch die ATP-Hydrolyse eine Bewegung. Durch einen rotierenden ATP-Hydrolyse-Mechanismus wird dann die RNA von den Schleifen durch den Ring gezogen (Thomsen und Berger, 2009).

Für die Translokation von Rho auf der RNA gibt es drei mögliche Mechanismen. Beim „*pure tracking*"-Mechanismus (Geiselmann *et al.*, 1993; Walstrom *et al.*, 1997) wird die RNA von den primären Bindestellen in den NTDs von Rho gebunden, die abhängig vom ATP-Hydrolysezustand der Untereinheiten stark oder schwach an die RNA binden und dadurch daran entlangwandern. Dieser Mechanismus wurde später erweitert, als bekannt wurde, dass der Hydrolysezustand hauptsächlich Einfluss auf die sekundäre Bindestelle von Rho hat. Demnach wandert Rho sowohl mit der sekundären als auch mit der primären Bindestelle an der RNA entlang. Nach dem „*tethered tracking*"-Mechanismus (Steinmetz und Platt, 1994) bleibt die primäre Bindestelle an der *rut*-Sequenz gebunden, während die RNA durch die sekundäre Bindestelle durchgezogen wird und eine größer werdende Schlaufe bildet. Bei dem dritten Mechanismus, dem „*rut-free tracking*" (Thomsen und Berger, 2009), löst sich die primäre Bindestelle von der *rut*-Sequenz; Rho bleibt nur mit der sekundären Bindestelle an die RNA gebunden und wandert an ihr entlang zur RNAP. Das wahrscheinlichste Modell ist der „*tethered tracking*"-Mechanismus, der stark durch Einzelmolekülanalysen unterstützt wird (Koslover *et al.*, 2012).

Für beide Arten der Termination ist der detaillierte Mechanismus der TEC-Dissoziation unklar. Es gibt aber drei Modelle für die RNA-Freisetzung (Tomar und Artsimovitch, 2013; Peters *et al.*, 2011). Im „*hybrid shearing*"-Modell (Toulokhonov und Landick, 2003; Richardson, 2002) wird die RNA aus dem aktiven Zentrum herausgezogen, was die Bindungen des DNA-RNA-Hybrids aufbricht. Dies soll bei der intrinsischen Termination die Bildung der stabilen RNA-Haarnadelschleife bewirken (Toulokhonov und Landick, 2003). Bei der Rho-abhängigen Termination wird angenommen, dass Rho auf der RNA zur RNAP transloziert und die RNA spannt (Richardson, 2002). Die Zugkraft von Rho wird dabei groß genug, um die RNA aus der RNAP herauszuziehen (Dalal *et al.*, 2006; Schwartz *et al.*, 2007).

Beim allosterischen Modell (Epshtein *et al.*, 2007; Toulokhonov *et al.*, 2001) wird angenommen, dass die Haarnadelschleife oder Rho mit dem TEC wechselwirken und dabei das aktive Zentrum verändern, sodass die Transkription beendet wird. Es ist möglich, dass Rho durch seine Helikaseaktivität direkt das DNA-RNA-Hybrid in der RNAP entwindet (Epshtein *et al.*, 2010). Dafür sind jedoch tiefgreifende strukturelle Änderungen in der RNAP notwendig, da sich das DNA-RNA-Hybrid normalerweise in der RNAP befindet und nicht zugänglich ist (Peters *et al.*, 2011). Tatsächlich sind strukturelle Änderungen des TEC der am wenigsten bekannte Teil der Termination (Peters *et al.*, 2011). Es wird jedoch davon ausgegangen, dass sich die Krebsscheren von β und β', die die dsDNA umschließen, öffnen und dass sich die TH zurück zur TS entfalten (Vassylyev *et al.*, 2007a; Epshtein *et al.*, 2010).

Im drittel Modell, dem Hypertranslokations-Modell (Yarnell und Roberts, 1999; Santangelo und Roberts, 2004), wird die RNAP auf dem Templat weitergeschoben, ohne dass weitere Nukleotide angefügt werden. Bei der intrinsischen Termination wird dabei wie beim „*hybrid shearing*"-Modell durch die Bildung der Haarnadelschleife die RNA aus der RNAP herausgezogen, jedoch bleiben dabei die Kontakte zur DNA bestehen und diese wird mit der RNA weitergeschoben. Einhergehend damit

wird die stromaufwärts gelegene DNA wieder zusammengewunden und die stromabwärts gelegene DNA aufgeschmolzen (Yarnell und Roberts, 1999; Park und Roberts, 2006). Bei der Rho-abhängigen Termination übt Rho eine Kraft auf die RNAP aus, die diese zur Translokation ohne Nukleotidaddition veranlasst (Park und Roberts, 2006).

Welcher Mechanismus tatsächlich stattfindet, ist unklar. Bei der intrinsischen Termination gibt es Anhaltspunkte, die den „*hybrid shearing*"- und den Hypertranslokations-Mechanismus favorisieren (Larson *et al.*, 2008). Dabei scheint der genaue Mechanismus von der Sequenz des DNA-RNA-Hybrids abhängig zu sein, je nachdem wie leicht die DNA-RNA-Kontakte brechen, und variiert von Terminator zu Terminator (Larson *et al.*, 2008). Ob auch der Mechanismus der Rho-abhängigen Termination abhängig von der Hybridsequenz ist, ist unklar (Peters *et al.*, 2011). Durch Vernetzungsexperimente konnte aber auch gezeigt werden, dass das 3'OH-Ende der naszierenden RNA im aktiven Zentrum verbleibt, was für den allosterischen Mechanismus der Termination spricht (Epshtein *et al.*, 2010). In jedem Fall wird die Interaktion zwischen DNA und RNA sowie zwischen RNAP und Nukleinsäuren geschwächt, was zur Termination der Transkription führt (von Hippel und Yager, 1992). Nach der Dissoziation des Elongationskomplexes kann die freigesetzte RNAP durch Bindung des σ-Faktors eine neue Transkription beginnen (Abb. 1-2, Mooney *et al.*, 2009a).

1.1.5 Antitermination der Transkription

Antitermination wird der Prozess genannt, bei dem die RNAP Terminationssignale überliest und die Transkription fortsetzt. Entdeckt wurde dies zuerst beim λ-Phagen, für dessen Lebenszyklen Antitermination zur Herstellung wichtiger Proteine notwendig ist (Roberts, 1969; Gottesman und Weisberg, 2004). Bei der Expression der frühen Gene des Phagen wird unter anderem das N-Protein hergestellt. N bindet an die RNAP und ist alleine in der Lage, die Transkription zu modifizieren (Mason *et al.*, 1992b). Dieser Elongationskomplex ist jedoch instabil und kann nur nahe gelegene Terminatoren überlesen (Mason *et al.*, 1992b; Rees *et al.*, 1996). Für effektive Antitermination ist die Ausbildung eines Antiterminationskomplexes bestehend aus mehreren *E. coli*-Proteinen, den N *utilisation substance* (Nus) Faktoren, und bestimmten RNA-Sequenzen notwendig (Abb. 1-5). Das Signal für die Antitermination auf der RNA wird N *utilisation site* (*nut*) genannt. Diese Erkennungsstelle umfasst eine einzelsträngige *BoxA*-Sequenz, eine darauffolgende Haarnadelschleife namens *BoxB* und eine konservierte alternierende GT-Sequenz, die als *BoxC* bezeichnet wird (Abb. 1-5B, Friedman und Olson, 1983; Berg *et al.*, 1989). λN bindet mit seiner NTD an *BoxB* und rekrutiert das *E. coli*-Protein NusA, das an die RNA-Sequenz zwischen *BoxB* und *BoxA* bindet (Lazinski *et al.*, 1989; Scharpf *et al.*, 2000). NusB und NusE binden als Heterodimer an *BoxA* und werden von NusG an der RNAP verankert (Nodwell und Greenblatt, 1993a; Friedman *et al.*, 1984). Dieser stabile Antiterminationskomplex ist in der Lage, intrinsische und Rho-abhängige Termination zu unterdrücken (Gottesman *et al.*, 1980).

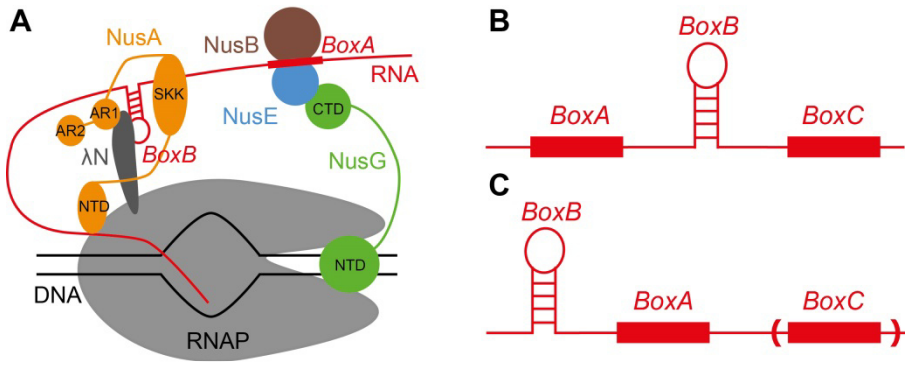


Abbildung 1-5: Antitermination. (A) Antiterminationskomplex des λ-Phagen. λN (dunkelgrau) bindet an die Haarnadelschleife *BoxB* auf der RNA (rot) und rekrutiert NusA (orange). NusB (braun) und NusE (blau) binden als Heterodimer an *BoxA* und werden von NusG (grün) an der RNAP (grau) verankert. Dadurch kann die RNAP die Transkription der DNA (schwarz) über Terminationssignale hinweg fortsetzen (nach Burmann *et al.*, 2010). **(B)** Aufbau der *nut*-Antiterminationssequenz. **(C)** Aufbau der Antiterminationssequenz in ribosomalen Operons von Bakterien. *BoxC* ist nur in der Leitsequenz zu finden.

Auch aus Bakterien ist eine derartige Antitermination bekannt. Zur effizienten Transkription der rRNA ist die Unterdrückung Rho-abhängiger Termination notwendig (Morgan, 1980; Aksoy *et al.*, 1984). In den Operons der ribosomalen RNA (*rrn* Operons) sind die Gene der 16S RNA und der 23S RNA, auf das direkt die 5S RNA folgt, operonspezifisch durch verschiedene tRNA-Gene getrennt (Morgan, 1986). In der Leitsequenz stromaufwärts der 16S RNA sowie in der Trennsequenz stromabwärts der tRNA-Gene und stromaufwärts der 23S RNA sind Antiterminationssequenzen ähnlich der des λ-Phagen lokalisiert (Abb. 1-5C). Diese bestehen ebenfalls aus den Elementen *BoxA* und *BoxB*, jedoch befindet sich hier die *BoxB* stromaufwärts von *BoxA*. *BoxA* ist innerhalb der Bakterien hochkonserviert und weist eine hohe Homologie zur *BoxA* des λ-Phagen auf. *BoxB* bildet wie beim λ-Phagen eine Haarnadelschleife, jedoch ist hier keine Sequenzhomologie erkennbar. *BoxC* ist nur in der Leitsequenz der *rrn* Operons stromabwärts von *BoxA* zu finden (Berg *et al.*, 1989; Li *et al.*, 1984). Für effiziente Antitermination ist aber nur das Vorhandensein von *BoxA* notwendig und ausreichend (Berg *et al.*, 1989; Heinrich *et al.*, 1995). Wie bei der Antitermination des λ-Phagen bilden die Proteine NusA, NusB, NusE und NusG mit diesen RNA-Sequenzen und der RNAP einen Antiterminationskomplex aus (Nodwell und Greenblatt, 1993a; Torres *et al.*, 2004; Vogel und Jensen, 1997). Diese Nus-Faktoren sind jedoch nicht ausreichend für effiziente Antitermination (Squires *et al.*, 1993). Es konnte bisher kein Protein identifiziert werden, das die Funktion des λN-Proteins einnimmt. Stattdessen sind mehrere ribosomale Proteine in die *rrn* Antitermination involviert. Eine wichtige Funktion konnte S4 zugewiesen werden, das an die RNAP bindet und die Termination verzögert, aber auch für L3, L4 und L13 konnte eine Beteiligung nachgewiesen werden (Torres *et al.*, 2001).

Die Funktion der für die Transkription der rRNA notwendigen Antitermination liegt vermutlich in der schnellen Anpassung an Wachstumsbedingungen begründet. Schnelles Wachstum von Bakterien erfordert eine hohe Proteinsyntheserate, für die viele Ribosomen benötigt werden. Daher gehören die *rrn* Operons in schnell wachsenden Bakterien zu den am meisten transkribierten Genen (Schaechter *et al.*, 1958; Klumpp und Hwa, 2008). Die rRNA-Gene weisen eine besonders hohe Dichte an transkri-

bierenden RNAPs (Gotta *et al.*, 1991) mit einer hohen Elongationsgeschwindigkeit auf (Vogel und Jensen, 1994). Diese Elongationsgeschwindigkeit wird durch die Ausbildung des Antiterminations-komplexes gewährleistet, der Pausieren der RNAP unterdrückt (Jin *et al.*, 1992). Derartige Pausen verlangsamen die Elongation deutlich, da schnell RNAP-Staus auf der DNA entstehen können. Für die pausierten RNAPs ist nun die Rho-abhängige Termination wichtig, die dort die Transkription beendet und eine schnelle Elongation der folgenden RNAPs gewährleistet (Klumpp und Hwa, 2008).

Weiterhin ist die Beteiligung von ribosomalen Proteinen an der Antitermination als Rückkopplung ein wichtiger Regulationsmechanismus. Die ribosomalen Proteine kommen in der Zelle nur frei vor, wenn nicht genügend rRNA vorhanden ist und sie nicht in Ribosomen verbaut werden können. In freier Form bewirken sie durch die Antitermination eine erhöhte Produktion von rRNA und dadurch von Ribosomen. Sind dagegen rRNA im Überschuss und ribosomale Proteine im Unterschuss vorhanden, kommen diese Proteine nicht frei vor und die Transkription von rRNA wird terminiert (Nodwell und Greenblatt, 1993b; Condon *et al.*, 1995).

1.2 Nus-Faktoren

Die Transkription wird durch eine Vielzahl von Transkriptionsfaktoren reguliert, die an die RNAP binden und diese modifizieren. Eine wichtige Gruppe stellen die Nus-Faktoren dar, die vor Allem in die Elongations- und die Terminationsphase involviert sind.

1.2.1 NusA

NusA ist ein wichtiger Transkriptionsfaktor, der konserviert aus zwei Domänen besteht (Abb. 1-6). Die NTD ist über einen flexiblen Linker mit der RNA-bindenden Domäne (SKK, bestehend aus S1 und zwei K homologen Domänen, KH1 und KH2) verbunden (Worbs *et al.*, 2001). In manchen α-, β- und γ-Proteobakterien wie *E. coli* sind zwei zusätzliche Domänen vorhanden, die stark negativ geladen sind und daher *acidic repeat* (AR1 und AR2) genannt werden (Eisenmann *et al.*, 2005).

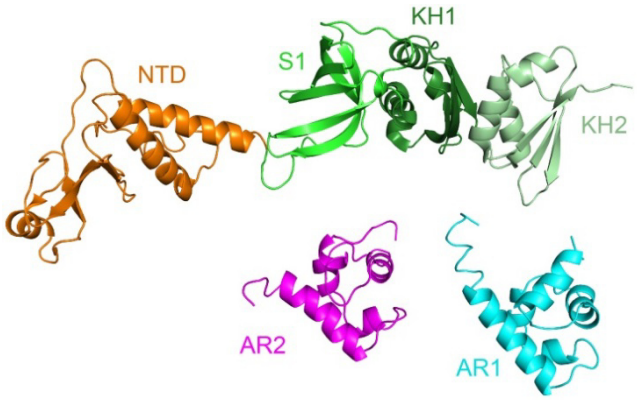


Abbildung 1-6: Struktur von NusA. Kristallstruktur der NTD (orange) und RNA-bindenden Domäne (S1, grün; KH1, dunkelgrün; KH2, hellgrün) von *T. maritima* (PDB-Code: 1HH2) und NMR-Strukturen von AR1 (cyan, PDB-Code: 1WCL) und AR2 (pink, PDB-Code: 1WCN) von *E. coli* sind in der *Cartoon*-Darstellung abgebildet.

NusA hat eine Vielzahl an Einflüssen auf die Transkription. Es modifiziert einerseits den TEC indem es seine Verweildauer bei Pausen verlängert und neue Pausen einführt, die in Abwesenheit von NusA nicht vorkommen. Dadurch wird die Elongationsgeschwindigkeit der RNAP herabgesetzt. Weiterhin moduliert NusA die Termination. Dabei kann es die Termination an intrinsischen und Rho-abhängigen Terminatoren begünstigen, aber auch unterdrücken (zusammengefasst in Borukhov *et al.*, 2005; Roberts *et al.*, 2008). Der durch NusA erzielte Effekt scheint dabei spezifisch vom Terminator abzuhängen (Carlomagno und Nappo, 2003).

NusA-NTD bindet an die RNAP. Ihre Bindestelle konnte über verschiedene Methoden wie Vernetzungsexperimente und eine elektronenmikroskopische Komplexstruktur mit schwacher Auflösung an der FTH am RNA-Austrittskanal (Abb. 1-1) lokalisiert werden (Toulokhonov *et al.*, 2001; Yang *et al.*, 2009; Ha *et al.*, 2010). Die Bindung der NTD an die RNAP ist notwendig und ausreichend für den Einfluss von NusA auf das Pausieren der RNAP (Ha *et al.*, 2010). Da die NTD NusA am RNA-Austrittskanal positioniert, kann die RNA-bindende Domäne SKK direkt mit der naszierenden RNA interagieren. Während der Antitermination (Abschnitt 1.1.5) bindet die SKK bevorzugt an die Region zwischen *BoxA* und *BoxB* (Prasch *et al.*, 2009). Die in z.B. *E. coli* vorhandene Domäne AR1 rekrutiert während der Antitermination des λ-Phagen das λN-Protein (Prasch *et al.*, 2006; Bonin *et al.*, 2004a). AR2 bindet an die SKK und verhindert die RNA-Bindung, was in einer Autoinhibition von NusA resultiert. Diese Autoinhibition kann dadurch aufgehoben werden, dass AR2 an die CTD der α-Untereinheit (α-CTD) der RNAP bindet und die SKK freisetzt (Schweimer *et al.*, 2011; Liu *et al.*, 1996).

1.2.2 NusB und NusE

NusB ist ein vollständig α-helikales Protein, das eine wichtige Rolle bei der Antitermination der Transkription spielt. Es bindet spezifisch an das *BoxA*-RNA-Motiv des Antiterminationssignals in *rrn* Operons oder von lambdoiden Phagen (Nodwell und Greenblatt, 1993b; Mühlberger *et al.*, 2003). Dafür ist vor allem der konservierte Arginin- und Lysin-reiche N-Terminus des Proteins verantwortlich, der an das negativ geladene Phosphatrückgrat der RNA bindet (Huenges *et al.*, 1998). Es konnten aber auch aromatische Aminosäuren als mögliche Stapelpartner für die Basen der RNA identifiziert werden (Bonin *et al.*, 2004b). In manchen Organsimen wie *Mycobacterium tuberculosis* (Gopal *et al.*, 2000) oder *Thermotoga maritima* (Bonin *et al.*, 2004b) bildet NusB Homodimere Dadurch wird die Nukleinsäurebindung verhindert, indem die für die RNA-Bindung identifizierte Fläche verdeckt wird. Dies repräsentiert vermutlich eine Inaktivierung von NusB, wenn es nicht für die Transkriptionsregulation gebraucht wird, was in einer kontrollierbaren RNA-Interaktion von NusB resultiert (Bonin *et al.*, 2004b).

NusB bildet zusammen mit dem Transkriptionsfaktor NusE ein Heterodimer, welches eine erhöhte Affinität zu *BoxA* besitzt. Dies liegt vor Allem an der durch NusE vergrößerten RNA-Bindefläche.

Isoliertes NusE weist nur eine sehr geringe unspezifische RNA-Bindung auf, die ohne NusB ineffizient ist (Greive *et al.*, 2005; Luo *et al.*, 2008). Da Antitermination aber auch in einem NusB-Deletionsstamm stattfinden kann, stellt NusE den aktiven Antiterminationsfaktor dar und die Funktion von NusB beschränkt sich auf die Rekrutierung von NusE (Luo *et al.*, 2008).

Zusätzlich zu seiner Funktion in der Antitermination ist NusE als ribosomales Protein S10 Teil der 30S-Untereinheit des Ribosoms und interagiert mit der 16S-RNA und mehreren Proteinen (Friedman *et al.*, 1981; Wimberly *et al.*, 2000). Dabei spielt es im endgültigen Zusammenbau der 30S-Untereinheit eine wichtige Rolle (Noller und Nomura, 1996; Squires und Zaporojets, 2000). NusE besteht aus einer globulären Domäne und einer Ribosomenbindungsschleife, die sich beim Einbau in das Ribosom zu einem zweisträngigen β-Faltblatt faltet (Wimberly *et al.*, 2000). Diese in Lösung ungefaltete Schleife hat in der Antitermination keine Funktion und sorgt für eine geringe Stabilität von freiem NusE, das zur Aggregation neigt. NusE aus *E. coli* z.B. ist nur im Komplex mit NusB stabil (Luo *et al.*, 2008). Durch seine beiden Funktionen in unterschiedlichen zellulären Zusammenhängen bei gleicher Struktur gehört NusE zu der Gruppe der sogenannten „*Moonlighting*“ Proteine (Jeffery, 1999).

1.2.3 NusG

NusG ist der einzige universell konservierte Transkriptionsfaktor, der auch in Archaeen und Eukaryonten homologe Proteine besitzt (Spt5, Werner, 2012; Hirtreiter *et al.*, 2010). Während der Transkription ist NusG ein essentieller Regulator mit gegensätzlichen Funktionen (Sullivan und Gottesman, 1992; Sullivan *et al.*, 1992). In *E. coli* besteht er aus einer NTD und einer CTD, die über einen flexiblen Linker verbunden sind (Abb. 1-7A, Mooney *et al.*, 2009b). Das konservierte Motiv der NusG N-*terminal domain* (NGN) kommt in allen NusG-ähnlichen Proteinen in Bakterien, Archaeen und Eukaryonten einmal vor (Abb. 1-7B). Diese Domäne bindet an die RNAP und reguliert die Transkription (Hirtreiter *et al.*, 2010; Mooney *et al.*, 2009b; Wada *et al.*, 1998). In manchen Bakterien wie *Thermotoga maritima* oder *Aquifex aeolicus* weist NusG eine nicht konservierte, in die NTD insertierte zusätzliche Domäne auf. Die Funktion dieser Domäne ist bis auf eine unspezifische Nukleinsäurebindung bisher jedoch unklar (Liao *et al.*, 1996; Steiner *et al.*, 2002; Knowlton *et al.*, 2003). Die CTD von NusG, die eine fassartige β-Faltblattstruktur aufweist (Mooney *et al.*, 2009b), umfasst ein Kyrpides-Ouzounis-Woese-Motiv (KOW, Kyrpides *et al.*, 1996), das eine Rekrutierungsplattform für akzessorische Proteine zum Elongationskomplex darstellt (Mooney *et al.*, 2009b; Lindstrom *et al.*, 2003). Archaeales Spt5 weist wie bakterielles NusG eine KOW-Domäne auf, in eukaryontischem Spt5 kommen dagegen mehrere KOW-Domänen vor (Abb. 1-7B, Hirtreiter *et al.*, 2010; Yamaguchi *et al.*, 1999). Archaeales und eukaryontisches Spt5 haben gemeinsam, dass sie ein Heterodimer mit Spt4 bilden (Spt4/5). Spt4 bindet an die NGN und stabilisiert diese. Weitere Funktionen konnten Spt4 bisher nicht zugeordnet werden (Werner, 2007; Guo *et al.*, 2008). Eukaryon-

tisches Spt5 besitzt zusätzlich zwei C-terminale Wiederholungen (C-*terminal repeats*, ctr), die phosphoryliert werden können und dadurch die Aktivität von Spt5 beeinflussen (Yamada *et al.*, 2006).

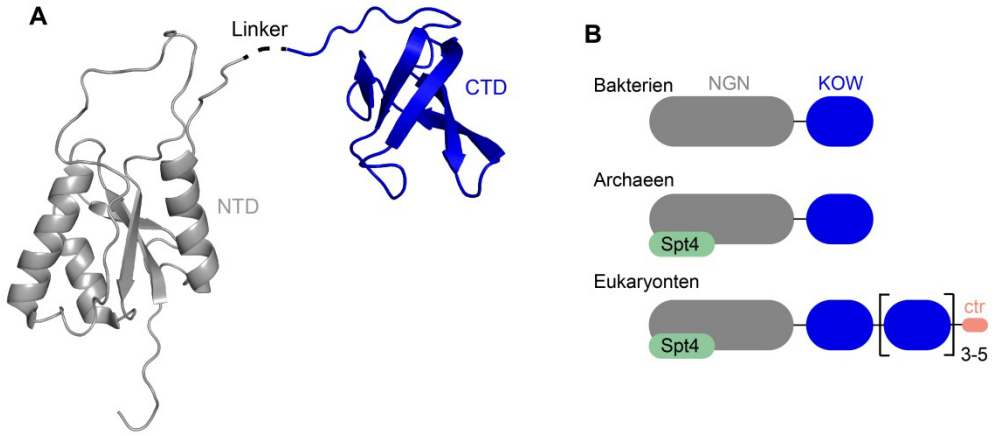


Abbildung 1-7: NusG und seine homologen Proteine. (A) *Cartoon*-Darstellung von *E. coli* NusG in Lösung. NTD (grau, PDB-Code: 2K06) und CTD (blau, PDB-Code: 2JVV) sind über einen flexiblen Linker verbunden. **(B)** Bakterielles NusG und archaeales Spt5 bestehen aus einer NGN- (grau) und einer KOW-Domäne (blau), während eukaryontisches Spt5 über mehrere KOW-Domänen verfügt. Zusätzlich folgen darauf zwei ctr (rosa). Die NGN von Spt5 aus Archaeen und Eukaryonten bildet einen Komplex mit Spt4 (grün; nach Werner, 2012).

NusG bindet nach der Initiation über die NTD an die β‘CH der RNAP, wie es für den Komplex aus Spt4/5 und die Klammerdomäne der RNAP aus *Pyrococcus furiosus* gezeigt werden konnte (Mooney *et al.*, 2009a; Martinez-Rucobo *et al.*, 2011). Als weitere Bindestelle für die NTD konnte für RfaH, ein paraloges Protein von NusG aus *E. coli,* die Schleife der β-Schranke (β *gate loop*, βGL) gefunden werden (Sevostyanova *et al.*, 2011). Der βGL liegt in der β-Krebsschere der RNAP gegenüber der β‘CH (Abb. 1-1). Durch Verbrückung dieser beiden Elemente verschließt die NusG-NTD die Krebsscheren, die die dsDNA halten. Dadurch wird die Mobilität der Krebsscheren eingeschränkt und die Freisetzung der DNA verhindert (Sevostyanova *et al.*, 2011). Dies erhöht die Prozessivität der RNAP und unterdrückt Pausen (Artsimovitch und Landick, 2000; Burova *et al.*, 1995), die vermutlich auch ein Öffnen der Krebsscheren beinhalten (Toulokhonov *et al.*, 2007). Die Krebsschere von β‘ ist über die BH mit der der TH verbunden und beide Elemente können sich gegenseitig beeinflussen (Sevostyanova *et al.*, 2011; Toulokhonov *et al.*, 2007). Daher wird durch das Arretieren der β‘CH die Umorganisation der TH im aktiven Zentrum verhindert, die während Pausen stattfindet.

In *E. coli* bindet NusG mit der CTD an NusE. Dies stellt zum einen eine Verbindung zwischen Transkription und Translation her. Das erste Ribosom wird über NusE (S10) gebunden und durch NusG auf der mRNA rekrutiert, was beide Prozesse der Genexpression kinetisch koppelt (Burmann *et al.*, 2010; Roberts, 2010). Zum anderen bewirkt die NusG-CTD:NusE-Bindung während der Antitermination der Transkription eine Verankerung von NusE an der RNAP und dadurch eine Stabilisierung des Antiterminationskomplexes (Abb. 1-5, Burmann *et al.*, 2010). Im Gegensatz dazu unterstützt NusG auch die Rho-abhängige Termination. Mit der CTD bindet es direkt an Rho und rekrutiert den Terminationsfaktor auf der RNA zur RNAP (Li *et al.*, 1993; Pasman und von Hippel, 2000). Dadurch konkurrieren NusE/S10 und Rho um die Bindung von NusG-CTD. Im Antiterminationskomplex unterstützt die

Anwesenheit der anderen Transkriptionsfaktoren vermutlich die Bindung von NusE, sodass Rho nicht mehr binden kann und die Termination unterdrückt wird (Burmann *et al.*, 2010). Auch während der Transkription eines translatierten Gens wird die Bindung von Rho wahrscheinlich verhindert. Ribosomen beginnen mit der Translation des Gens, sobald die mRNA lang genug ist. Dadurch wird ein großer Teil der mRNA von Ribosomen verdeckt und Rho kann nicht an die RNA binden. Zusätzlich ist die NusG-CTD vom ersten Ribosom besetzt. Wird dieses Ribosom am Ende des offenen Leserasters freigesetzt, kann NusG-CTD an Rho binden und die Termination fördern (Burmann *et al.*, 2010).
Auch Spt4/5 aus Archaeen und Eukaryonten besitzen Funktionen von NusG und können die Prozessivität der RNAP erhöhen und Pausen unterdrücken (Hirtreiter *et al.*, 2010; Wada *et al.*, 1998). Für Eukaryonten konnte außerdem gezeigt werden, dass Spt4/5 an Enzyme bindet, die dafür verantwortlich sind, das 3'OH-Ende der mRNA mit der für die RNA-Prozessierung notwendigen Kappe zu versehen. Außerdem können Mutationen in Spt4/5 das Spleißen der mRNA unterdrücken (Lindstrom *et al.*, 2003). Somit spielt Spt4/5 eine wichtige Rolle in der RNA-Prozessierung, welche in Eukaryonten durch die räumliche Trennung von Transkription und Translation den direkt auf die Transkription folgenden Prozess darstellt. NusG ist daher sowohl strukturell als auch funktional konserviert, was seine Bedeutung für die zelluläre Funktion unterstreicht.

1.3 *Thermotoga maritima*

Thermotoga maritima ist ein gramnegatives, strikt anaerobes Bakterium (Huber *et al.*, 1986). Der Name *Thermotoga* kommt daher, dass es von einer mantelartigen Proteinhülle umgeben ist (Huber *et al.*, 1986; Rachel *et al.*, 1990). Unter optimalen Wachstumsbedingungen verlieren die Bakterien diesen Mantel (Jiang *et al.*, 2006). Gefunden wurde der hyperthermophile Organismus, der bei Temperaturen zwischen 55 °C und 90 °C leben kann, in geothermal erhitztem Meeressediment. Sein Temperaturoptimum liegt bei 80 °C (Huber *et al.*, 1986). Daher benötigt das Bakterium thermostabile Proteine, die besonders widerstandsfähig sind. Diese Proteine ermöglichen einfache und preiswerte Analysen, wodurch *Thermotoga maritima* als Modellorganismus für die Grundlagenforschung dient.

1.4 NMR-Spektroskopie an großen Molekülen

Wie im Abschnitt 1.1 beschrieben, ist strukturell und mechanistisch bereits viel über die RNAP und den Prozess der Transkription bekannt. Die katalytisch wichtigen Bereiche wurden intensiv analysiert und es konnten strukturelle Umlagerungen während der Katalyse identifiziert werden. Für das volle Verständnis des Mechanismus der Transkription ist es jedoch wichtig, Informationen über die Dynamik von Bewegungen innerhalb und zwischen verschiedenen Domänen zu erhalten und transiente Interaktionen der RNAP mit Transkriptionsfaktoren zu untersuchen. Diese Informationen lassen sich weder über Röntgenkristallographie noch über Elektronenmikroskopie erhalten, welches

die beiden bevorzugten Methoden zur Untersuchung großer Proteinkomplexe sind. Die ideale Methode hierfür ist *Nuclear magnetic resonance* (NMR) Spektroskopie. Für NMR-Messungen spielt jedoch die Proteingröße eine entscheidende Rolle. Dipolare ^{1}H-^{1}H- und ^{1}H-^{13}C-Wechselwirkungen führen zu einer schnellen Relaxation der Magnetisierung und dadurch zu einem Signalverlust, der mit zunehmender Molekülmasse des Proteins schwerwiegender wird (LeMaster und Richards, 1988; Torchia *et al.*, 1988). Das Feld der NMR-Spektroskopie an großen Molekülen hat sich in den letzten zwei Jahrzehnten stark weiterentwickelt. Deuterieren des Proteins und dadurch Eliminieren der dipolaren Wechselwirkungen und Verwendung von optimierten Pulssequenzen wie in der *transverse relaxation optimized spectroscopy* (TROSY) führen zu deutlich höheren Signalintensitäten und schärferen Signalen und dadurch zu einer höheren Qualität der Spektren (LeMaster und Richards, 1988; Gardner *et al.*, 1997). Damit sind NMR-Messungen auch mit Proteinen bis 100 kDa möglich (Salzmann *et al.*, 2000; Tugarinov *et al.*, 2002). Weiterhin haben sich Methylgruppen als exzellente NMR-Sonden herausgestellt (Tugarinov und Kay, 2004). Aufgrund ihrer flexiblen Rotation um die Methylachse sind sie auch in Proteinen > 100 kDa beweglich genug um hochaufgelöste Spektren zu liefern (Nicholson *et al.*, 1992). Da Methylgruppen häufig im hydrophoben Proteininneren oder in den Interaktionsflächen von Proteinen lokalisiert sind (Janin *et al.*, 1988) und dabei meistens in Kontakt zueinander stehen, können sie strukturelle Informationen liefern (Gardner *et al.*, 1997; Rosen *et al.*, 1996). Außerdem sind Methylgruppen sehr gut geeignet für die Messung von molekularer Dynamik in einem großen Bereich von Zeitskalen (Mittermaier und Kay, 2006). Der Einsatz von Methylgruppen als NMR-Sonden ist auch dadurch begünstigt, dass ihre Markierung im ansonsten deuterierten Protein einfach durchzuführen ist (Gardner und Kay, 1998; Goto *et al.*, 1999). Für die Markierung der Methylgruppen von Ile, Leu und Val, aber auch die von Ala und Met sind Aminosäurevorstufen in unterschiedlichen Markierungsmustern erhältlich, die für verschiedene Ziele eingesetzt werden können. Diese Vorstufen werden einfach dem zur Zellanzucht verwendeten Minimalmedium zugefügt und dann von den Bakterien bevorzugt verwendet. Die am häufigsten verwendeten Aminosäurevorstufen für Ile, Leu und Val sind α-Ketosäuren, wobei für Leu und Val eine gemeinsame Vorstufe verwendet wird, die die letzte Zwischenstufe in der Valinsynthese darstellt (α-Ketoisovalerat, Tugarinov und Kay, 2003).

Natürlich ist es mit der Markierung von Ile, Leu und Val nicht möglich, wie bei ^{13}C- oder ^{15}N-Markierung eines Proteins, Aussagen über jede Aminosäure zu bekommen. Normalerweise sind Ile, Leu und Val aber relativ gleichmäßig über die Sequenz verteilt, sodass keine größeren Bereiche im Protein ohne Aussage entstehen (Janin *et al.*, 1988) und häufig in Bindungsflächen zu anderen Proteinen oder Liganden zu finden. Zur Bestimmung dieser Bindungsflächen oder um die Dynamik bestimmter Proteinbereiche zu analysieren ist diese Verteilung ausreichend. Die Markierung zusätzlicher Aminosäuren wie Ala und Met führt weiter zur Erhöhung der NMR-Sondendichte im

Protein. Mit dieser Art der Markierung konnten bedeutende NMR-Experimente mit sehr großen Systemen wie dem 20S Proteasom (670 kDa) durchgeführt werden (Sprangers und Kay, 2007).

2 Ziele

Die Transkription ist einer der zentralen Prozesse des Lebens, der strukturell und mechanistisch bereits intensiv untersucht wurde. Auch über die Regulation durch Nus-Faktoren ist besonders bei *E. coli* als Modellorganismus viel bekannt. Trotzdem existiert noch großer Informationsbedarf, z.B. über die Dynamik und transiente Interaktionen der RNAP mit Regulatoren sowie die genauen Bindestellen vieler Transkriptionsfaktoren sowohl auf RNAP- als auch auf Transkriptionsfaktor-Seite, die eine wichtige strukturelle Basis für das Verständnis der Regulation auf molekularer Ebene in atomarem Detail darstellen. Weiterhin ist bisher wenig über Unterschiede der Transkriptionsfaktoren aus *E.* coli zu denen anderer Organismen aufgeklärt. Daher sollten in dieser Arbeit Nus-Faktoren aus dem hyperthermophilen Bakterium *Thermotoga maritima* näher untersucht werden und dabei vor Allem das Protein NusG hinsichtlich der Dynamik seiner Domäneninteraktion und seiner Bindungspartner analysiert werden. Ein besonderer Fokus sollte auf die zusätzliche Domäne von NusG aus *T. maritima* (tmNusG) gelegt werden, deren Funktion noch ungeklärt ist.

Weiterhin sollte der Grundstein für die Untersuchung der *E. coli* RNAP mittels NMR-Spektroskopie gelegt werden. Dazu sollte ein neues Protokoll entwickelt werden, um das aktive Enzym ohne σ-Faktor aus den einzelnen Untereinheiten zu assemblieren. Anschließend sollte es deuteriert und bei einzelnen Untereinheiten selektiv die Methylgruppen ^{1}H,^{13}C-markiert werden, um NMR-Messungen durchzuführen. Auch die isolierten RNAP-Untereinheiten sollten gereinigt und hinsichtlich ihrer Faltung analysiert werden. Weiterhin sollte eine Methode entwickelt werden, um die mit einem Transkriptionsfaktor interagierende Domäne der RNAP zu ermitteln. Auch für die Bestimmung der Bindungsfläche eines Transkriptionsfaktors an die RNAP mittels NMR-Spektroskopie sollte eine Methode erarbeitet werden.

3 Zusammenfassung und Diskussion der Ergebnisse

3.1 Transkriptionsfaktoren aus *Thermotoga maritima*

3.1.1 tmNusEΔ ist in Lösung stabil und interagiert mit tmNusG-CTD

Der für die Antitermination der Transkription (Abschnitt 1.1.5) essentielle Transkriptionsfaktor NusE aus *E. coli* (ecNusE) ist alleine unlöslich und auch nach dem Ersatz der Ribosomenbindungsschleife durch ein einzelnes Serin (ecNusEΔ) ist es nur im Komplex mit ecNusB löslich (Luo *et al.*, 2008). Auch tmNusE neigte zur Aggregation. Im Gegensatz dazu ließ sich tmNusE nach dem Ersatz der Ribosomenbindungsschleife durch ein Serin (tmNusEΔ) in *E. coli*-Zellen produzieren und in für NMR-Messungen geeigneten Mengen löslich reinigen (Einzelarbeit A). Mit einer ^{13}C,^{15}N-markierten Probe wurden bei 50 °C Standard-NMR-Experimente zur Strukturbestimmung gemessen, die Rückgrat- und Seitenkettenresonanzen zugeordnet und mit Hilfe von Abstandsinformationen durch *Nuclear Overhauser Effects* (NOEs) die Struktur bestimmt (PDB-Code: 2MEW, Abb. 3-1, Einzelarbeit A). tmNusEΔ besteht aus einem viersträngigen, antiparallelen β-Faltblatt, das an einer Seite von zwei α-Helices flankiert wird. Damit ist freies tmNusEΔ strukturell fast identisch mit ecNusEΔ, das an ecNusB gebunden ist (Standardabweichung, *root mean square deviation* (*r.m.s.d.*): 1.0 Å, Abb. 3-1B, Luo *et al.*, 2008). Auch die charakteristische cis-Konformation von Pro39, das in der Bindungsfläche von ecNusB liegt, ist in tmNusEΔ vorhanden (hier Pro40) und kein Resultat der Komplexbildung mit ecNusB. Bei der optimalen Wachstumstemperatur für *T. maritima* von 80 °C ist tmNusEΔ gefaltet, da NMR-Spektren auch bei dieser Temperatur noch die charakteristische Dispersion eines strukturierten Proteins zeigen (Abb. 3-1C). Durch eine [^{1}H,^{15}N]- *hetero single quantum coherence* (HSQC)-Titration wurde schließlich nachgewiesen, dass tmNusEΔ an tmNusB bindet (Einzelarbeit A).

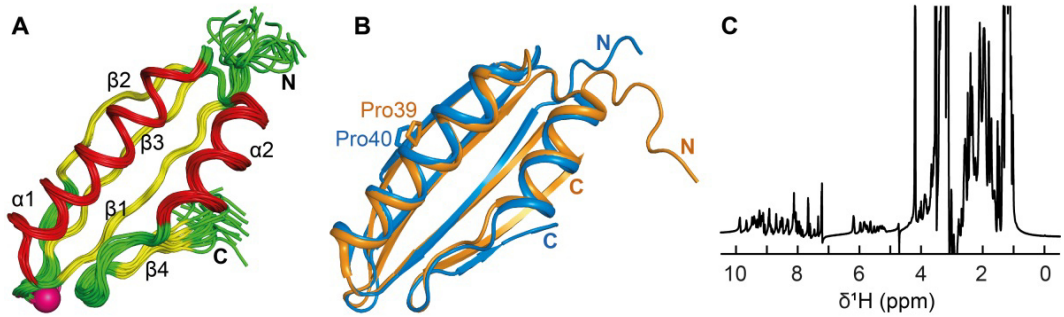


Abbildung 3-1: NMR-Struktur von tmNusEΔ. (A) Überlagerung von 20 akzeptierten Strukturen von tmNusEΔ in der *Cartoon*-Darstellung; Helices: rot, Faltblattstränge: gelb; Schleifen: grün. Das Serin, durch das die Ribosomenbindungsschleife ersetzt ist, ist als pinke Kugel markiert. **(B)** Überlagerung von tmNusEΔ (blau) mit ecNusEΔ (orange) aus dem NusEΔ:NusB-Komplex (PDB-Code: 3D3B). NusB, das an die linke Seite von NusE bindet, ist nicht gezeigt. **(C)** 1D ^{1}H-NMR-Sprektrum von tmNusEΔ bei 80 °C.

Für ecNusE ist bekannt, dass es an die CTD von NusG aus *E. coli* (ecNusG) bindet (Burmann *et al.*, 2010). Auch für tmNusEΔ konnte die Interaktion mit tmNusG-CTD nachgewiesen werden (Einzelarbeit A). Hierfür wurden HSQC-Titrationen von ^{15}N-tmNusEΔ mit tmNusG-CTD und ^{15}N-tmNusG-CTD mit tmNusEΔ durchgeführt und nach jedem Titrationsschritt ein [^{1}H,^{15}N]-HSQC-Spektrum aufgenommen. Bei diesen Spektren, die einen charakteristischen „Fingerabdruck“ eines Proteins darstellen,

wird die chemische Verschiebung von Wasserstoffatomen mit der von direkt gebundenen Stickstoffatomen korreliert, wodurch sich aus dem Peptidrückgrat ein Signal für jede Aminosäure mit Ausnahme von Prolin ergibt. Da die chemischen Verschiebungen der Signale abhängig von der chemischen Umgebung der Aminosäuren sind, verändern sich die Signale im HSQC-Spektrum bei einer Änderung der chemischen Umgebung, wie z.B. bei der Bindung eines anderen Proteins (Skinner und Laurence, 2008). Bei einer solchen Bindung beeinflusst die Austauschrate zwischen freiem und gebundenem Zustand des Proteins die Signale im Spektrum. Bei schnellem Austausch mitteln sich die chemischen Verschiebungen von freiem und gebundenem Zustand und das Signal verschiebt sich mit steigendem Ligand:Protein-Verhältnis. Ist der Austausch zwischen beiden Zuständen langsam, entsteht ein Signal für jeden Zustand und die Signalintensität hängt von der Population des Zustands ab. Beim intermediären Austausch kommt es dagegen zur Linienverbreiterung und dadurch zu einer Intensitätsabnahme der Signale (Skinner und Laurence, 2008; Zuiderweg, 2002). Der jeweilige Austausch ist dabei jedoch nicht nur vom Charakter der Protein:Ligand-Interaktion, sondern auch von der Stärke der Veränderung der chemischen Verschiebung und der Stärke des Magnetfelds abhängig.

Bei den Titrationen von tmNusEΔ und tmNusG-CTD waren deutliche Änderungen der Spektren zu beobachten. Es traten sowohl Signalverschiebungen als auch Intensitätsabnahmen auf. Basierend auf der normierten Änderung der chemischen Verschiebung der Signale wurden die durch die Bindung betroffenen Aminosäuren identifiziert. Damit ließen sich die Bindungsflächen für beide Bindungspartner bestimmen (Abb. 3-2). Bei tmNusEΔ war hauptsächlich der C-Terminus betroffen (Faltblattstrang β4 und die vorhergehende Schleife), aber auch β1 und viele Aminosäuren aus Helix α2. In tmNusG-CTD besteht die Bindestelle aus β4, der Schleife zwischen β1 und β2, sowie Aminosäuren in β3 (Einzelarbeit A). Diese beiden Oberflächen für die Bindung zwischen tmNusEΔ und tmNusG-CTD stimmen sehr gut mit den für die *E. coli*-Proteine bestimmten Bindestellen überein (Burmann *et al.*, 2010).

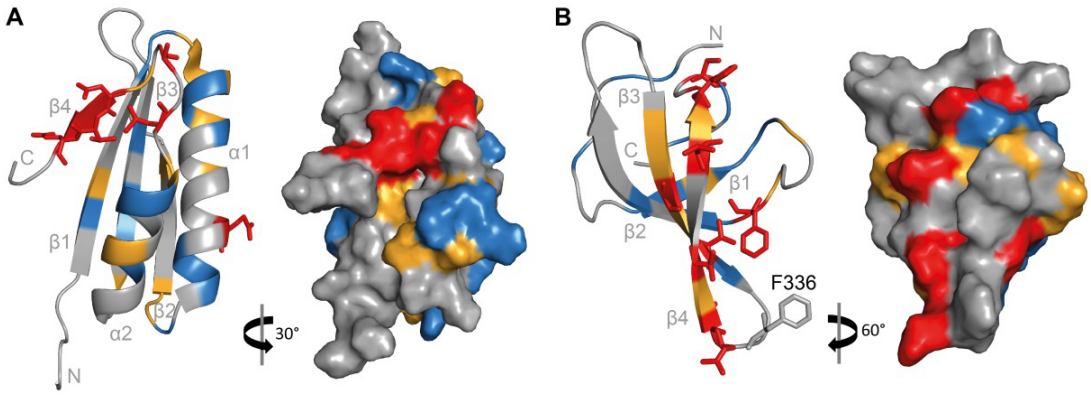


Abbildung 3-2: Interaktionsflächen von tmNusEΔ und tmNusG-CTD. *Cartoon*- und Oberflächendarstellung von **(A)** tmNusEΔ (PDB-Code: 2MEW) und **(B)** tmNusG-CTD (PDB-Code: 2LQ8). Der für die Bindung von tmNusE wichtige Rest F336 in tmNusG-CTD ist als Stäbchen dargestellt und gekennzeichnet. Durch die Bindung stark betroffene Aminosäuren sind rot und durch Stäbchendarstellung hervorgehoben, mittelmäßig stark betroffene Aminosäuren sind orange und schwach betroffene blau.

Für die Bindung zwischen tmNusEΔ und tmNusG-CTD wurde aus den Signalverschiebungen bei steigendem Ligand:Protein-Verhältnis ein K_D-Wert von 13 µM berechnet. Da für NMR-Messungen hohe

Konzentrationen benötigt werden, können nur schwache Wechselwirkungen mit K_D-Werten im höheren mikromolaren bis millimolaren Bereich exakt analysiert werden (Zuiderweg, 2002). Daher ist der tatsächliche K_D-Wert vermutlich kleiner als 13 µM. Im Vergleich zu der Bindung der *E. coli*-Proteine, deren K_D-Wert bei 50 µM liegt (Burmann *et al.*, 2010), scheint die Bindung der Proteine aus *T. maritima* somit stärker zu sein. Obwohl die Messungen mit den *T. maritima*-Proteinen bei 50 °C durchgeführt wurden, einer Temperatur am unteren Ende der Lebensbedingungen von *T. maritima*, ist es möglich, dass die Affinität unter optimalen Lebensbedingungen (80 °C, Huber *et al.*, 1986) anders ist. Die guten Übereinstimmungen der Bindungsflächen von NusEΔ und NusG-CTD aus *T. maritima* und *E. coli* sowie die Tatsache, dass beide Organismen evolutionär nicht nah verwandt sind (Woese, 1987), sprechen für eine konservierte Interaktion zwischen NusE und NusG-CTD. Somit scheinen NusE und NusG in der Antitermination und der Rekrutierung des Ribosoms für Bakterien übereinstimmende Aufgaben zu besitzen.

3.1.2 tmNusG ist in Lösung autoinhibiert

Der Transkriptionsfaktor NusG spielt eine wichtige Rolle in der Rho-abhängigen Termination, der Antitermination der Transkription und der Rekrutierung des Ribosoms, indem er an die RNAP, Rho und NusE bindet (Mooney *et al.*, 2009a; Li *et al.*, 1993; Burmann *et al*., 2010). Die Struktur des NusG-Proteins wurde bisher für mehrere Organismen gelöst und die beiden konservierten Domänen (NTD und CTD) wurden als flexibel verbunden beschrieben (*E. coli*: Mooney *et al.*, 2009b, *T. thermophilus*: Reay *et al.*, 2004, *A. aeolicus*: Steiner *et al.*, 2002). Untersuchungen zeigten jedoch auch, dass NusG aus *A. aeolicus* (aaNusG) in Kristallen teilweise dimerisiert oder oligomerisiert. Dabei binden NTD und CTD verschiedener NusG-Moleküle aneinander. Diese Interaktion scheint aber instabil zu sein und ist leicht aufzubrechen (Knowlton *et al.*, 2003). NMR-spektroskopische Analysen von ecNusG ergaben ebenfalls eine transiente intermolekulare Interaktion zwischen NTD und CTD, die jedoch schwer zu beobachten und vermutlich nicht funktional ist (Burmann *et al.*, 2011). Dagegen wurde für tmNusG in Kristallen nur eine Form gefunden, in der NTD und CTD intramolekular interagieren (Abb. 3-3A, Einzelarbeit B). Hierbei sind die Bindestellen an NusE, Rho und die RNAP verdeckt, wodurch diese Domäneninteraktion möglicherweise eine Autoinhibition darstellt. Auch die NMR-Struktur einer Deletionsvariante von tmNusG (tmNusGΔ), in der die zusätzliche Domäne (DII, Aminosäuren 43-230) durch die an dieser Stelle in ecNusG vorkommende Schleife (Aminosäuren 52-61) ersetzt wurde, zeigt klar eine feste Interaktion zwischen NTD und CTD (Abb. 3-3B). Beide Domänen bewegen sich auf der Zeitskala der Molekülrotation gemeinsam und binden daher fest aneinander (Einzelarbeit B). Der Grund für die Domäneninteraktion ist vermutlich die resultierende, deutliche Stabilisierung von tmNusG, die bei den hohen Temperaturen benötigt wird. tmNusG steht in einem Gleichgewicht von offenem und geschlossenem Zustand (ca. 2 % der Moleküle sind bei 50 °C geöffnet) und weist eine

etwa 50-fach höhere Rate für das Schließen der Domänen auf als für deren Öffnen (Einzelarbeit B). Auch bei 80 °C, dem Temperaturoptimum von *T. maritima* (Huber *et al.*, 1986) interagieren tmNusG-NTD und tmNusG-CTD miteinander, wie die [^{1}H,^{15}N]-HSQC-Spektren von tmNusGΔ und tmNusG-CTD zeigen. Diese weisen große Unterschiede zueinander auf, was bedeutet, dass die Aminosäuren in beiden tmNusG-Konstrukten unterschiedliche chemische Umgebungen erfahren (Einzelarbeit B).

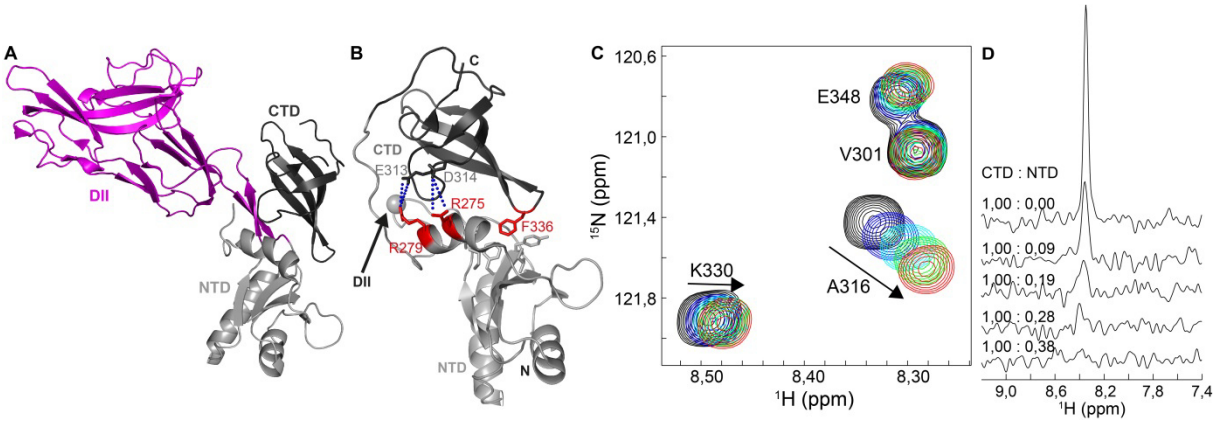


Abbildung 3-3: Domäneninteraktion von tmNusG. (A) *Cartoon*-Darstellung der Kristallstruktur von tmNusG (PDB-Code: 2XHC). NTD: hellgrau, DII: magenta, CTD: dunkelgrau. **(B)** *Cartoon*-Darstellung der NMR-Struktur von tmNusGΔ (PDB-Code 2LQ8), Farbgebung wie in (A). Für die Domäneninteraktion wichtige Aminosäuren sind als Stäbchen dargestellt. R275, R279 und F336 sind rot hervorgehoben. Die Stelle, an der DII insertiert ist, ist markiert. **(C)** Ausschnitt aus den [^{1}H,^{15}N]-HSQC-Spektren der Titration von ^{15}N-tmNusG-CTD mit tmNusG-NTD. Die Pfeile zeigen die Richtung, in die sich die Signale verschieben. **(D)** 1D-Spuren der [^{1}H,^{15}N]-HSQC-Spektren aus (C) von Gly310. Das Verhältnis von CTD:NTD ist angegebenen.

Zur Untersuchung der Dynamik der Domäneninteraktion von tmNusGΔ im Mikro- bis Millisekundenbereich, wurden Relaxationsdispersionsexperimenten bei 50 °C durchgeführt. In der Grenzfläche der Domänen konnte jedoch auf dieser Zeitskala keine Dynamik detektiert werden. Da im [^{1}H,^{15}N]-HSQC-Spektrum nicht für beide Zustände Signale sichtbar sind, muss die Austauschrate zwischen offenem und geschlossenem Zustand in tmNusGΔ im Bereich für schnellen Austausch liegen (Einzelarbeit B). Zur näheren Charakterisierung des Gleichgewichts wurde eine HSQC-Titration von ^{15}N-tmNusG-CTD mit tmNusG-NTD bei 50 °C durchgeführt. Zunächst wurde damit der K_D-Wert für die Bindung der isolierten Domänen auf etwa 5 µM abgeschätzt (Abb. 3-3C, Einzelarbeit B). Die Affinität wird jedoch im Gesamtprotein, in dem tmNusG-NTD und tmNusG-CTD durch einen Linker verbunden sind, vermutlich höher sein. Die verschwindenden Signale von Aminosäuren im intermediären Austausch (Abb. 3-3D) lieferten mit der Dissoziationsrate k_{off} und der Assoziationsrate k_{on} weitere Informationen über die Dynamik der Domäneninteraktion. Die bestimmten Werte ($k_{off} \approx 1000\ s^{-1}$, $k_{on} \approx 2 \cdot 10^{8}\ M^{-1}\ s^{-1}$) sind typisch für diffusionslimitierte Assoziation. Unter der Annahme, dass die Dissoziationsrate des hier verwendeten binären Komplexes identisch ist mit der Öffnungsrate von tmNusG, kann die 50-fach höhere Schließrate des monomeren tmNusGs auf 50.000 s^{-1} abgeschätzt werden. Diese ist höher als die Assoziationsrate im binären Komplex, was den Einfluss des Linkers auf die Domäneninteraktion reflektiert (Einzelarbeit B). Diese Ergebnisse zeigen, dass tmNusGΔ in einem dynamischen Gleichgewicht zwischen offenem und geschlossenem Zustand vorliegt.

Um herauszufinden, ob die Öffnungsrate von etwa 1000 s^{-1} eine kinetische Barriere für die Interaktion von tmNusG mit seinen Bindungspartnern darstellt, wurden NMR-Titrationen mit tmNusEΔ und tmRho durchgeführt. Dazu wurden von ^{15}N-tmNusG und ^{15}N-tmNusGΔ nach jedem Titrationsschritt mit tmNusEΔ 2D [^{1}H,^{15}N]-HSQC-Spektren aufgenommen, nach jedem Titrationsschritt mit tmRho 1D [^{1}H,^{15}N]-HSQC-Spektren. Bindet tmRho an eine tmNusG-Vaiante, erhöht sich die molare Masse drastisch. Dies führt zu einer schnellen Relaxation der Magnetisierung und dadurch zu einer starken Linienverbreiterung, wodurch keine Signale mehr detektiert werden können. Da weder bei der Titration mit tmNusEΔ Signale von tmNusG und tmNusGΔ verschoben wurden oder verschwanden, noch bei der Titration mit tmRho die Signalintensität verringert wurde, können beide Proteine nicht an diese tmNusG-Konstrukte binden, obwohl eine Bindung an die isolierte tmNusG-CTD gezeigt wurde (Einzelarbeit A). Somit stellt der geschlossene Zustand tatsächlich wie postuliert eine autoinhibierte Form vom tmNusG dar, da die tmNusG-CTD in dieser Form weder an tmNusEΔ noch an tmRho binden kann. tmRho und tmNusEΔ können also keine Domänenöffnung induzieren. Genauso wie die Interaktion mit tmRho wurde die Bindung an die RNAP aus *E. coli* (ecRNAP) getestet, da tmRNAP nicht verfügbar war. Die Signalintensität von ^{15}N-tmNusGΔ nahm signifikant ab, was andeutet, dass ecRNAP wenigstens teilweise die Domänenöffnung induziert, sodass tmNusG-NTD an die ecRNAP binden kann. Auch wenn unspezifische Interaktionen der autoinhibierten Form von tmNusGΔ nicht ausgeschlossen werden können, könnte die RNAP das Signal zur Domänenöffnung von tmNusG sein.
Eine vergleichbare Domäneninteraktion konnte bisher nur für das zu NusG paraloge Protein RfaH aus *E. coli* gefunden werden, bei dem die CTD ebenfalls an die RNAP-Bindestelle der NTD bindet. Jedoch weist die CTD dabei eine α-helikale Struktur auf (Belogurov *et al.*, 2007). RfaH ist eine spezialisierte Form von NusG und ist nur in Operons aktiv, die die sogenannte *operon polarity suppressor* (*ops*) Sequenz enthalten. RfaH-NTD interagiert spezifisch mit dem an dieser Stelle pausierten TEC, der die *ops*-Sequenz des einzelsträngigen DNA-Gegenstrangs an der RNAP-Bindestelle der RfaH-NTD exponiert (Artsimovitch und Landick, 2002). Dadurch wird RfaH an die RNAP rekrutiert und durch die Bindung der RfaH-NTD an die β‘CH wird die autoinhibierende Domäneninteraktion aufgehoben, sodass die RfaH-CTD sich in die für NusG typische β-Fassstruktur umfalten und ein Ribosom binden kann (Belogurov *et al.*, 2007; Burmann *et al.*, 2012). Eine derartige regulatorische Funktion der Domäneninteraktion ist für tmNusG aber unwahrscheinlich.

3.1.3 Domänenöffnung von tmNusG ermöglicht die Interaktion mit seinen Bindungspartnern

In den Strukturen von tmNusG konnten in der Interaktionsfläche zwischen NTD und CTD drei Elemente identifiziert werden, die vermutlich für die Domäneninteraktion essentiell sind. Phe336 befindet sich in tmNusG-CTD und ragt in die hydrophobe Tasche von tmNusG-NTD, die an die β‘CH bindet (Abb. 3-3B, Martinez-Rucobo *et al.*, 2011). Außerdem gibt es zwischen tmNusG-NTD und

tmNusG-CTD zwei Salzbrücken, Arg275-Asp314 und Arg279-Glu313 (Abb. 3-3B, Einzelarbeit B). Der Austausch von Arg275 gegen Ala in tmNusGΔ (tmNusGΔ^{R275A}) hatte keine signifikanten Auswirkungen auf das [^{1}H,^{15}N]-HSQC-Spektrum, was bedeutet, dass tmNusGΔ^{R275A} in der geschlossenen Konformation verblieb. Dagegen führte sowohl Eliminierung der Salzbrücke Arg279-Glu313 (tmNusGΔ^{R279A}), als auch die Zerstörung beider Salzbrücken (tmNusGΔ^{R275A,R279A}) zu signifikanten Signalverschiebungen in den [^{1}H,^{15}N]-HSQC-Spektren. In beiden Fällen waren vor Allem Signale der C-terminalen Helix der NTD, in der Arg275 und Arg279 lokalisiert sind, betroffen. Dennoch gab es deutliche Unterschiede zum [^{1}H,^{15}N]-Spektrum der isolierten CTD, d.h. ein Großteil der Moleküle von tmNusGΔ^{R279A} lag noch in der geschlossen Konformation vor und auch die Eliminierung beider Salzbrücken konnte keine komplette Domänenöffnung bewirken (Einzelarbeit B). Der zusätzliche Austausch von Phe336 gegen Ala (tmNusGΔ^{R275A,R279A,F336A}) hatte größere Auswirkungen. Der Vergleich der [^{1}H,^{15}N]-HSQC-Spektren zeigt, dass sich die CTD-Signale mit jedem weiteren Aminosäureaustausch von ihren Resonanzen, die typisch für die geschlossene Form sind, zu den Resonanzen der isolierten tmNusG-CTD bewegen (Abb. 3-4A). In tmNusGΔ^{R275A,R279A,F336A} konnten die meisten CTD-Signale durch Vergleich mit dem Spektrum von isolierter tmNusG-CTD direkt zugeordnet werden (Einzelarbeit A). Somit stellen tmNusGΔ^{R279A}, tmNusGΔ^{R275A,R279A} und tmNusGΔ^{R275A,R279A,F336A} Zwischenstufen des Übergangs von der autoinhibierten zur offenen, nicht inhibierten Form von tmNusG dar und F336 scheint maßgeblich für die Domäneninteraktion verantwortlich zu sein.

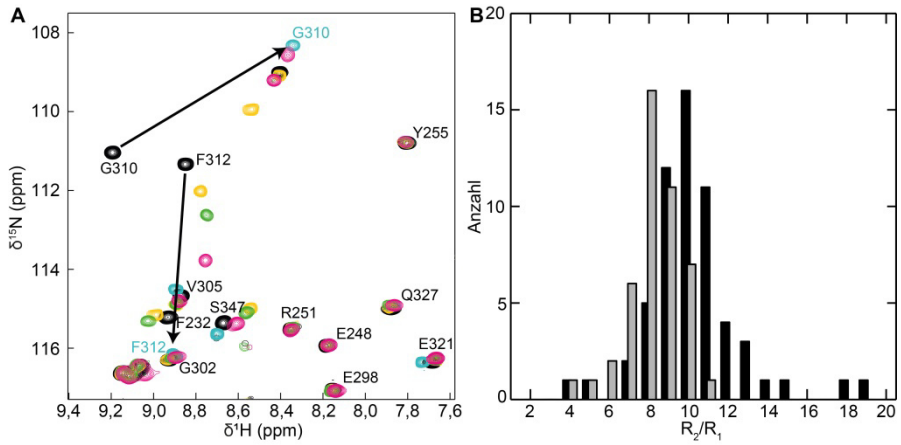


Abbildung 3-4: Domänenöffnung von tmNusGΔ. (A) Ausschnitt der Überlagerung der [^{1}H,^{15}N]-HSQC-Spektren von tmNusGΔ (schwarz), tmNusG-CTD (cyan), tmNusGΔ^{R279A} (gelb), tmNusGΔ^{R275A,R279A} (grün) und tmNusGΔ^{R275A,R279A,F336A} (pink). Die Pfeile zeigen die Verschiebung von Signalen zwischen tmNusGΔ und tmNusG-CTD. **(B)** Auftragung der Verteilung von R_2/R_1 für tmNusGΔ^{R275A,R279A,F336A} bei 50 °C. Daten der NTD sind schwarz, Daten der CTD grau dargestellt.

Die Domäneninteraktion stabilisiert tmNusG-CTD um 10,3 kJ/mol (Einzelarbeit B). Um die Stabilisierung der CTD in den drei tmNusGΔ-Varianten mit geschwächter Domäneninteraktion zu bestimmen, wurde H/D-Austausch gemessen. Bei dieser Methode wird die Tatsache ausgenutzt, dass Amidwasserstoffatome schnell mit den Protonen des Lösungsmittels austauschen. Sind sie jedoch in Wasserstoffbrückenbindungen involviert, ist dieser Austausch stark verlangsamt (Berger und Linderstrom-Lang, 1957; Englander *et al.*, 1996) und kann z.B. mittels NMR-Spektroskopie untersucht werden, da

Deuterium bei Standardexperimenten nicht beobachtet werden kann. Dazu wurden die lyophilisierten, $^{1}H,^{15}N$-markierten Proteine in D_2O gelöst und der Austausch der Amidwasserstoffatome gegen Deuterium durch Aufnahme einer Serie von [$^{1}H,^{15}N$]-HSQC-Spektren verfolgt. Anschließend konnte aus der Signalabnahme ein Schutzfaktor (*protection factor*, PF) für die Amidwasserstoffatome berechnet werden, der eine Messgröße für die Stabilität einer bestimmten Konformation verglichen mit dem ungefalteten Zustand ist (Bai *et al.*, 1993; Huyghues-Despointes *et al.*, 1999). Durch die größere Stabilität sind die PFs in tmNusGΔ signifikant höher als in der isolierten tmNusG-CTD (Eizelarbeit B). Die sukzessive Zerstörung beider Salzbrücken und der F336A-Austausch verringerte die PFs schrittweise, bis die PFs schließlich denen der isolierten CTD entsprachen (Einzelarbeit A). Die Abnahme des Schutzfaktors korreliert mit der Schwächung der Domäneninteraktion. In tmNusGΔ^{R275A,R279A} und tmNusGΔ^{R275A,R279A,F336A} war keine signifikante Stabilisierung mehr messbar, was bedeutet, dass die geschlossene und die offene Form in einem 1:1-Gleichgewicht zueinander stehen (Einzelarbeit A).

Um die Domänenbewegungen von tmNusGΔ^{R275A,R279A,F336A} näher zu untersuchen, wurden Relaxationsmessungen durchgeführt. Hierfür wurden die Raten der longitudinalen Relaxation (R_1) und der transversalen Relaxation (R_2) der Kernspins nach Anregung gemessen. Der Quotient R_2/R_1 liefert Informationen über die Gesamtrotation des Proteins und die Interdomänenbewegungen auf einer Zeitskala schneller als diese Gesamtrotation (Horstmann *et al.*, 2006). Dabei ergab sich eine bimodale Verteilung von R_2/R_1, die die individuelle Rotation der Domänen (NTD und CTD) widerspiegelt (Abb. 3-4B, Einzelarbeit A). Dies ist jedoch nicht so eindeutig wie bei ecNusG (Burmann *et al.*, 2011), bei dem NTD und CTD völlig unabhängig voneinander rotieren und sich nur durch die Linker-Verbindung gegenseitig beeinflussen. Im Gegensatz dazu weist tmNusGΔ eine einheitliche Verteilung für NTD und CTD mit einer Rotationskorrelationszeit von 7,4 ns auf. NTD und CTD in tmNusGΔ bewegen sich auf der Zeitskala der Molekülrotation also gemeinsam und binden fest aneinander (Einzelarbeit B). Zwar deutet die bimodale Verteilung von R_2/R_1 bei tmNusGΔ^{R275A,R279A,F336A} darauf hin, dass NTD und CTD nicht mehr als starre Einheit rotieren wie bei tmNusGΔ, jedoch sind sich die Rotationskorrelationszeiten von 15,5 ns (NTD) und 13,4 ns (CTD) noch sehr ähnlich. Daher ist davon auszugehen, dass sie sich weiterhin gegenseitig beeinflussen, wie auch aus dem Vergleich der [$^{1}H,^{15}N$]-HSQC-Spektren geschlossen werden kann. Die Domäneninteraktion ist somit nicht vollständig auf die Einflüsse der beiden Salzbrücken und des Phe336 zurückzuführen, sondern es sind zusätzliche Kontakte beteiligt. Trotzdem wird vermutlich der größte Beitrag von diesen drei Faktoren geliefert.

Das durch Eliminierung der Salzbrücken in der Domäneninteraktion geschwächte tmNusGΔ^{R275A,R279A} war, im Gegensatz zu tmNusGΔ^{R279A}, in der Lage, mit tmNusEΔ und tmRho zu interagieren (Einzelarbeit A). Hier ist die Domäneninteraktion offensichtlich trotz der Anwesenheit von Phe336 ausreichend geschwächt, um eine Bindung mit den Interaktionspartnern der CTD zu ermöglichen. tmNusGΔ^{R275A,R279A,F336A} zeigte weder eine Interaktion mit tmNusEΔ noch mit tmRho. Phe336 konnte

in der isolierten CTD nicht zugeordnet werden, liegt aber in der Interaktionsfläche zu tmNusEΔ (Abb. 3-2B). In ecNusG-CTD ist das entsprechende Phe (Phe165) essentiell für die Bindung zu ecNusEΔ (Burmann *et al.*, 2010). Somit führt die Eliminierung dieses Phenylalanins auch in tmNusG zur Zerstörung der Interaktion mit tmNusEΔ. Da die Bindestellen von NusG-CTD an NusE und Rho überlappen (Burmann *et al.*, 2010), ist Phe336 vermutlich auch für die Bindung an tmRho wichtig und folglich kann tmNusGΔ^{R275A,R279A,F336A} weder mit tmNusEΔ noch mit tmRho interagieren.

Die Domäneninteraktion von tmNusG ist vermutlich die Basis für die erforderliche Thermostabilität des Proteins. Während intermolekulare Interaktionen zwischen NTD und CTD bei bakteriellen NusGs beschrieben wurden (Knowlton *et al.*, 2003; Burmann *et al.*, 2011), ist die intramolekulare Bindung von NTD und CTD in tmNusG einzigartig. Die beiden Salzbrücken, die die Domänen verbinden, (Arg275-Asp314, Arg279-Glu313, Abb. 3-3B) und Phe336, das in die hydrophobe Tasche der NTD ragt, leisten wichtige Beiträge zur Domäneninteraktion. Phe336 ist in Bakterien hochkonserviert, während die Aminosäurekombination für die Bildung der Salzbrücken nicht konserviert ist. Nur in NusG aus *T. thermophilus* (ttNusG) sind ein Arg und ein Asp vorhanden, die Arg275 und Asp314 entsprechen, jedoch konnte keine NTD:CTD-Interaktion für ttNusG beobachtet werden (Reay *et al.*, 2004). Obwohl transiente intra- oder intermolekulare NTD:CTD-Interaktionen in allen NusG-Proteinen vorkommen könnten, hat nur tmNusG Merkmale entwickelt, um die autoinhibierte Form zu stabilisieren. Diese Form dient vermutlich nur der Thermostabilität und hat nicht wie in RfaH regulatorische Funktion (Belogurov *et al.*, 2007; Burmann *et al.*, 2012)

3.1.4 DII rekrutiert tmNusG an den TEC

Wie auch aaNusG verfügt tmNusG über die zusätzliche Domäne DII. Als Hinweis auf ihre Funktion konnte bisher nur eine unspezifische Nukleinsäurebindung gefunden werden (Liao *et al.*, 1996; Steiner *et al.*, 2002). Mit NMR-Messungen wurde deshalb zunächst getestet, ob tmNusG-DII wie auch tmNusG-NTD, an die RNAP bindet. Dazu wurden die Untereinheiten der tmRNAP einzeln in *E. coli* überproduziert und gereinigt. Da tmβ‘ jedoch während der Genexpression abgebaut wurde, wurde die daran bindende Untereinheit tmω koexprimiert, um den Abbau zu verhindern (Einzelarbeit A, vgl. Abschnitt 1.1.1). tmα und tmβ wurden von Martin Strauß (Lehrstuhl Biopolymere, Universität Bayreuth) gereinigt. Die Intaktheit der tmRNAP-Untereinheiten wurde durch Bindung an tmNusG-NTD und tmNusG-CTD überprüft. 1D [^{1}H,^{15}N]-HSQC-Experimente zeigten, dass tmNusG-NTD mit tmβ und tmβ‘:tmω, jedoch nicht mit tmα interagiert, während tmNusG-CTD an keine Untereinheit der tmRNAP band (Einzelarbeit A). Daher kann davon ausgegangen werden, dass die isolierten Untereinheiten der tmRNAP funktional sind. Da assemblierte RNAP aus *T. maritima* nicht verfügbar war, wurde zunächst eine Interaktion von tmNusG-DII mit ecRNAP nachgewiesen. Die Experimente mit

den einzelnen tmRNAP-Untereinheiten zeigten, dass tmNusG-DII, wie auch tmNusG-NTD, an tmβ und tmβ‘, aber nicht an tmα bindet (Einzelarbeit A).

Um die Nukleinsäureinteraktion von tmNusG zu charakterisieren, wurden Fluoreszenz-Anisotropie-Messungen mit 6-Carboxyfluorescein (6-FAM)-markierten, zufälligen DNA- und RNA-Sequenzen mit einer Länge von etwa 20 nt durchgeführt, die mit den einzelnen Domänen von tmNusG titriert wurden (Einzelarbeit A). Dabei konnte sowohl für tmNusG-NTD als auch tmNusG-DII eine sequenzunabhängige Bindung an alle verwendeten Nukleinsäuren ermittelt werden, während tmNusG-CTD keine Bindung zeigte. Für die nachgewiesene Nukleinsäureinteraktion einer DII-Deletionsvariante von tmNusG (vergleichbar mit tmNusGΔ, Liao *et al.*, 1996) ist somit tmNusG-NTD verantwortlich. Diese unterscheidet jedoch nicht zwischen einzelsträngiger (*single stranded*, ss) DNA, dsDNA und RNA, sondern bindet alle Nukleinsäuren mit einer Affinität im niedrigen mikromolaren Bereich (Einzelarbeit A). tmNusG-DII bindet ebenfalls mit einem K_D im niedrigen mikromolaren/hohen nanomolaren Bereich an ssDNA und RNA, jedoch deutlich stärker an dsDNA (K_D = 0,04 µM). tmNusG-NTD könnte, gebunden an den TEC, den Gegenstrang kontaktieren, so wie es für RfaH vorgeschlagen ist (Artsimovitch und Landick, 2002). Möglicherweise gilt dies generell für NusG-Proteine, auch wenn es bisher für andere NusG-NTDs nicht gezeigt wurde. Die hauptsächliche Nukleinsäurebindung von tmNusG kann aber tmNusG-DII zugeordnet werden, die hochaffin dsDNA bindet. Die Bindestelle von tmNusG-DII an Nukleinsäuren wurde mittels NMR-Titration von ^{15}N-tmNusG-DII mit ssDNA bestimmt (Einzelarbeit A). Obwohl die betroffenen Aminosäuren keine komplett durchgehende Fläche ergeben, zeigt der Vergleich mit dem elektrostatischen Oberflächenpotenzial, dass die bestimmte Bindestelle mit einer positiv geladenen Fläche überlagert und mit hoher Wahrscheinlichkeit der Interaktionsfläche für Nukleinsäuren entspricht (Abb. 3-5).

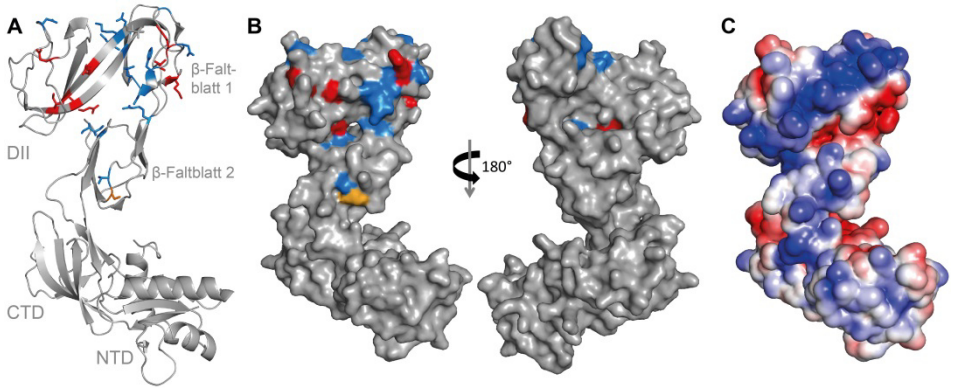


Abbildung 3-5: Nukleinsäurebindestelle von tmNusG-DII. *Cartoon*- **(A)** und Oberflächendarstellung **(B)** von tmNusG (PDB-Code: 2XHC). Durch die Bindung von ssDNA stark betroffene Aminosäuren sind rot hervorgehoben, weniger stark betroffene Aminosäuren sind orange und schwach betroffene blau gefärbt. **(C)** Elektrostatisches Oberflächenpotenzial von tmNusG, berechnet mit dem *Adaptive Poisson-Boltzmann Solver* (APBS)-Programm (Baker *et al.*, 2001), gefärbt von -3 (rot) bis +3 kT/e (dunkelblau).

Die Domäneninteraktion in tmNusG, durch die die Bindestellen an NusE, Rho und RNAP verdeckt sind, führt zur Autoinhibition von tmNusG. Eine Aufhebung dieser Autoinhibition ist aber notwendig, damit tmNusG die für ecNusG bekannten Funktionen erfüllen kann. Vermutlich sind mehrere Fakto-

ren notwendig, um die Autoinhibition aufzulösen. Möglicherweise ist die tmRNAP selbst in der Lage, eine Domänendissoziation zu bewirken. Unter physiologischen Bedingungen bindet tmNusG aber nicht an die RNAP sondern den TEC. Daher kann eine zusätzliche Bindung von tmNusG-NTD an den einzelsträngigen DNA-Gegenstrang, der an der Bindestelle der NTD exponiert ist, die Affinität von tmNusG an den TEC wie bei RfaH (Artsimovitch und Landick, 2002) sowie die Stabilität des gesamten Komplexes erhöhen. Da tmNusG-DII jedoch auch an die RNAP bindet, könnte diese Wechselwirkung tmNusG-NTD in die Nähe des TEC bringen und somit die Bindung an die β'CH deutlich begünstigen. Ist tmNusG-NTD an die RNAP gebunden, so liegt tmNusG-CTD frei vor und kann seine Bindungspartner rekrutieren. Auch tmNusG-DII ist für Interaktionen verfügbar, wie das Modell in Abb. 3-6 zeigt. Aufgrund ihrer bevorzugten Bindung an dsDNA könnte tmNusG-DII mit stromaufwärts gelegener dsDNA interagieren, wenn tmNusG an den TEC rekrutiert wurde. Diese Wechselwirkung sowie die Bindung an die RNAP (β- und/oder β'-Untereinheit) dienen wahrscheinlich der Stabilisierung der tmNusG:RNAP-Interaktion, die bei Temperaturen bis zu 90 °C, bei denen *T. maritima* leben kann (Huber *et al.*, 1986), nötig ist.

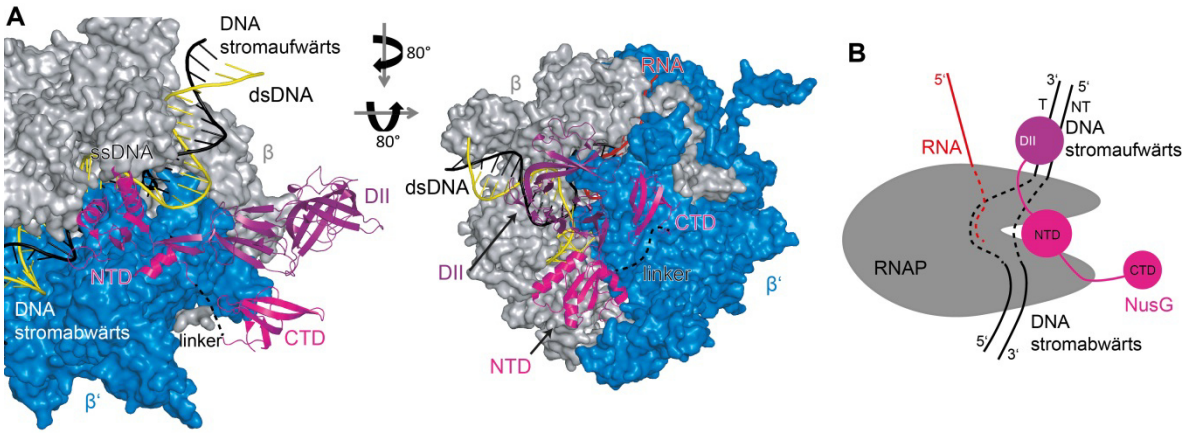


Abbildung 3-6: Modell der Bindung von tmNusG an den TEC von *T. thermophilus* (ttTEC). (A) Oberflächendarstellung des ttTEC (grau: β, blau: β', schwarz: codierender DNA-Strang, gelb: Gegenstrang, rot: RNA; PDB-Code: 2O5I). tmNusG in *Cartoon*-Darstellung (pink, PDB-Code: 2XHC) und der ttTEC wurden mit der Kristallstruktur des *P. furiosus* Spt4/5-Komplexes gebunden an die RNAP-Klammerdomäne (Martinez-Rucobo *et al.*, 2011) überlagert. **(B)** Schematische Darstellung der RNAP- und Nukleinsäurebindung von tmNusG. Über die NTD an die Scheren der RNAP gebunden kann tmNusG-DII an stromaufwärts gelegene dsDNA und/oder die RNAP binden.

Wie *T. maritima* ist *A. aeolicus* ein hyperthermophiles Bakterium und kann bei Temperaturen zwischen 67 und 95 °C leben, wobei das Optimum bei 85 °C liegt (Huber *et al.*, 1992). Wie tmNusG besitzt auch aaNusG eine zusätzliche Domäne, die ebenfalls unspezifisch an Nukleinsäuren bindet. Mit nur einem β-Faltblatt ist sie jedoch deutlich kleiner als tmNusG-DII und weist keine Sequenzhomologie zu tmNusG-DII auf (Steiner *et al.*, 2002, Einzelarbeit B). Außerdem unterliegt aaNusG keiner Autoinhibition (Steiner *et al.*, 2002; Knowlton *et al.*, 2003). Daher könnte die zusätzliche Domäne von NusG in manchen (hyper)thermophilen Organismen eine Anpassung an die hohen Temperaturen ihrer Lebensräume sein, da sie durch Nukleinsäure- und/oder RNAP-Bindung eine Stabilisation der Interaktion mit dem TEC erlaubt.

3.2 Analyse der RNAP aus *E. coli*

3.2.1 Zusammenbau der RNAP aus den isolierten Untereinheiten

Mechanistisch und strukturell konnte bereits viel über den Prozess der Transkription in Erfahrung gebracht werden. Die Regulation der Transkription basiert auf transienten Interaktionen und der Dynamik des Proteins. Eine gute Methode, um Informationen darüber zu erhalten, ist die NMR-Spektroskopie. Da die RNAP mit etwa 400 kDa (Gasteiger *et al.*, 2003) eine sehr große molare Masse besitzt, sind spezielle NMR-Techniken zu ihrer Untersuchung notwendig (Abschnitt 1.4). Doch auch bei selektiver Markierung einzelner Aminosäuren wie Ile, Leu oder Val enthält das NMR-Spektrum bei einem derartig großen Protein eine sehr großen Anzahl an Signalen, sodass es zu vielen Signalüberlagerungen kommt und das Spektrum unübersichtlich wird. Um gezielt einzelne Untereinheiten im Gesamtprotein beobachten zu können, wurden die Gene der RNAP-Untereinheiten aus *E. coli* zunächst einzeln kloniert und das Enzym aus den isoliert überexprimierten Untereinheiten zusammengebaut (Einzelarbeit C). Diese Assemblierung wurde schon mehrfach in der Literatur beschrieben (Tang *et al.*, 1995; Heil und Zillig, 1970; Borukhov und Goldfarb, 1993), jedoch wurde dabei immer der σ-Faktor verwendet, aber nicht ω, welches erst in den späten 1990er Jahren als RNAP-Untereinheit identifiziert wurde (Gentry und Burgess, 1993; Dove und Hochschild, 1998). Wie in Abschnitt 1.1.1 beschrieben, ist ω sehr wichtig für die korrekte Faltung von β‘ und deren Assemblierung mit dem $\alpha_2\beta$-Komplex (Ghosh *et al.*, 2001). Daher wurde für Einzelarbeit C die ω-Untereinheit mit einbezogen, aber gleichzeitig der σ-Faktor ausgeschlossen, da hauptsächlich die elongierende RNAP und nicht der Initiationskomplex von Interesse war. Die Gene der vier RNAP-Untereinheiten (α, β, β‘ und ω) wurden separat in *E. coli* exprimiert und die Zellpellets in Puffer mit 8 M Harnstoff gelöst und vor dem Zellaufschluss vereinigt. Die Denaturierung war notwendig, da β nicht löslich exprimiert werden konnte. Der Harnstoff wurde durch schrittweise Dialyse entfernt und mit der dabei zusammengebauten RNAP wurde eine Affinitätschromatographie an einer Ni^{2+}-Säule durchgeführt, da β‘ C-terminal mit einem His_6-Anhang fusioniert war. Abschließender wurde eine Größenausschlusschromatographie (*size exclusion chromatography*, SEC) durchgeführt, bei der die RNAP in zwei Elutionsgipfeln mit vergleichbarer Absorption bei 280 nm eluierte (Einzelarbeit C). Molekulargewichtsanalyse mit Standardproteinen und Vergleich mit einer RNAP, deren Gene aus einem Operon auf einem Plasmid überexprimiert und die in Zellen assembliert wurde ($RNAP^{nativ}$), ergab, dass die RNAP aus Elutionsgipfel 2 die richtige Größe besaß. Ein Aktivitätstest zur Verlängerung eines fluoreszenzmarkierten RNA-Primers zeigte für die RNAP aus Elutionsgipfel 1 keine Aktivität, für die RNAP aus Elutionsgipfel 2 aber eine Aktivität vergleichbar zu der von $RNAP^{nativ}$ (Abb. 3-7A). Daher war die RNAP aus Elutionsgipfel 2 ($RNAP^{aktiv}$) im Gegensatz zur RNAP aus Elutionsgipfel 1 ($RNAP^{inaktiv}$) korrekt zusammengebaut und aktiv (Einzelarbeit C). Die Sekundärstrukturanalyse mittels Circulardichroismus (CD) Spektroskopie zeigte für $RNAP^{aktiv}$ und $RNAP^{inaktiv}$ Spektren mit Charakteristika gefalteter

Proteine. Dabei ist das Fern-Ultraviolett (UV) CD-Spektrum der RNAP^{aktiv} dem der RNAP^{nativ} sehr ähnlich, während das CD-Spektrum von RNAP^{inaktiv} deutlich geringere Intensitäten und für α-helikale Strukturen weniger ausgeprägte charakteristische Minima aufweist (Abb. 3-7B, Einzelarbeit C). Dies zeigt, dass RNAP^{aktiv} genauso gefaltet ist wie RNAP^{nativ}, RNAP^{inaktiv} dagegen fehlgefaltet vorliegt. Zusätzliche SEC-Läufe zeigten weiterhin, dass RNAP^{aktiv} und RNAP^{inaktiv} nicht in einem Gleichgewicht zueinander stehen, sondern RNAP^{aktiv} auf der Zeitskala dieser Chromatographiemethode stabil ist. De- und renaturierte RNAP^{inaktiv} dagegen eluierte wieder in zwei Elutionsgipfeln als RNAP^{aktiv} und RNAP^{inaktiv}, wodurch die Ausbeute an aktiver RNAP erhöht werden konnte (Einzelarbeit C).

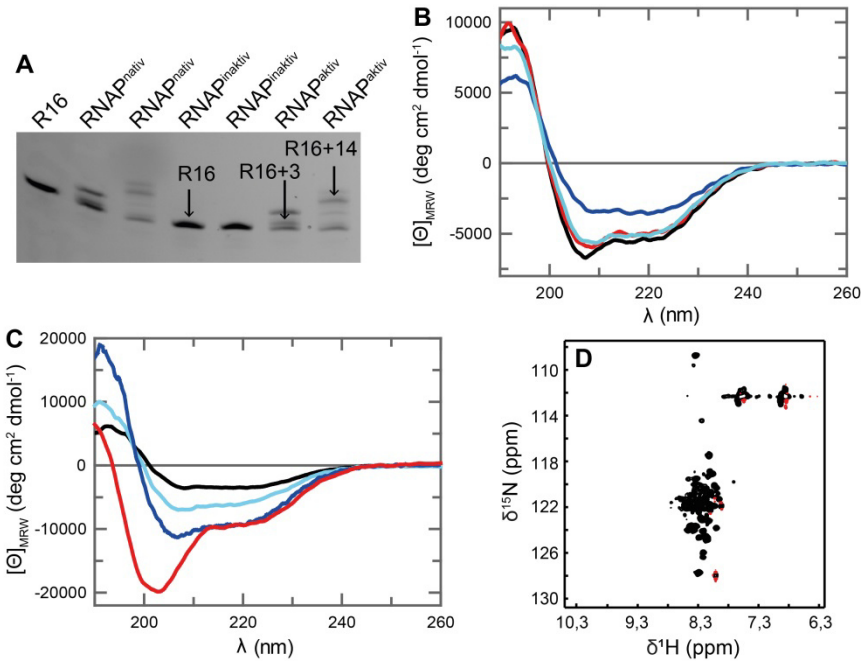


Abbildung 3-7: Aktivitäts- und Faltungsanalyse der RNAP. (A) 20 % (w/v) Polyacrylamidgel des Aktivitätstest der RNAP. ATP und CTP (Bahnen 2, 4, 6, 8) oder ATP, CTP und GTP (Bahnen 3, 5, 7, 9) wurden für eine Verlängerung der RNA (R16) um 3 bzw. 14 nt hinzugefügt. Es wurden 3 pmol RNA aufgetragen. **(B)** Fern-UV CD-Spektren von 0,6 µM RNAP^{nativ} (schwarz), 0,6 µM RNAP^{inaktiv} (blau), 0,5 µM RNAP^{aktiv} (rot) und 0,6 µM ββ'-Komplex (cyan). **(C)** Fern-UV CD-Spektren der isolierten RNAP-Untereinheiten: 2,5 µM α (blau), 0,6 µM β (cyan), 1,1 µM β' (schwarz), 10 µM ω (rot). **(D)** [^{1}H,^{15}N]-HSQC-Spektrum von 300 µM ω.

Die Gesamtausbeute und Reinheit der assemblierten RNAP^{aktiv} ist vergleichbar mit früheren Protokollen ohne die ω-Untereinheit (Tang *et al.*, 1995; Heil und Zillig, 1970; Borukhov und Goldfarb, 1993). Trotzdem hat die in Einzelarbeit C gezeigte Assemblierung der RNAP mehrere Vorteile. In den vorherigen Veröffentlichungen wurden eine oder alle Untereinheiten der RNAP vor der Assemblierung gereinigt. Durch Zusammenbau der Untereinheiten aus den Zellpellets ohne vorherige Reinigung konnten wiederholte De- und Renaturierungsschritte vermieden werden und viel Zeit gespart werden. Die Verwendung der ω-Untereinheit für die Assemblierung führte möglicherweise zu einer Erhöhung des Anteils an korrekt gefalteter aktiver RNAP. Aus Bakterienstämmen, in denen das Gen für ω (*rpoZ*) ausgeknockt ist, kann die RNAP nur im Komplex mit dem Chaperon GroEL gewonnen werden. Die Entfernung von GroEL führt zur Inaktivierung der RNAP, die Aktivität kann aber durch Renaturierung in Gegenwart von ω wiedergewonnen werden (Mukherjee *et al.*, 1999). Somit hat ω

durch seine Rolle im Zusammenbau der RNAP auch einen gewissen Einfluss auf die Aktivität der RNAP. Durch die Verwendung der SEC als letzten Reinigungsschritt konnte selektiv $RNAP^{aktiv}$ gereinigt werden und alle fehlgefalteten RNAPs ausgeschlossen werden, was in den meisten früheren Protokollen nicht der Fall war. Es konnte also eine schnelle, einfache Methode entwickelt werden, um aktive, korrekt gefaltete RNAP aus isolierten Untereinheiten zu assemblieren. Diese Reinigung erlaubt einfache Markierung einzelner Untereinheiten im Kontext der gesamten funktionalen RNAP.

3.2.2 Reinigung der isolierten RNAP-Untereinheiten und Bindung von Transkriptionsfaktoren

Zusätzlich zur Assemblierung der RNAP aus den isolierten Untereinheiten wurden diese auch einzeln gereinigt, wobei die Reinigungen von α und β von Martin Strauß (Lehrstuhl Biopolymere, Universität Bayreuth) durchgeführt worden sind. Es war im Gegensatz zu früheren Publikationen (z.B. Borukhov und Goldfarb, 1993; Igarashi und Ishihama, 1991; Katayama *et al.*, 2000) möglich, β‘ löslich zu exprimieren. Der Komplex aus β und β‘ konnte analog zur gesamten RNAP assembliert werden. Dessen Fern-UV CD-Spektrum war nahezu identisch mit denen von $RNAP^{nativ}$ und $RNAP^{aktiv}$ (Abb. 3-7B, Einzelarbeit C), da β und β‘ die größten Untereinheiten der RNAP sind und am meisten zur globalen Faltung beitragen. Dies deutet darauf hin, dass dieser Komplex genauso assembliert werden kann wie die gesamte RNAP (Abschnitt 3.2.1). Die CD-Spektren von α, β und β‘ zeigten deutliche Charakteristika gefalteter Proteine (Abb. 3-7C). ω dagegen besitzt eine strukturierte NTD mit einem unstrukturierten C-Terminus (Ghosh *et al.*, 2001). Das CD-Spektrum glich nicht vollständig dem eines entfalteten Proteins (Greenfield und Fasman, 1969), deutete jedoch auf wenig Sekundärstruktur hin. Auch das [^{1}H,^{15}N]-HSQC-Spektrum zeigte sehr geringe Signaldispersion (Abb. 3-7D), was ebenfalls ein Zeichen für wenige bis keine stabilen Sekundärstrukturelemente ist. Die isolierte ω-Untereinheit ist also kaum gefaltet und erreicht seine Struktur wahrscheinlich erst durch Bindung von β‘ oder der gesamten RNAP. Da ω für die Faltung von β‘ wichtig ist und es vor Aggregation schützt (Ghosh *et al.*, 2001), wurde die Interaktion von isoliertem ω mit isoliertem β‘ untersucht. Wie Messungen von 1D [^{1}H,^{15}N]-HSQC-Spektren von ^{15}N-ω vor und nach Zugabe von β‘ zeigten, interagieren die isolierten Untereinheiten nicht miteinander (Einzelarbeit C). Beim Zusammenbau der RNAP eluiert ω mit den anderen RNAP-Untereinheiten von der Ni^{2+}-Säule und ist in $RNAP^{aktiv}$ enthalten (Abschnitt 3.2.1, Einzelarbeit C). Daher kann ω vermutlich nur an ungefaltetes oder teilweise gefaltetes β‘ binden. Dies und die Sekundärstrukturanalyse von ω deuten darauf hin, dass ω seine finale Struktur während der Assemblierung der RNAP oder bei der Faltung von β‘ durch seine Bindung annimmt.

Die Sekundärstrukturanalyse der einzelnen RNAP-Untereinheiten zeigte, dass α, β und β‘ zwar gefaltet sind, jedoch ist kein Aktivitätstest für die isolierten Untereinheiten möglich. Daher wurde getestet, ob α, β, β‘ und ω in der Lage sind, mit Transkriptionsfaktoren zu interagieren, deren Bindestellen bekannt sind. NusG-NTD bindet an die β‘CH und den βGL und verbindet dadurch die

Scheren der RNAP, die die dsDNA umschließen (Abschnitt 1.1, Martinez-Rucobo *et al.*, 2011; Sevostyanova *et al.*, 2011). Dadurch wird die Freisetzung der DNA verhindert und die RNAP in einem effektiv transkribierenden Zustand gehalten (Artsimovitch und Landick, 2000; Burova *et al.*, 1995). Mit 1D [^{1}H,^{15}N]-HSQC-Spektren konnte die Bindung von NusG-NTD an isoliertes β und isoliertes β‘ nachgewiesen werden (Einzelarbeit C). Daher scheinen β und β‘ in Abwesenheit der anderen RNAP-Untereinheiten korrekt gefaltet zu sein, zumindest in den Regionen, die für die NusG-NTD-Bindung verantwortlich sind (vgl. Abb. 3-8). Mit α und ω konnte keine Interaktion beobachtet werden.

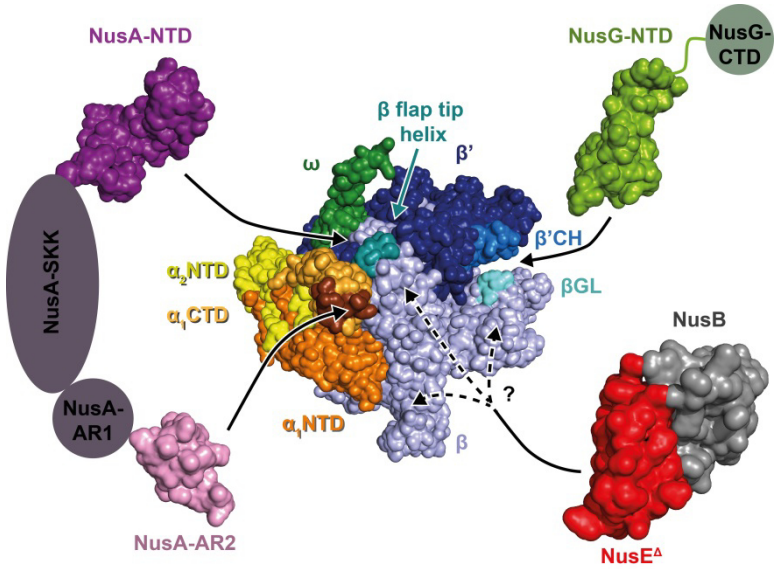


Abbildung 3-8: Bindestellen der Nus-Faktoren auf der RNAP. RNAP aus *E. coli* (PDB-Code: 4KMU) in der Oberflächendarstellung (α_1-NTD: orange, α_1-CTD hellorange mit der Bindestelle an AR2 in braun, α_2: gelb, β: hellblau, β‘: dunkelblau, ω: dunkelgrün. Die Bindestellen der Nus-Faktoren sind hervorgehoben und durch schwarze Pfeile gekennzeichnet. Die Nus-Faktoren sind in der Oberflächendarstellung, wobei Linkerregionen oder nicht untersuchte Domänen schematisch dargestellt sind. NusG: grün, NusA: lila, NusEΔ: rot, NusB: grau. Die exakte Bindestelle von NusEΔ auf β ist unklar.

Auch für NusA-NTD und NusA-AR2 ist bekannt, dass sie mit der RNAP interagieren. NusA-AR2 bindet an die α-CTD, wodurch NusA-AR2 von NusA-SKK abgelöst wird und die Autoinhibition von NusA aufgehoben wird (Mah *et al.*, 2000; Schweimer *et al.*, 2011). NusA-NTD kontaktiert die FTH der β-Untereinheit (Toulokhonov *et al.*, 2001; Yang *et al.*, 2009), wodurch Pausen der RNAP verlängert und neue Pausen eingeführt werden. Durch Signalabnahmen in 2D [^{1}H,^{15}N]-HSQC-Spektren konnte die Interaktion von NusA-AR2 mit der α-Untereinheit bestätigt werden (Abb. 3-8). Auch nach Zugabe von β kam es zu einem leichten Signalverlust im 1D [^{1}H,^{15}N]-HSQC-Spektrum, was aber auch bei Vorlage von ^{15}N-NusA-AR1 der Fall war. Somit ist diese Interaktion vermutlich unspezifisch und erfolgt auf Grund der negativen Ladung beider Domänen. Es konnte keine Wechselwirkung zwischen NusA-AR2 und β‘ bzw. ω nachgewiesen werden. Anhand von 1D [^{1}H,^{15}N]-HSQC-Spektren wurde gezeigt, dass NusA-NTD nur unvollständig an sowohl isoliertes β als auch isoliertes β‘ bindet, wobei die Signalabnahme nach Zugabe von β leicht stärker war. Der Signalverlust nach Zugabe des ββ‘-Komplexes dagegen war signifikant stärker, was aber durch Relaxationsexperimente auf die höhere molare Masse zurückgeführt werden konnte (Einzelarbeit C). Da für α und ω keine Interaktion beobachtet wurde, bestätigt

dies, dass NusA-NTD mit der β-Klappenregion interagiert (Abb. 3-8). Eine Beteiligung der β'-Untereinheit, die zusammen mit β den RNA-Austrittskanal bildet, kann aber nicht ausgeschlossen werden.
Zusammen mit den CD-Daten zeigen diese Bindungsexperimente, dass die isolierten RNAP-Untereinheiten korrekt gefaltet sind. Da alle Untereinheiten, besonders β und β', in der zusammengebauten RNAP große Berührungsflächen zueinander aufweisen, kann natürlich nicht ausgeschlossen werden, dass Teile der Untereinheiten, die nicht für die Bindung von NusG oder NusA verantwortlich sind, fehlgefaltet oder anders als im gesamten Enzym gefaltet sind. Diese Analyse ist eine schnelle Methode, um die mit einem Transkriptionsfaktor interagierende RNAP-Untereinheit zu ermitteln, solange der Transkriptionsfaktor eine Größe von ~50 kDa nicht überschreitet. Dies ist ein erster Schritt in der Bestimmung der genauen Bindestelle und eine gute Basis für weitere Analysen.
NusE ist Teil des Ribosoms und des Antiterminationskomplexes, in dem es zusammen mit NusB an die *BoxA*-RNA bindet (Luo *et al.*, 2008; Friedman *et al.*, 1981). Es wird über NusG an die RNAP gekoppelt (Burmann *et al.*, 2010), kann aber auch selbst an die RNAP binden (Mason und Greenblatt, 1991). Mit der gleichen Methode, mit der die NusG- und NusA-Interaktionen mit den RNAP-Untereinheiten analysiert wurden, wurde auch β als die bindende Untereinheit von NusE im NusEΔ:NusB-Komplex bestimmt (Abb. 3-8, Einzelarbeit C). Die Funktion dieser Bindung ist unklar, scheint jedoch eher für die Antitermination von Bedeutung zu sein als für die Transkriptions-Translationskopplung, da eine direkte Verbindung der beiden supramolekularen Komplexe RNAP und Ribosom sehr steif und unflexibel wäre und die Genexpression vermutlich hemmen würde.

3.2.3 NMR-Untersuchungen der RNAP

Um die RNAP mit NMR-Spektroskopie zu untersuchen, wurden [^{1}H,^{13}C]-TROSY *heteronuclear multiple quantum correlation* (HMQC)-Spektren von der deuterierten RNAP^{nativ} und von der isolierten, deuterierten β'-Untereinheit gemessen, bei denen die Methylgruppen von Ile, Leu und Val ^{1}H,^{13}C markiert waren (Einzelarbeit C). Das Spektrum der RNAP^{nativ} zeigte hohe Signaldispersion, jedoch überlagerten sehr viele Signale, was bei einem derart großen System (287 Val, 230 Ile, 349 Leu) zu erwarten war (Abb. 3-9A). Trotzdem konnten nach Spektrenüberlagerung viele Signale der α-CTD zugeordnet werden, die hochflexibel mit der α-NTD verbunden ist (Jeon *et al.*, 1997; Blatter *et al.*, 1994) und auch in der gesamten RNAP scharfe Signale mit hoher Intensität liefert.
Das Spektrum der isolierten β'-Untereinheit zeigte ebenfalls Signaldispersion, die jedoch nicht so groß war wie bei RNAP^{nativ}. Dies deutet wie die CD-Spektren (Abschnitt 3.2.2) auf korrekte Faltung von β' hin. Einige der Signale überlappten gut mit Signalen der RNAP^{nativ} und konnten folglich β' zugeordnet werden, während andere Signale nicht übereinstimmten (Abb. 3-9B, Einzelarbeit C). Diese stammen wahrscheinlich von Aminosäuren in Interaktionsflächen zu anderen RNAP-Untereinheiten oder erfahren allgemein eine andere chemische Umgebung als in der assemblierten RNAP. Durch

Zugabe von NusG-NTD zum methylgruppenmarkierten β‘ konnten signifikante Intensitätsabnahmen von Signalen in der Ile- (δ(^{13}C) = 9-16 ppm) und in der Leu/Val-Region (δ(^{13}C) = 17-29 ppm) des Spektrums beobachtet werden (Abb. 3-9C, Einzelarbeit C), die darauf hindeuten, dass diese Aminosäuren an der Bindung von NusG-NTD beteiligt sind. In den β‘CH, der Bindestelle von NusG-NTD, befinden sich zwei Ile und zwei Leu in der direkten Interaktionsfläche zu NusG-NTD, die zu zwei Signalen in der Ile-Region und vier Signalen in der Leu/Val-Region führen. Weitere Ile, Leu und Val-Aminosäuren sind in direkter Nähre lokalisiert und dadurch ebenfalls durch die Bindung von NusG-NTD beeinflusst (Abb. 3-8E). Daher ist davon auszugehen, dass bei der Titration von NusG-NTD zu β‘ deren Bindung beobachtet werden konnte und dass β‘ zumindest in dieser Region korrekt gefaltet ist.

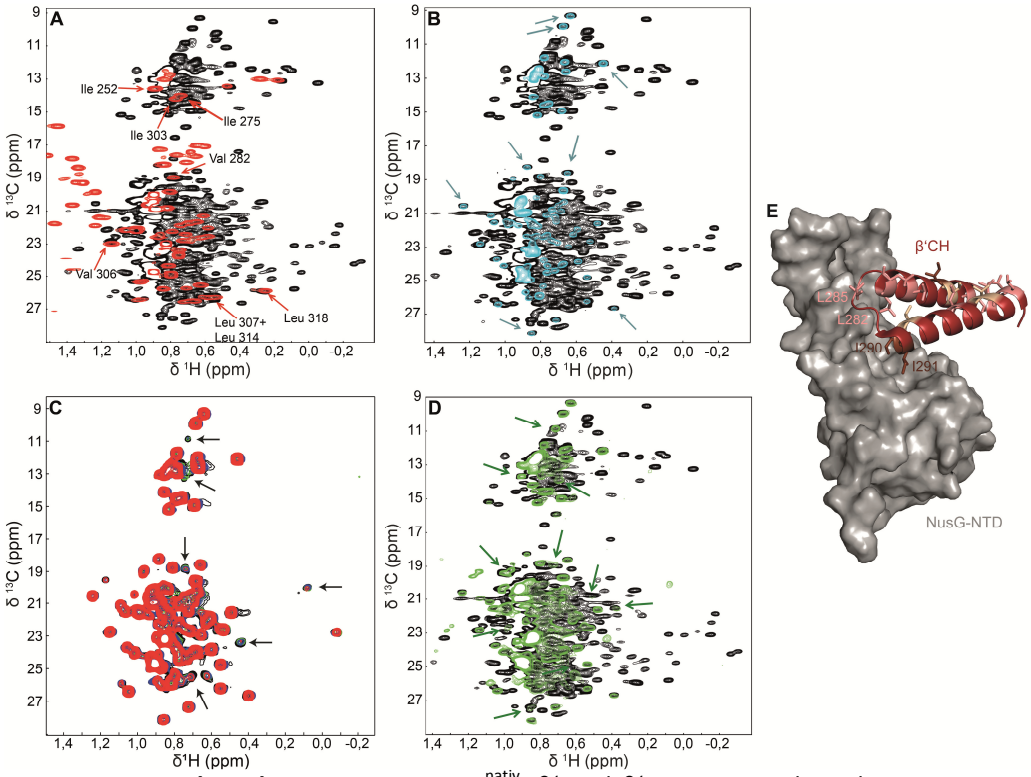


Abbildung 3-9: NMR-Analyse der RNAP. Bei RNAP^{nativ}, β‘ und β‘ in RNAP sind im deuterierten Protein die Methylgruppen von Ile, Leu und Val ^{1}H,^{13}C markiert. **(A)** Überlagerung des [^{1}H,^{13}C]-HMQC-Spektrums von RNAP^{nativ} (schwarz, 30 µM) und des [^{1}H,^{13}C]-HSQCs von ^{13}C,^{15}N α-CTD (rot, 700 µM). Direkt zugeordnete Signale sind gekennzeichnet. **(B)** Überlagerung der [^{1}H,^{13}C]-HMQC-Spektren von RNAP^{nativ} (schwarz) und β‘ (cyan, 2 µM). Einige Signale mit gleichen chemischen Verschiebungen sind gekennzeichnet. **(C)** Überlagerung der [^{1}H,^{13}C]-HMQC-Spektren von β‘ (2 µM) vor (schwarz) und nach Zugabe von NusG-NTD im molaren Verhältnis von 1:1 (grün), 1:2 (blau) und 1:10 (rot). Signale mit verminderter Intensität sind gekennzeichnet. **(D)** Überlagerung der [^{1}H,^{13}C]-HMQC-Spektrum von RNAP^{nativ} (schwarz) und β‘ in der RNAP (grün). Grüne Pfeile kennzeichnen Signale, die mit dem Spektrum von RNAP^{nativ} übereinstimmen, aber nicht mit isoliertem β‘. **(E)** Modell der Bindung von ecNusG-NTD (PDB-Code:2K06, grau) an die *E. coli* β’CH (PDB-Code: 4KMU, braun), erstellt durch Überlagerung mit der Kristallstruktur des *P. furisosus* Spt4/5 Komplexes gebunden an die Klammerdomäne der RNAP (Martinez-Rucobo *et al.*, 2011). Ile (dunkelbraun), Leu (rosa) und Val (beige) in den β’CH sind als Stäbchen dargestellt.

Um die Anzahl der Signale im Spektrum von RNAP^{nativ} zu reduzieren, wurde die RNAP aus deuterierten Untereinheiten assembliert, wobei die Methylgruppen von Ile, Leu und Val in β‘ ^{1}H,^{13}C markiert waren. Wieder war eine gute Signaldispersion erkennbar und mehr Signale überlappten mit

Signalen der methylgruppenmarkierten RNAP^{nativ} als beim Spektrum von isoliertem β‘, da β‘ nun in seinem physiologischen Kontext war (Abb. 3-9D, Einzelarbeit C).

Diese Daten zeigen einen ersten Versuch, die RNAP mittels NMR-Spektroskopie zu analysieren. Bisher untersuchte supramolekulare Systeme wie die Aspartat-Transcarbamoylase (300 kDa, Velyvis *et al.*, 2009), das „halbe“ Proteasom (360 kDa, Tugarinov *et al.*, 2007) oder das 20S Proteasom (670 kDa, Sprangers und Kay, 2007) haben vergleichbar hohe molare Massen wie die RNAP, sind aber Homooligomere, die mit NMR-Spektroskopie deutlich leichter zu untersuchen sind. Trotzdem war es möglich, vielversprechende [^{1}H,^{13}C]-Korrelationsspektren der Methylgruppen von Ile, Leu und Val der kompletten RNAP und der β‘-Untereinheit im Gesamtprotein zu erhalten. Dies ist die Basis weiterer Untersuchungen in Lösung und wird die Analyse von Intra- und Interdomänenbewegungen sowie transienten Interaktionen mit regulatorischen Faktoren erlauben, die mit anderen Methoden nur schwer zugänglich sind.

3.2.4 Bestimmung der Bindestellen verschiedener Transkriptionsfaktoren an die RNAP

Für die Regulation der Transkription sind Interaktionen der RNAP zu Transkriptionsfaktoren essentiell. Bisher gibt es jedoch nur wenig strukturelle Informationen über diese Bindungen. Daher wurde eine Methode entwickelt, um die RNAP-Bindestellen von Transkriptionsfaktoren zu bestimmen. Dazu wurden zunächst die Methylgruppen von Ile-, Leu- und Val-Resten im deuterierten Transkriptionsfaktor ^{1}H,^{13}C markiert. Anschließend erfolgte die Titration des methylgruppenmarkierten Transkriptionsfaktors mit protonierter RNAP mit Aufnahme von [^{1}H,^{13}C]-TROSY-Spektren nach jedem Titrationsschritt. Validiert wurde diese Methode mit NusG-NTD, deren Bindestelle an die RNAP bekannt ist. Die dabei beobachtete heterogene Intensitätsabnahme der Signale (Abb. 3-10A+B) lässt sich durch mehrere Effekte erklären. Zum einen ist ein allgemeiner Signalverlust auf die Komplexbildung mit der RNAP zurückzuführen, da die molare Masse drastisch erhöht wird ($MW_{NusG\ NTD}$ = 14 kDa; MW_{RNAP} = 390 kDa) und es dadurch zu schnellerer Relaxation und geringerer Magnetisierungstransfereffizienz kommt, was die Signale verbreitert. Zum anderen gelangen Methylgruppen des Transkriptionsfaktors, die sich in der RNAP-Bindestelle befinden, in die Nähe von Wasserstoffatomen der protonierten RNAP. Die dabei entstehenden intermolekularen Dipol-Dipol-Wechselwirkungen leisten einen zusätzlichen Beitrag zur Relaxation, wodurch die Signalintensität von Methylgruppen in der Bindungsfläche deutlich stärker abnimmt als die von anderen Methylgruppen. Außerdem kann die Signalintensität wie bei anderen Titrationen (vgl. Abschnitt 3.1.1) durch unvorteilhafte chemische Austauschraten beeinflusst werden. Zur Auswertung wurde zunächst der Mittelwert aller Restsignalintensitäten gebildet. Aminosäuren mit einer Restintensität von etwa 60 % dieses Mittelwerts wurden als stark betroffen angesehen, Aminosäuren mit einer Restintensität von etwa 80 % des Mittelwerts als schwach betroffen (Abb. 3-10B, Einzelarbeit D). Da mit dieser Methode nur Methylgruppen

untersucht werden können, aber wahrscheinlich Aminosäuren in unmittelbarer Umgebung ebenfalls von der Bindung betroffen sind, wurden zwei benachbarte Reste von betroffenen Methylgruppen zur Bindefläche dazu gezählt. Spt5 bindet mit Helix α3 und den Faltblattsträngen β1 und β3 an die β'CH der RNAP und mit dem N-terminalen Bereich von Helix α2 an den βGL (Martinez-Rucobo *et al.*, 2011; Sevostyanova *et al.*, 2011). Die NMR-Analyse zeigt klar zwei betroffene Regionen auf der NusG-NTD. Die erste entspricht der Bindestelle an die β'CH und umfasst Helix α3 und die Faltblattstränge β1 und β3, die zweite wird von Helices α1 und α2 geformt und liegt in der Nähe des RfaH-Motifs, das für die Bindung an den βGL identifiziert wurde (Abb. 3-10C+D, Einzelarbeit D, Sevostyanova *et al.*, 2011). Im Bindungsmodell von NusG-NTD an die RNAP, erstellt durch Überlagerung mit der Kristallstruktur von *P. furiosus* Spt4/5 gebunden an die Klammerdomäne der RNAP (Martinez-Rucobo *et al.*, 2011), liegt die erste betroffene Region klar an der Grenzfläche zu den β'CH, während die zweite betroffene Region in Kontakt zum βGL steht. (Abb. 3-10E). Diese entspricht nicht exakt der von Sevostyanova *et al.* bestimmten Bindestelle, was darauf zurückzuführen ist, dass in NusG-NTD keine Ile/Leu/Val-Reste in der entsprechenden Region zu finden sind. Daher ist in den NMR-Experimenten vermutlich die benachbarte Region betroffen.

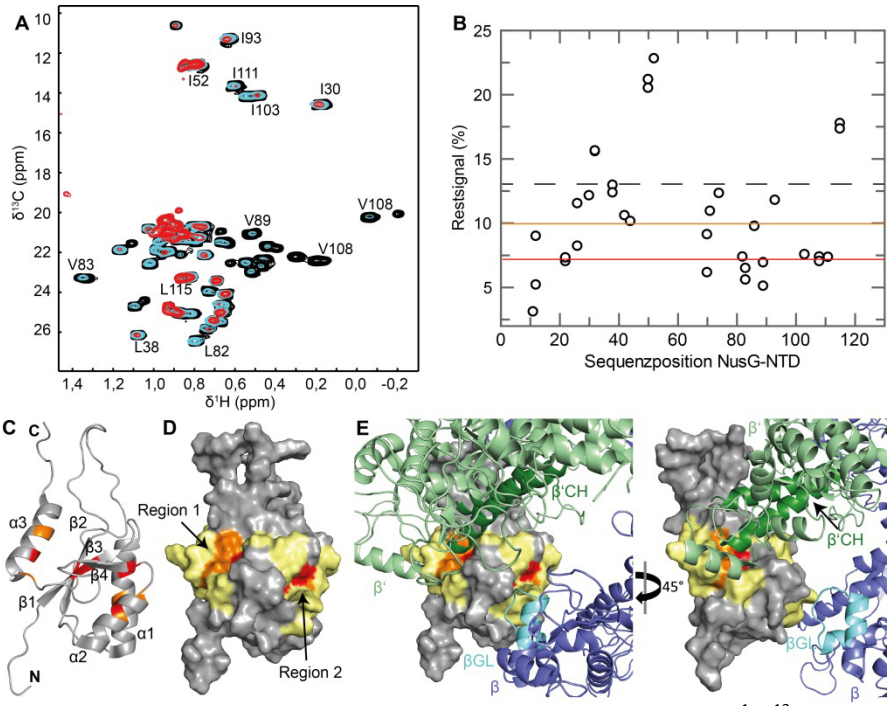


Abbildung 3-10: Bindung von NusG-NTD an die RNAP. (A) Überlagerung der [^{1}H,^{13}C]-TROSY-Spektren von methylgruppenmarkierter NusG-NTD (schwarz) und nach Zugabe von RNAP im molaren Verhältnis von 1:1 (cyan) und 1:2 (rot). Ausgewählte Signale sind beschriftet. **(B)** Prozentuale Restsignalintensität der NusG-NTD-Methylgruppen nach der Zugabe von RNAP im 1:1-Verhältnis. Die schwarze gestrichelte Linie zeigt den Mittelwert der Restsignalintensitäten, während die roten und orangen Linien die Grenzen von stark betroffenen (55 % des Mittelwerts der Restintensitäten) und schwach betroffenen (75 % des Mittelwerts der Restintensitäten) Signalen darstellen. **(C)** *Cartoon*-Darstellung der NusG-NTD (PDB-Code: 2K06). Durch die Bindung der RNAP stark betroffene Aminosäuren sind rot, schwach betroffene orange gekennzeichnet. **(D)** Oberflächendarstellung von NusG-NTD wie in (C), zwei Aminosäuren auf jeder Seite eines betroffenen Rests sind gelb hervorgehoben, es sei denn es waren nicht betroffene Ile/Leu/Val-Reste. **(E)** Bindung von NusG-NTD aus (D) an die ecRNAP (PDB-Code: 4KMU), erstellt durch Überlagerung mit der Kristallstruktur von *P. furiosus* Spt4/5 gebunden an die Klammerdomäne der RNAP (PDB-Code: 3QQC). Blau: β, hellgrün: β', dunkelgrün: β'CH, cyan: βGL.

Während des Transkriptionszyklus erfahren die β'CH durch das Öffnen und Schließen der Krebsscheren um die dsDNA strukturelle Änderungen (Cramer *et al.*, 2001; Darst *et al.*, 2002; Landick, 2001). NusG arretiert die β'CH im geschlossenen Zustand, indem es an die β'CH und den βGL bindet. Dies stabilisiert den Elongationskomplex und unterdrückt Pausen (Artsimovitch und Landick, 2000; Burova *et al.*, 1995). Die NMR-Experimente aus Einzelarbeit C deuten an, dass NusG auch in Abwesenheit von DNA an die β'CH und den βGL binden kann. Somit sind die Krebsscheren entweder in einer Konformation, die diese Bindung erlaubt oder NusG ist in der Lage, deren geschlossenen Zustand zu induzieren. Alles in allem konnte die Interaktionsflächen der NusG-NTD an die RNAP bestätigt werden, was bedeutet, dass die Methode der Bindestellenbestimmung mittels NMR-Spektroskopie funktioniert. Daraufhin konnte sie auf weitere Transkriptionsfaktoren angewendet werden.

Der Transkriptionsfaktor NusE/S10 bindet während der Antitermination zusammen mit NusB die RNA-Sequenz *BoxA* (Luo *et al.*, 2008; Mason *et al.*, 1992a). Dabei wird es, wie auch als Teil des Ribosoms, über NusG an der RNAP verankert (Burmann *et al.*, 2010). NusE kann jedoch auch alleine an die β-Untereinheit der RNAP binden (Einzelarbeit C, Mason und Greenblatt, 1991). Mittels der oben genannten Methode wurde gezeigt, dass die RNAP-Bindestelle auf NusE vor allem Helix α2 und Faltblattstränge β1 und β4, aber auch Helix α1 umfasst (Einzelarbeit D, Abb. 3-11A).

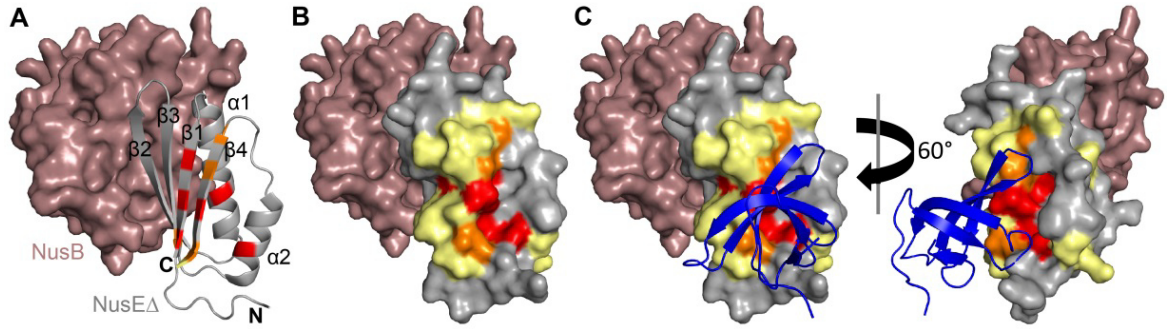


Abbildung 3-11: Bindestelle von NusE an die RNAP. (A) *Cartoon*-Darstellung von NusEΔ (hellgrau) im NusEΔ:NusB-Komplex (NusB: Oberflächendarstellung, dunkelrosa; PDB-Code: 3D3B). Durch die RNAP-Bindung stark betroffene Aminosäuren sind rot, schwach betroffene orange hervorgehoben. **(B)** Zwei benachbarte Aminosäuren (außer nicht betroffene Ile/Leu/Val-Reste) sind gelb hervorgehoben, wodurch eine zusammenhängende Bindestelle entsteht. **(C)** Die Überlagerung mit dem NusEΔ:NusG-CTD-Komplex (PDB-Code: 2KVQ, NusG-CTD: blau) zeigt die Überlappung der Bindestellen von NusG-CTD und RNAP.

Die Bindestelle (Abb. 3-11B) liegt gegenüber der Bindefläche von NusB und sollte auch bei in das Ribosom integriertem NusE noch zugänglich sein (Schuwirth *et al.*, 2005). Die direkte Bindung des Ribosoms an die RNAP ist jedoch höchst unwahrscheinlich, da der resultierende Komplex sehr steif und unflexibel wäre und vermutlich die Genexpression beeinträchtigen würde. Interessanterweise überlappt die RNAP-Bindestelle auf NusE stark mit der Bindestelle von NusE an NusG-CTD (Burmann *et al.*, 2010, Abb. 3-11C). Wie NMR-Verdrängungsexperimente zeigten, schließen sich die Bindungen von NusE an die RNAP und NusG-CTD gegenseitig aus und beide Proteine können durch einen Überschuss des anderen Proteins von NusEΔ verdrängt werden (Einzelarbeit D). Beide Komplexe, d.h. NusE:RNAP bzw. NusE:NusG:RNAP, scheinen in einem Gleichgewicht zueinander zu stehen, das

möglicherweise durch andere Regulatoren wie Transkriptionsfaktoren oder gewisse RNA-Sequenzen wie *BoxA* beeinflusst werden kann. Dies zeigt, dass NusG-CTD auch in der Gegenwart der RNAP mit NusE interagieren kann, was die Transkriptions-Translations-Kopplung durch NusG ermöglicht. Die direkte Bindung von NusE an die RNAP wird eher in der Antitermination der Transkription von Bedeutung sein. Möglicherweise ist die Bindung von NusE oder vom NusE:NusB-Komplex ein frühes Ereignis der Antitermination. Da die geschätzte Menge an freiem NusE in der Zelle sehr gering ist (Greive *et al.*, 2005), würde die direkte Bindung der RNAP die lokale NusE-Konzentration und die Effizienz der Assemblierung des Antiterminationskomplexes erhöhen. Die hohe Dichte von transkribierenden RNAPs in *rrn* Operons (Gotta *et al.*, 1991) mit hohen Elongationsraten (Vogel und Jensen, 1994) lässt die Bindung von NusE an die RNAP als gute Lösung erscheinen, um effiziente Antitermination in diesen Operons zu ermöglichen.

Die Bindestelle der NTD von NusA konnte durch Vernetzungsexperimente und eine elektronenmikroskopische Komplexstruktur mit schwacher Auflösung am RNA-Austrittskanal nahe der FTH lokalisiert werden (Toulokhonov *et al.*, 2001; Yang *et al.*, 2009), wobei jedoch die betroffenen Aminosäuren sowie die Orientierung der NusA-NTD unklar blieben. Mittels NMR-Titrationen analog zu NusG und NusE wurde eine durchgehende Bindefläche an die RNAP auf einer Seite von NusA-NTD bestimmt (Einzelarbeit D). Die meisten durch die Bindung der RNAP betroffenen Aminosäuren sind im globulären Kopf lokalisiert, aber auch in Helix α4 und der Schleife zwischen α2 und β1 (Einzelarbeit D). Anhand der betroffenen Reste wurde NusA-NTD durch eine HADDOCK-Simulation an den TEC von *T. thermophilus* modelliert (Abb. 3-12A, Einzelarbeit D). Im Gegensatz zu anderen Analysen bezüglich der Orientierung von NusA-NTD (Yang *et al.*, 2009; Ha *et al.*, 2010) gibt die mit NMR-Spektroskopie bestimmte Bindestelle ein deutlich detaillierteres Bild der Bindung von NusA-NTD an die RNAP. Die globale Orientierung von NusA-NTD entspricht der von Ha 2010 (Abb. 3-12B), jedoch ist die NusA-NTD gedreht, sodass die konkave Seite der NusA-NTD an die FTH bindet. Der negativ geladene Kopf von NusA-NTD kontaktiert dabei eine positiv geladene Oberfläche von β‘. Schon in Einzelarbeit C wurde festgestellt, dass auch die β‘-Untereinheit der RNAP von der NusA-NTD-Bindung betroffen sein könnte. Dabei moduliert NusA-NTD möglicherweise den RNA-Austrittskanal und drückt die RNA in Richtung von NusA-SKK (Yang *et al.*, 2009; Yang und Lewis, 2010). Diese ist so positioniert, dass sie austretende RNA direkt binden kann (Abb. 3-12C), die basierend auf RNA-Bindungsstudien (Schweimer *et al.*, 2011; Beuth *et al.*, 2005) im Modell um die NusA-SKK gewunden wird (Abb. 3-12C). Um Kollisionen von NusA-SKK mit der RNAP zu verhindern, muss die Verbindung zwischen NusA-NTD und NusA-SKK flexibel sein. Dies kann zum einen durch Rotation des 3-4 Aminosäuren umfassenden Linkers zwischen NusA-NTD und NusA-SKK oder zum anderen durch eine Reorientierung der Linkerhelix in NusA-NTD erreicht werden.

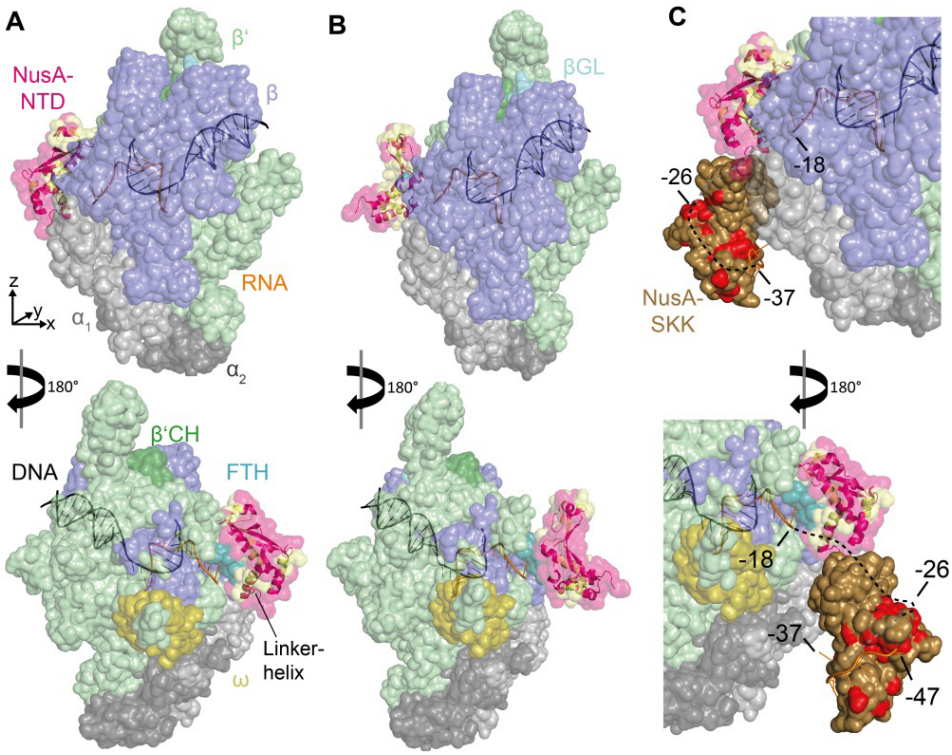


Abbildung 3-12: Modell der NusA-Bindung an den ttTEC. (A) NusA-NTD (pink, durch die RNAP-Bindung betroffenen Aminosäuren: gelb) in der *Cartoon*- und Oberflächendarstellung wurde, basierend auf der durch NMR-Experimente bestimmten RNAP-Bindestelle, durch eine HADDOCK-Simulation an die β-Untereinheit des ttTEC (PDB-Code: 2O5I; grau: α_1 und α_2, blau: β, türkis: FTH, cyan: βGL, hellgrün: β', dunkelgrün: β'CH, ocker: ω, schwarz: DNA, orange: RNA) modelliert. **(B)** Modell der ecNusA-NTD-Bindung an die RNAP nach Ha *et al.*, 2010. **(C)** Bindung der austretenden RNA durch NusA. ecNusA-NTD ist orientiert wie in (A). NusA-SKK aus *T. maritima* (PDB-Code: 1L2F, braun) ist in der Oberflächendarstellung, während durch die RNA-Bindung betroffene Reste nach Schweimer *et al.*, 2011 rot markiert sind. Die RNA der Kristallstruktur von NusA-SKK aus *Mycobacterium tuberculosis* ist in der *Cartoon*-Darstellung, während die schwarze und graue gestrichelte Linie einen möglichen Weg der austretenden RNA darstellt. Die abgeschätzten Basennummern sind angegeben.

Um die Regulation der Transkription zu verstehen, sind strukturelle Informationen über die Interaktionen von Transkriptionsfaktoren und RNAP sehr wichtig. Insgesamt stellt das Konzept mit einfachen NMR-Experimenten die Bindestelle der RNAP auf methylgruppenmarkierten Transkriptionsfaktoren zu bestimmen, einen sehr überzeugender Ansatz dar, um Wechselwirkungen zwischen Transkriptionsfaktor und RNAP zu untersuchen. Die Leistung dieser Methode zeigt sich darin, dass sogar transiente Interaktionen analysiert werden können und sie erweiterbar auf andere Teile des Transkriptions-/Translationsapparates ist. Darüber hinaus kann sie leicht auf andere Systeme übertragen werden, in denen kleine Proteine mit supramolekularen Komplexen wechselwirken, und kann auch dort wichtige Informationen über Bindungen liefern.

4 Abkürzungsverzeichnis

aa	*Aquifex aeolicus*
α-CTD	C-terminale Domäne der α-Untereinheit der RNAP
AR	*acidic repeat*, Domäne von NusA
ATP	Adenosintriphosphat
β'CH	β' *clamp helices*, Helices der Klammer der β'-Untereinheit
βGL	β *gate loop*, Schleife der β-Schranke
BH	Brückenhelix
bp	Basenpaare
CD	Circulardichroismus
CTD	C-terminale Domäne
ctr	*c-terminal repeat*, Bestandteil von NusG-homologem Stp5 aus Eukaryonten
DNA	Desoxyribonukleinsäure
dNTP	Desoxyribonukleosidtriphosphat
ds	doppelsträngig
ec	*Escherichia coli*
6-FAM	6-Carboxyfluorescein, Fluoreszenzfarbstoff
FTH	*flap tip helix*, flexible Helix der β-Klammerdomäne
H/D-Austausch	Wasserstoff/Deuterium Austausch
HMQC	*heteronuclear multiple quantum coherence*
HSQC	*heteronuclear single quantum coherence*
KH	K homolog, Domäne von NusA
KOW	Kyrpides-Ouzounis-Woese-Motiv, Rekrutierngsplattform von NusG für akzessorische Proteine
LUCA	*last universal common ancestor*, der letzte universelle gemeinsame Vorfahre von Bakterien, Archaeen und Eukaryonten
Mg1	fest von der katalytischen Aspartattriade von β' gebundenes Magnesiumion
Mg2	von β' lose gebundenes Magnesiumion, gelangt mit NTP zum aktiven Zentrum
mRNA	*messenger* RNA
NGN	NusG *N-terminal domain*-Motiv
NMR	*nuclear magnetic resonance*
NOE	*Nulcear Overhauser Effect*
NOESY	*Nuclear Overhauser Effect Spectroscopy*
nt	Nukleotide

NTD	N-terminale Domäne
NTP	Nukleosidtriphosphat
Nus	N *utilisation substance*
NusEΔ	die Ribosomenbindungsschleife von NusE wurde durch ein Serin ersetzt
nut	N *utilisation site*
ops	*operon polarity suppressor*, Bindesequenz auf der DNA für RfaH-NTD
PP_i	Pyrophosphat
rmsd	*root mean square deviation*, Standardabweichung
RNA	Ribonukleinsäure
rRNA	ribosomale RNA
RNAP	RNA-Polymerase
$RNAP^{aktiv}$	aktive Fraktion der aus einzelnen Untereinheiten assemblierten RNAP
$RNAP^{inaktiv}$	inaktive Fraktion der aus einzelnen Untereinheiten assemblierten RNAP
$RNAP^{nativ}$	von einem Plasmid überexprimierte und in *E.* coli assemblierte RNAP
rrn Operons	ribosomale RNA Operons
rut	Rho *utilization site*
SEC	*size exclusion chromatography*, Gelfiltration
SKK	RNA-bindende Domänen von NusA: S1, KH1 und KH2
ss	*single stranded*, einzelsträngig
TEC	Transkriptions-Elongations-Komplex
TH	Trigger-Helices
tm	*thermotoga maritima*
tmNusGΔ	DII von tmNusG wurde durch die an dieser Stelle in ecNusG vorkommende Schleife ersetzt
tRNA	Transfer-RNA
TROSY	*transverse relaxation optimized spectroscopy*
TS	Trigger-Schleife
tt	*Thermus thermophilus*

5 Literaturverzeichnis

Aksoy S, Squires CL und Squires C (1984) Evidence for antitermination in *Escherichia coli* RRNA transcription. J Bacteriol **159:** 260-4

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

Artsimovitch I, Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci U S A **97:** 7090-7095

Bai Y, Milne JS, Mayne L und Englander SW (1993) Primary structure effects on peptide group hydrogen exchange. Proteins **17:** 75-86

Baker NA, Sept D, Joseph S, Holst MJ und McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A **98:** 10037-10041

Belogurov GA, Mooney RA, Svetlov V, Landick R und Artsimovitch I (2009) Functional specialization of transcription elongation factors. EMBO J **28:** 112-122

Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin NV, Vassylyev DG und Artsimovitch I (2007) Structural Basis for Converting a General Transcription Factor into an Operon-Specific Virulence Regulator. Mol Cell **26:** 117-129

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

Berger A, Linderstrom-Lang K (1957) Deuterium exchange of poly-DL-alanine in aqueous solution. Arch Biochem Biophys **69:** 106-118

Beuth B, Pennell S, Arnvig KB, Martin SR und Taylor IA (2005) Structure of a *Mycobacterium tuberculosis* NusA-RNA complex. EMBO J **24:** 3576-3587

Blatter EE, Ross W, Tang H, Gourse RL und Ebright RH (1994) Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell **78:** 889-896

Bogden CE, Fass D, Bergman N, Nichols MD und Berger JM (1999) The Structural Basis for Terminator Recognition by the Rho Transcription Termination Factor. Mol Cell **3:** 487-493

Bonin I, Muhlberger R, Bourenkov GP, Huber R, Bacher A, Richter G und Wahl MC (2004a) Structural basis for the interaction of *Escherichia coli* NusA with protein N of phage lambda. Proc Natl Acad Sci U S A **101:** 13762-13767

Bonin I, Robelek R, Benecke H, Urlaub H, Bacher A, Richter G und Wahl MC (2004b) Crystal structures of the antitermination factor NusB from *Thermotoga maritima* and implications for RNA binding. Biochem J **383:** 419-428

Borukhov S, Goldfarb A (1993) Recombinant *Escherichia coli* RNA polymerase: purification of individually overexpressed subunits and in vitro assembly. Protein Expr Purif **4:** 503-511

Borukhov S, Lee J und Laptenko O (2005) Bacterial transcription elongation factors: new insights into molecular mechanism of action. Mol Microbiol **55:** 1315-1324

Borukhov S, Severinov K (2002) Role of the RNA polymerase sigma subunit in transcription initiation. Res Microbiol **153:** 557-562

Brendel V, Hamm GH und Trifonov EN (1986) Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J Biomol Struct Dyn **3:** 705-723

Brennan CA, Dombroski AJ und Platt T (1987) Transcription termination factor rho is an RNA-DNA helicase. Cell **48:** 945-952

Briercheck DM, Wood TC, Allison TJ, Richardson JP und Rule GS (1998) The NMR structure of the RNA binding domain of *E. coli* rho factor suggests possible RNA-protein interactions. Nat Struct Biol **5:** 393-399

Buck M, Gallegos MT, Studholme DJ, Guo Y und Gralla JD (2000) The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. J Bacteriol **182:** 4129-4136

Burgess RR (1969) Separation and characterization of the subunits of ribonucleic acid polymerase. J Biol Chem **244:** 6168-6176

Burgess RR, Travers AA, Dunn JJ und Bautz EK (1969) Factor stimulating transcription by RNA polymerase. Nature **221:** 43-46

Burmann BM, Knauer SH, Sevostyana A, Schweimer K, Mooney RA, Landick R, Artsimovitch I und Rösch P (2012) An α-helix to β-barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell **150:** 291-303

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

Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME und Rösch P (2010) A NusE:NusG Complex Links Transcription and Translation. Science **328:** 501-504

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

Carlomagno MS, Nappo A (2003) NusA modulates intragenic termination by different pathways. Gene **308:** 115-28

Condon C, French S, Squires C und Squires CL (1993) Depletion of functional ribosomal RNA operons in *Escherichia coli* causes increased expression of the remaining intact copies. EMBO J **12:** 4305-4315

Condon C, Squires C und Squires CL (1995) Control of rRNA transcription in *Escherichia coli*. Microbiol Rev **59:** 623-645

Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, Jawhari A, Jennebach S, Kamenski T, Kettenberger H, Kuhn CD, Lehmann E, Leike K, Sydow JF und Vannini A (2008) Structure of eukaryotic RNA polymerases. Annu Rev Biophys **37:** 337-352

Cramer P, Bushnell DA und Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science **292:** 1863-1876

Dalal RV, Larson MH, Neuman KC, Gelles J, Landick R und Block SM (2006) Pulling on the nascent RNA during transcription does not alter kinetics of elongation or ubiquitous pausing. Mol Cell **23:** 231-239

Darst SA, Opalka N, Chacon P, Polyakov A, Richter C, Zhang G und Wriggers W (2002) Conformational flexibility of bacterial RNA polymerase. Proc Natl Acad Sci U S A **99:** 4296-4301

Das A, Court D und Adhya S (1976) Isolation and characterization of conditional lethal mutants of *Escherichia coli* defective in transcription termination factor rho. Proc Natl Acad Sci U S A **73:** 1959-1963

Daube SS, von Hippel PH (1999) Interactions of *Escherichia coli* sigma(70) within the transcription elongation complex. Proc Natl Acad Sci U S A **96:** 8390-8395

Davis CA, Bingman CA, Landick R, Record MT,Jr und Saecker RM (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 U S A **104:** 7833-7838

Dove SL, Hochschild A (1998) Conversion of the omega subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. Genes Dev **12:** 745-754

Eisenmann A, Schwarz S, Prasch S, Schweimer K und Rösch P (2005) The *E. coli* NusA carboxy-terminal domains are structurally similar and show specific RNAP- and λ N interaction. Protein Sci **14:** 2018-2029

Englander SW, Sosnick TR, Englander JJ und Mayne L (1996) Mechanisms and uses of hydrogen exchange. Curr Opin Struct Biol **6:** 18-23

Epshtein V, Cardinale CJ, Ruckenstein AE, Borukhov S und Nudler E (2007) An allosteric path to transcription termination. Mol Cell **28:** 991-1001

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

Friedman DI, Olson ER (1983) Evidence that a nucleotide sequence, "boxA," is involved in the action of the NusA protein. Cell **34:** 143-149

Friedman DI, Olson ER, Georgopoulos C, Tilly K, Herskowitz I und Banuett F (1984) Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. Microbiol Rev **48:** 299-325

Friedman DI, Schauer AT, Baumann MR, Baron LS und Adhya SL (1981) Evidence that ribosomal protein S10 participates in control of transcription termination. Proc Natl Acad Sci U S A **78:** 1115-1118

Gardner KH, Kay LE (1998) The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct **27:** 357-406

Gardner KH, Rosen MK und Kay LE (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. Biochemistry (N Y) **36:** 1389-1401

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD und Bairoch A (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res **31:** 3784-3788

Geiselmann J, Wang Y, Seifried SE und von Hippel PH (1993) A physical model for the translocation and helicase activities of *Escherichia coli* transcription termination protein Rho. Proc Natl Acad Sci U S A **90:** 7754-7758

Gentry D, Xiao H, Burgess R und Cashel M (1991) The omega subunit of *Escherichia coli* K-12 RNA polymerase is not required for stringent RNA control in vivo. J Bacteriol **173:** 3901-3903

Gentry DR, Burgess RR (1993) Cross-linking of *Escherichia coli* RNA polymerase subunits: identification of beta' as the binding site of omega. Biochemistry **32:** 11224-11227

Ghosh P, Ishihama A und Chatterji D (2001) *Escherichia coli* RNA polymerase subunit omega and its N-terminal domain bind full-length beta' to facilitate incorporation into the $alpha_2$beta subassembly. Eur J Biochem **268:** 4621-4627

Goldman SR, Ebright RH und Nickels BE (2009) Direct Detection of Abortive RNA Transcripts *in Vivo*. Science **324:** 927-928

Gopal B, Haire LF, Cox RA, Jo Colston M, Major S, Brannigan JA, Smerdon SJ und Dodson G (2000) The crystal structure of NusB from *Mycobacterium tuberculosis*. Nat Struct Biol **7:** 475-8

Goto NK, Gardner KH, Mueller GA, Willis RC und Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. J Biomol NMR **13:** 369-374

Gotta SL, Miller OL,Jr und French SL (1991) rRNA transcription rate in *Escherichia coli*. J Bacteriol **173:** 6647-6649

Gottesman ME, Adhya S und Das A (1980) Transcription antitermination by bacteriophage lambda N gene product. J Mol Biol **140:** 57-75

Gottesman ME, Weisberg RA (2004) Little lambda, who made thee? Microbiol Mol Biol Rev **68:** 796-813

Greenfield N, Fasman GD (1969) Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry **8:** 4108-4116

Greive SJ, Lins AF und von Hippel PH (2005) Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to boxA RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in *Escherichia coli*. J Biol Chem **280:** 36397-36408

Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol **57:** 441-466

Gunnelius L, Hakkila K, Kurkela J, Wada H, Tyystjarvi E und Tyystjarvi T (2014) The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. Nucleic Acids Res **42:** 4606-4614

Guo M, Xu F, Yamada J, Egelhofer T, Gao Y, Hartzog GA, Teng M und Niu L (2008) Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. Structure **16:** 1649-1658

Ha KS, Toulokhonov I, Vassylyev DG und Landick R (2010) The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. J Mol Biol

Heil A, Zillig W (1970) Reconstitution of bacterial DNA-dependent RNA-polymerase from isolated subunits as a tool for the elucidation of the role of the subunits in transcription. FEBS Lett **11:** 165-168

Heinrich T, Condon C, Pfeiffer T und Hartmann RK (1995) Point mutations in the leader boxA of a plasmid-encoded *Escherichia coli* rrnB operon cause defective antitermination in vivo. J Bacteriol **177:** 3793-3800

Hirtreiter A, Damsma GE, Cheung ACM, Klose D, Grohmann D, Vojnic E, Martin ACR, Cramer P und Werner F (2010) Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. Nucleic Acids Research **38:** 4040-4051

Horstmann M, Ehses P, Schweimer K, Steinert M, Kamphausen T, Fischer G, Hacker J, Rösch P und Faber C (2006) Domain Motions of the Mip Protein from *Legionella pneumophila*. Biochemistry **45:** 12303-12311

Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, König H, Reinhard R, Rockinger I, Fricke H und Stetter KO (1992) A*quifex pyrophilus* gen. nov. sp. nov., Represents a Novel Group of Marine Hyperthermophilic Hydrogen-Oxidizing Bacteria. Systematic and Applied Microbiology **15 (3):** 340-351

Huber R, Langworthy TA, König H, Thomm M, Wöse CR, Sleytr UB und Stetter KO (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic euhacteria growing up to 90° C. Arch Microbiol **144:** 324-333

Huenges M, Rolz C, Gschwind R, Peteranderl R, Berglechner F, Richter G, Bacher A, Kessler H und Gemmecker G (1998) Solution structure of the antitermination protein NusB of *Escherichia coli*: a novel all-helical fold for an RNA-binding protein. EMBO J **17:** 4092-4100

Huyghues-Despointes BM, Scholtz JM und Pace CN (1999) Protein conformational stabilities can be determined from hydrogen exchange rates. Nat Struct Biol **6:** 910-912

Igarashi K, Ishihama A (1991) Bipartite functional map of the *E. coli* RNA polymerase alpha subunit: involvement of the C-terminal region in transcription activation by cAMP-CRP. Cell **65:** 1015-1022

Ishiguro A, Nogi Y, Hisatake K, Muramatsu M und Ishihama A (2000) The Rpb6 subunit of fission yeast RNA polymerase II is a contact target of the transcription elongation factor TFIIS. Mol Cell Biol **20:** 1263-1270

Ito K, Iwakura Y und Ishihama A (1975) Biosynthesis of RNA polymerase in *Escherichia coli*. III. Identification of intermediates in the assembly of RNA polymerase. J Mol Biol **96:** 257-271

Janin J, Miller S und Chothia C (1988) Surface, subunit interfaces and interior of oligomeric proteins. J Mol Biol **204:** 155-164

Jeffery CJ (1999) Moonlighting proteins. Trends Biochem Sci **24:** 8-11

Jeon YH, Yamazaki T, Otomo T, Ishihama A und Kyogoku Y (1997) Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain. J Mol Biol **267:** 953-962

Jiang Y, Zhou Q, Wu K, Li XQ und Shao WL (2006) A highly efficient method for liquid and solid cultivation of the anaerobic hyperthermophilic eubacterium *Thermotoga maritima*. FEMS Microbiol Lett **259:** 254-259

Jin DJ, Burgess RR, Richardson JP und Gross CA (1992) Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. Proc Natl Acad Sci U S A **89:** 1453-1457

Katayama A, Fujita N und Ishihama A (2000) Mapping of subunit-subunit contact surfaces on the beta' subunit of *Escherichia coli* RNA polymerase. J Biol Chem **275:** 3583-3592

Kim D, Patel SS (2001) The Kinetic Pathway of RNA Binding to the *Escherichia coli* Transcription Termination Factor Rho. J Biol Chem **276:** 13902-13910

Kim TK, Ebright RH und Reinberg D (2000) Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science **288:** 1418-1422

Klumpp S, Hwa T (2008) Stochasticity and traffic jams in the transcription of ribosomal RNA: Intriguing role of termination and antitermination. Proceedings of the National Academy of Sciences **105:** 18159-18164

Knowlton JR, Bubunenko M, Andrykovitch M, Guo W, Routzahn KM, Waugh DS, Court DL und 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

Ko DC, Marr MT, Guo J und Roberts JW (1998) A surface of *Escherichia coli* sigma 70 required for promoter function and antitermination by phage lambda Q protein. Genes Dev **12:** 3276-85

Kontur WS, Capp MW, Gries TJ, Saecker RM und Record MT,Jr (2010) Probing DNA binding, DNA opening, and assembly of a downstream clamp/jaw in *Escherichia coli* RNA polymerase-lambdaP(R) promoter complexes using salt and the physiological anion glutamate. Biochemistry **49:** 4361-4373

Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A und Darst SA (2000) A structural model of transcription elongation. Science **289:** 619-625

Koslover DJ, Fazal FM, Mooney RA, Landick R und Block SM (2012) Binding and translocation of termination factor rho studied at the single-molecule level. J Mol Biol **423:** 664-676

Kyrpides NC, Woese CR und Ouzounis CA (1996) KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. Trends Biochem Sci **21:** 425-426

Landick R (2006) The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans **34:** 1062-1066

Landick R (2001) RNA polymerase clamps down. Cell **105:** 567-570

Larson MH, Greenleaf WJ, Landick R und Block SM (2008) Applied Force Reveals Mechanistic and Energetic Details of Transcription Termination. Cell **132:** 971-982

Lazinski D, Grzadzielska E und Das A (1989) Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. Cell, **59:** 207-218

Lee S, Nguyen HM und Kang C (2010) Tiny abortive initiation transcripts exert antitermination activity on an RNA hairpin-dependent intrinsic terminator. Nucleic Acids Res **38:** 6045-6053

LeMaster DM, Richards FM (1988) NMR sequential assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteriation. Biochemistry **27:** 142-150

Li J, Mason SW und Greenblatt J (1993) Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. Genes Dev **7:** 161-172

Li SC, Squires CL und Squires C (1984) Antitermination of *E. coli* rRNA transcription is caused by a control region segment containing lambda nut-like sequences. Cell **38:** 851-860

Liao D, Lurz R, Dobrinski B und Dennis PP (1996) A NusG-like protein from *Thermotoga maritima* binds to DNA and RNA. J Bacteriol **178:** 4089-4098

Lindstrom DL, Squazzo SL, Muster N, Burckin TA, Wachter KC, Emigh CA, McCleery JA, Yates JR,3rd und Hartzog GA (2003) Dual roles for Spt5 in pre-mRNA processing and transcription elongation revealed by identification of Spt5-associated proteins. Mol Cell Biol **23:** 1368-1378

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

Liu X, Bushnell DA, Silva DA, Huang X und Kornberg RD (2011) Initiation complex structure and promoter proofreading. Science **333:** 633-637

Lonetto M, Gribskov M und Gross CA (1992) The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol **174:** 3843-3849

Luo X, Hsiao HH, Bubunenko M, Weber G, Court DL, Gottesman ME, Urlaub H und Wahl MC (2008) Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. Mol Cell **32:** 791-802

Magill CP, Jackson SP und Bell SD (2001) Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. J Biol Chem **276:** 46693-46696

Mah TF, Kuznedelov K, Mushegian A, Severinov K und Greenblatt J (2000) The alpha subunit of *E. coli* RNA polymerase activates RNA binding by NusA. Genes Dev **14:** 2664-2675

Martinez-Rucobo FW, Sainsbury S, Cheung AC und Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J **30:** 1302-1310

Mason SW, Greenblatt J (1991) Assembly of transcription elongation complexes containing the N protein of phage lambda and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. Genes Dev **5:** 1504-1512

Mason SW, Li J und Greenblatt J (1992a) Direct interaction between two *Escherichia coli* transcription antitermination factors, NusB and ribosomal protein S10. J Mol Biol **223:** 55-66

Mason SW, Li J und Greenblatt J (1992b) Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage lambda. J Biol Chem **267:** 19418-19426

Mathew R, Chatterji D (2006) The evolving story of the omega subunit of bacterial RNA polymerase. Trends Microbiol **14:** 450-455

Mittermaier A, Kay LE (2006) New Tools Provide New Insights in NMR Studies of Protein Dynamics. Science **312:** 224-228

Miwa Y, Horiguchi T und Shigesada K (1995) Structural and functional dissections of transcription termination factor rho by random mutagenesis. J Mol Biol **254:** 815-837

Mooney RA, Artsimovitch I und Landick R (1998) Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. J Bacteriol **180:** 3265-3275

Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ und Landick R (2009a) Regulator Trafficking on Bacterial Transcription Units In Vivo. Mol Cell **33:** 97-108

Mooney RA, Schweimer K, Rösch P, Gottesman ME und Landick R (2009b) 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

Morgan EA (1986) Antitermination mechanisms in rRNA operons of *Escherichia coli*. J Bacteriol **168:** 1-5

Morgan EA (1980) Insertions of Tn 10 into an *E. coli* ribosomal RNA operon are incompletely polar. Cell **21:** 257-265

Morgan WD, Bear DG, Litchman BL und von Hippel PH (1985) RNA sequence and secondary structure requirements for rho-dependent transcription termination. Nucleic Acids Res **13:** 3739-3754

Mühlberger R, Robelek R, Eisenreich W, Ettenhuber C, Sinner EK, Kessler H, Bacher A und Richter G (2003) RNA DNA Discrimination by the Antitermination Protein NusB. Journal of Molecular Biology, **327:** 973-983

Mukherjee K, Nagai H, Shimamoto N und Chatterji D (1999) GroEL is involved in activation of *Escherichia coli* RNA polymerase devoid of the omega subunit in vivo. Eur J Biochem **266:** 228-235

Mukhopadhyay J, Kapanidis AN, Mekler V, Kortkhonjia E, Ebright YW und Ebright RH (2001) Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell **106:** 453-463

Murakami KS, Masuda S, Campbell EA, Muzzin O und Darst SA (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science **296:** 1285-1290

Nicholson LK, Kay LE, Baldisseri DM, Arango J, Young PE, Bax A und Torchia DA (1992) Dynamics of methyl groups in proteins as studied by proton-detected 13C NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. Biochemistry **31:** 5253-5263

Nickels BE, Garrity SJ, Mekler V, Minakhin L, Severinov K, Ebright RH und Hochschild A (2005) The interaction between sigma70 and the beta-flap of *Escherichia coli RNA* polymerase inhibits extension of nascent RNA during early elongation. Proc Natl Acad Sci U S A **102:** 4488-4493

Nickels BE, Hochschild A (2004) Regulation of RNA polymerase through the secondary channel. Cell **118:** 281-284

Nodwell JR, Greenblatt J (1993a) Recognition of boxA antiterminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. Cell **72:** 261-268

Nodwell JR, Greenblatt J (1993b) Recognition of boxA antiterminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. Cell **72:** 261-8

Noller HF, Nomura M (1996) Ribosomes. In FC Neidhardt, R Curtiss, J Ingraham, ECC Lin, KB Low, eds, *Escherichia Coli* and *Salmonella*: Cellular and Molecular Biology, Ed 2nd Vol 1. Am. Soc. Microbiol., Washington, D.C., USA, pp 167-186

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

Paget MS, Helmann JD (2003) The sigma70 family of sigma factors. Genome Biol **4:** 203

Park JS, Roberts JW (2006) Role of DNA bubble rewinding in enzymatic transcription termination. Proc Natl Acad Sci U S A **103:** 4870-4875

Pasman Z, von Hippel PH (2000) Regulation of Rho-Dependent Transcription Termination by NusG Is Specific to the *Escherichia coli* Elongation Complex†. Biochemistry (N Y) **39:** 5573-5585

Peters JM, Vangeloff AD und Landick R (2011) Bacterial Transcription Terminators: The RNA 3'-end Chronicles. J Mol Biol **in press:**

Platt T (1981) Termination of transcription and its regulation in the tryptophan operon of *E. coli*. Cell **24:** 10-23

Prasch S, Jurk M, Washburn RS, Gottesman ME, Wöhrl BM und Rösch P (2009) RNA-binding specificity of *E. coli* NusA. Nucl Acids Res **37:** 4736-4742

Prasch S, Schwarz S, Eisenmann A, Wöhrl BM, Schweimer K und Rösch P (2006) Interaction of the intrinsically unstructured phage lambda N Protein with *E. coli* NusA. Biochemistry **45:** 4542-4549

Rachel R, Engel AM, Huber R, Stetter K- und Baumeister W (1990) A porin-type protein is the main constituent of the cell envelope of the ancestral eubacterium *Thermotoga maritima*. FEBS **262 (1):** 64-68

Reay P, Yamasaki K, Terada T, Kuramitsu S, Shirouzu M und Yokoyama S (2004) Structural and sequence comparisons arising from the solution structure of the transcription elongation factor NusG from *Thermus thermophilus*. Proteins **56:** 40-51

Rees WA, Weitzel SE, Yager TD, Das A und von Hippel PH (1996) Bacteriophage lambda N protein alone can induce transcription antitermination in vitro. Proc Natl Acad Sci U S A **93:** 342-346

Richardson JP (2002) Rho-dependent termination and ATPases in transcript termination. Biochim Biophys Acta **1577:** 251-260

Ring BZ, Yarnell WS und Roberts JW (1996) Function of *E. coli* RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. Cell **86:** 485-493

Roberts JW (2010) Molecular biology. Syntheses that stay together. Science **328:** 436-437

Roberts JW (1969) Termination factor for RNA synthesis. Nature **224:** 1168-1174

Roberts JW, Shankar S und Filter JJ (2008) RNA polymerase elongation factors. Annu Rev Microbiol **62:** 211-233

Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T und Kay LE (1996) Selective methyl group protonation of perdeuterated proteins. J Mol Biol **263:** 627-636

Sakata-Sogawa K, Shimamoto N (2004) RNA polymerase can track a DNA groove during promoter search. Proceedings of the National Academy of Sciences of the United States of America **101:** 14731-14735

Salzmann M, Pervushin K, Wider G, Senn H und WÃ¼thrich K (2000) NMR Assignment and Secondary Structure Determination of an Octameric 110 kDa Protein Using TROSY in Triple Resonance Experiments. J Am Chem Soc **122:** 7543-7548

Santangelo TJ, Roberts JW (2004) Forward translocation is the natural pathway of RNA release at an intrinsic terminator. Mol Cell **14:** 117-126

Schaechter M, Maaloe O und Kjeldgaard NO (1958) Dependency on medium and temperature of cell size and chemical composition during balanced grown of *Salmonella typhimurium*. J Gen Microbiol **19:** 592-606

Scharpf M, Sticht H, Schweimer K, Boehm M, Hoffmann S und Rosch P (2000) Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. Eur J Biochem **267:** 2397-2408

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM und Cate JHD (2005) Structures of the Bacterial Ribosome at 3.5 A Resolution. Science **310:** 827-834

Schwartz A, Margeat E, Rahmouni AR und Boudvillain M (2007) Transcription termination factor rho can displace streptavidin from biotinylated RNA. J Biol Chem **282:** 31469-31476

Schweimer K, Prasch S, Santhanam SP, Bubunenko M, Gottesman ME und Rösch P (2011) NusA interaction with the α-subunit of *E. coli* RNA Polymerase is via the UP-element site and releases autoinhibition. Structure **19:** 945-954

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

Skinner AL, Laurence JS (2008) High-field solution NMR spectroscopy as a tool for assessing protein interactions with small molecule ligands. J Pharm Sci **97:** 4670-4695

Skordalakes E, Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. Cell **114:** 135-146

Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature **445:** 618-622

Squires CL, Greenblatt J, Li J, Condon C und Squires CL (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

Squires CL, Zaporojets D (2000) Proteins shared by the transcription and translation machines. Annu Rev Microbiol **54:** 775-798

Steiner T, Kaiser JT, Marinkovic S, Huber R und Wahl MC (2002) Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. EMBO J **21:** 4641-4653

Steinmetz EJ, Platt T (1994) Evidence supporting a tethered tracking model for helicase activity of *Escherichia coli* Rho factor. Proc Natl Acad Sci U S A **91:** 1401-1405

Steitz TA (1998) A mechanism for all polymerases. Nature **391:** 231-232

Sullivan SL, Gottesman ME (1992) Requirement for *E. coli* NusG protein in factor-dependent transcription termination. Cell **68:** 989-994

Sullivan SL, Ward DF und Gottesman ME (1992) Effect of *Escherichia coli* nusG function on lambda N-mediated transcription antitermination. J Bacteriol **174:** 1339-44

Tang H, Severinov K, Goldfarb A und Ebright RH (1995) Rapid RNA polymerase genetics: one-day, no-column preparation of reconstituted recombinant *Escherichia coli* RNA polymerase. Proc Natl Acad Sci U S A **92:** 4902-4906

Thomsen ND, Berger JM (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. Cell **139:** 523-534

Tomar SK, Artsimovitch I (2013) NusG-Spt5 Proteins-Universal Tools for Transcription Modification and Communication. Chem Rev

Torchia DA, Sparks SW und Bax A (1988) Delineation of α-Helical Domains in Deuterated Staphylococcal Nuclease by 2D NOE NMR Spectroscopy. JACS **110:** 2321-2322

Torres M, Balada JM, Zellars M, Squires C und Squires CL (2004) *In vivo* effect of NusB and NusG on rRNA transcription antitermination. J Bacteriol **186:** 1304-10

Torres M, Condon C, Balada JM, Squires C und Squires CL (2001) Ribosomal protein S4 is a transcription factor with properties remarkably similar to NusA, a protein involved in both non-ribosomal and ribosomal RNA antitermination. Embo J **20:** 3811-20

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

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

Toulokhonov I, Zhang J, Palangat M und Landick R (2007) A Central Role of the RNA Polymerase Trigger Loop in Active-Site Rearrangement during Transcriptional Pausing. Mol Cell **27:** 406-419

Tugarinov V, Kay LE (2004) An isotope labeling strategy for methyl TROSY spectroscopy. J Biomol NMR **28:** 165-172

Tugarinov V, Kay LE (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc **125:** 13868-13878

Tugarinov V, Muhandiram R, Ayed A und Kay LE (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g. J Am Chem Soc **124:** 10025-10035

Tugarinov V, Sprangers R und Kay LE (2007) Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. J Am Chem Soc **129:** 1743-1750

Vassylyev DG, Sekine S, Laptenko O, Lee J, Vassylyeva MN, Borukhov S und Yokoyama S (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature **417:** 712-719

Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH und Artsimovitch I (2007a) Structural basis for transcription elongation by bacterial RNA polymerase. Nature **448:** 157-162

Vassylyev DG, Vassylyeva MN, Zhang J, Palangat M, Artsimovitch I und Landick R (2007b) Structural basis for substrate loading in bacterial RNA polymerase. Nature **448:** 163-168

Velyvis A, Schachman HK und Kay LE (2009) Assignment of Ile, Leu, and Val methyl correlations in supra-molecular systems: an application to aspartate transcarbamoylase. J Am Chem Soc **131:** 16534-16543

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

Vogel U, Jensen KF (1994) The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. J Bacteriol **176:** 2807-2813

von Hippel PH, Yager TD (1992) The elongation-termination decision in transcription. Science **255:** 809-812

Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F, Buratowski S und Handa H (1998) 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

Walstrom KM, Dozono JM und von Hippel PH (1997) Kinetics of the RNA-DNA helicase activity of *Escherichia coli* transcription termination factor rho. 2. Processivity, ATP consumption, and RNA binding. Biochemistry **36:** 7993-8004

Wang D, Bushnell DA, Westover KD, Kaplan CD und Kornberg RD (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. Cell **127:** 941-954

Wenzel S, Martins BM, Rösch P und Wöhrl BM (2009) Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. Biochem J **425:** 373-380

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

Werner F (2007) Structure and function of archaeal RNA polymerases. Mol Microbiol **65:** 1395-1404

Werner F, Grohmann D (2011) Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Microbiol **9:** 85-98

Wigneshweraraj S, Bose D, Burrows PC, Joly N, Schumacher J, Rappas M, Pape T, Zhang X, Stockley P, Severinov K und Buck M (2008) Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. Mol Microbiol **68:** 538-546

Wimberly BT, Brodersen DE, Clemons WMJ, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T und Ramakrishnan V (2000) Structure of the 30S ribosomal subunit. Nature **407:** 327-339

Winkler ME, Yanofsky C (1981) Pausing of RNA polymerase during in vitro transcription of the tryptophan operon leader region. Biochemistry **20:** 3738-3744

Woese CR (1987) Bacterial evolution. Microbiol Rev **51:** 221-271

Worbs M, Bourenkov GP, Bartunik HD, Huber R und Wahl MC (2001) An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. Mol Cell **7:** 1177-1189

Yakhnin AV, Yakhnin H und Babitzke P (2006) RNA Polymerase Pausing Regulates Translation Initiation by Providing Additional Time for TRAP-RNA Interaction. Mol Cell **24:** 547-557

Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T und Handa H (2006) P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. Mol Cell **21:** 227-237

Yamaguchi Y, Wada T, Watanabe D, Takagi T, Hasegawa J und Handa H (1999) Structure and function of the human transcription elongation factor DSIF. J Biol Chem **274:** 8085-8092

Yang X, Lewis PJ (2010) The interaction between RNA polymerase and the elongation factor NusA. RNA Biol **7:** 272-275

Yang X, Molimau S, Doherty GP, Johnston EB, Marles-Wright J, Rothnagel R, Hankamer B, Lewis RJ und Lewis PJ (2009) The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. EMBO Rep **10:** 997-1002

Yarnell WS, Roberts JW (1999) Mechanism of intrinsic transcription termination and antitermination. Science **284:** 611-615

Yee D, Armstrong VW und Eckstein F (1979) Mechanistic studies on deoxyribonucleic acid dependent ribonucleic acid polymerase from *Escherichia coli* using phosphorothioate analogues. 1. Initiation and pyrophosphate exchange reactions. Biochemistry **18:** 4116-4120

Yin YW, Steitz TA (2004) The structural mechanism of translocation and helicase activity in T7 RNA polymerase. Cell **116:** 393-404

Zalenskaya K, Lee J, Gujuluva CN, Shin YK, Slutsky M und Goldfarb A (1990) Recombinant RNA polymerase: inducible overexpression, purification and assembly of *Escherichia coli* rpo gene products. Gene **89:** 7-12

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K und Darst SA (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. Cell **98:** 811-824

Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E und Ebright RH (2012) Structural basis of transcription initiation. Science **338:** 1076-1080

Zhu AQ, von Hippel PH (1998) Rho-dependent termination within the trp t' terminator. I. Effects of rho loading and template sequence. Biochemistry **37:** 11202-11214

Zuiderweg ER (2002) Mapping protein-protein interactions in solution by NMR spectroscopy. Biochemistry **41:** 1-7

6 Publikationsliste

6.1 Einzelarbeit A

Johanna Drögemüller, Christin Schneider, Kristian Schweimer, Martin Strauß, Birgitta M. Wöhrl, Stefan H. Knauer und Paul Rösch (2016): *Thermotoga maritima* NusG: domain interaction mediates autoinhibition and thermostability. *Nucleic Acids Research, accepted*.

Christin Schneider hat unter meiner Anleitung die NMR-Titrationen von tmNusG-CTD, tmNusG-NTD und tmNusGΔ mit tmNusEΔ durchgeführt und ausgewertet. Die NMR-Experimente zur Strukturbestimmung von tmNusEΔ wurden von mir und Kristian Schweimer durchgeführt, die Rückgrat- und Seitenkettenzuordnung wurde von mir vorgenommen, während Kristian Schweimer die Struktur gelöst hat. Auch die Moleküldynamik-Simulationen wurden von Kristian Schweimer durchgeführt. Martin Strauß hat die α- und β-Untereinheiten der tmRNAP gereinigt. Alle weiteren Experimente wurden von mit geplant und durchgeführt. Birgitta M. Wöhrl hat mit mir die Klonierung der tmRNAP-Untereinheiten geplant. Stefan H. Knauer hat mit mir die Fluoreszenz-Anisotropie-Messungen geplant. Das Manuskript wurde von mir, Kristian Schweimer, Stefan H. Knauer und Paul Rösch verfasst.

6.2 Einzelarbeit B

Johanna Drögemüller*, Christian M. Stegmann*, Angshuman Mandal, Thomas Steiner, Björn M. Burmann, Max E. Gottesman, Birgitta M. Wöhrl, Paul Rösch, Markus C. Wahl und Kristian Schweimer (2013): An Autoinhibited State in the Structure of *Thermotoga maritima* NusG. *Structure* **21,** 365-375

* beide Autoren haben im gleichen Maße zur Arbeit beigetragen

Die Kristallstrukturanalyse von tmNusG wurde von Christian M. Stegmann, Angshuman Mandal und Thomas Steiner durchgeführt. Björn M. Burmann und Kristian Schweimer haben tmNusGΔ zugeordnet. Max E. Gottesman hat die *in vivo*-Tests in *E. coli* vorgenommen. Birgitta Wöhrl hat mich bei der Planung der Klonierungen und Reinigungen unterstützt. Während meiner Masterarbeit habe ich die NMR-Struktur von tmNusGΔ, die Relaxationsmessungen und den HD-Austausch durchgeführt. Während meiner Doktorarbeit habe ich tmNusG-NTD hergestellt und ihre Interaktion mit tmNusG-CTD analysiert, ebenso wie die Dynamik der Domäneninteraktion von tmNusGΔ. Auch die Mutanten von tmNusGΔ habe ich während meiner Doktorarbeit hergestellt und analysiert. Die NMR-Messungen und Auswertungen wurden von Kristian Schweimer unterstützt. Das Manuskript wurde von mir, Kristian Schweimer, Max E. Gottesman, Markus C. Wahl und Paul Rösch verfasst.

6.3 Einzelarbeit C

Johanna Drögemüller*, Martin Strauß*, Kristian Schweimer, Birgitta M. Wöhrl, Stefan H. Knauer und Paul Rösch (2015): Exploring RNA Polymerase Regulation by NMR Spectroscopy. *Scientific Reports* **5**: 10825

* beide Autoren haben im gleichen Maße zur Arbeit beigetragen

Martin Strauß hat die RNAP-Untereinheiten α und β kloniert und gereinigt, während ich β‘ und ω kloniert und gereinigt habe. Alle Messungen mit diesen beiden Untereinheiten wurden von mir geplant und durchgeführt. Der Zusammenbau und die Reinigung der RNAP und alle anschließenden Messungen wurden von Martin Strauß und mir vorgenommen. Die selektive Markierung von Methylgruppen wurde von mir etabliert. Birgitta M. Wöhrl hat bei der Planung der Klonierungen und Reinigungen geholfen, Kristian Schweimer bei der Planung und Durchführung der NMR-Experimente. Stefan H. Knauer unterstützte die Planung der Experimente. Das Manuskript wurde von mir, Martin Strauß, Stefan H. Knauer und Paul Rösch verfasst.

6.4 Einzelarbeit D

Martin Strauß*, Johanna Drögemüller*, Kristian Schweimer, Marcel Jurk, Stefan H. Knauer und Paul Rösch (2015): Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Scientific Reports* **5**: 16428

* beide Autoren haben im gleichen Maße zur Arbeit beigetragen

Die Struktur von NusA-NTD wurde von Marcel Jurk und Kristian Schweimer gelöst. Die Proteinreinigungen wurden von Martin Strauß und mir vorgenommen, genauso wie die Entwicklung der Messmethode zur Bestimmung der Bindestellen. Die Messungen haben Martin Straß und ich durchgeführt und die Auswertung der Ergebnisse von NusG-NTD und NusEΔ:NusB erfolgte durch mich, während Martin Strauß die Ergebnisse von NusA-NTD ausgewertet hat. Die Verdrängungsexperimente mit RNAP, NusEΔ:NusB und NusG-CTD wurden von mir und Martin Strauß geplant, durchgeführt und ausgewertet. Die Docking-Modelle wurden von Stefan H. Knauer erzeugt. Das Manuskript wurde von Martin Strauß, Stefan H. Knauer, Paul Rösch und mir verfasst.

7 Einzelarbeit A

Johanna Drögemüller, Christin Schneider, Kristian Schweimer, Martin Strauß, Birgitta M. Wöhrl, Stefan H. Knauer und Paul Rösch (2015): *Thermotoga maritima* NusG: domain interaction mediates autoinhibition and thermostability. Nucleic Acids Research, accepted.

Thermotoga maritima NusG: domain interaction mediates autoinhibition and thermostability

Johanna Drögemüller, Christin Schneider, Kristian Schweimer, Martin Strauß, Birgitta M. Wöhrl, Paul Rösch, and Stefan H. Knauer*

Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany

* To whom correspondence should be addressed. Tel: 0049 921 553868; Fax: 0049 921 16490459; Email: stefan.knauer@uni-bayreuth.de

Present Address: Christin Schneider, Lehrstuhl für Biochemie und Molekulare Medizin, Friedrich-Alexander-Universität Erlangen-Nürnberg, Fahrstraße 17, 91054 Erlangen, Germany

ABSTRACT

NusG, the only universally conserved transcription factor, comprises an N- and a C-terminal domain (NTD, CTD) that are flexibly connected and move independently in *Escherichia coli* and other organisms. In NusG from the hyperthermophilic bacterium *Thermotoga maritima* (tmNusG), however, NTD and CTD interact tightly. This closed state stabilizes the CTD, but masks the binding sites for the interaction partners Rho, NusE, and RNA polymerase (RNAP), suggesting that tmNusG is autoinhibited. Furthermore, tmNusG and some other bacterial NusGs have an additional domain, DII, of unknown function.

Here we demonstrate that tmNusG is indeed autoinhibited and that binding to RNAP may stabilize the open conformation. We identified two interdomain salt bridges as well as Phe336 as major determinants of the domain interaction. By successive weakening of this interaction we show that after domain dissociation tmNusG-CTD can bind to Rho and NusE, similar to the *Escherichia coli* NusG-CTD, suggesting that these interactions are conserved in bacteria. Furthermore, we show that tmNusG-DII interacts with RNAP as well as nucleic acids with a clear preference for double stranded DNA. We suggest that tmNusG-DII supports tmNusG recruitment to the transcription elongation complex and stabilizes the tmNusG:RNAP complex, a necessary adaptation to high temperatures.

INTRODUCTION

Transcription is catalyzed by the enzyme RNA polymerase (RNAP) in all domains of life. RNAP is highly regulated by numerous transcription factors, amongst them the N-utilization substance (Nus) factor NusG, the only universally conserved transcription factor (called Spt5 in *archaea* and *eukarya*) (1, 2). *Escherichia coli* (*E. coli*) NusG (ecNusG) comprises an N-terminal and a C-terminal domain (NTD, CTD) that are connected by a flexible 15 amino acid linker (3). ecNusG-NTD binds to the clamp of RNAP by making bridging contacts between the β' clamp helices (β'CH) and the β gate loop (βGL) (4, 5). This interaction is proposed to lock the clamp in a closed state to keep RNAP processive and to suppress pauses (5). ecNusG-CTD is the target of several interaction partners. It either binds to termination factor Rho, supporting Rho-dependent termination, or it interacts with protein S10 of the 30S subunit of the ribosome, thus coupling transcription and translation (6, 7). S10 is identical to antitermination factor NusE, and as such it can form a complex with NusB. This heterodimer is part of the multiprotein antitermination complex in which RNAP is modified to read through termination signals, a process necessary for the transcription of ribosomal RNA operons as well as of lambdoid phages (8-10).

The NTD and CTD of NusG from the hyperthermophilic bacterium *Thermotoga maritima* (*T. maritima*, tmNusG) tightly interact with each other, in striking contrast to the corresponding domains of ecNusG (11, 12). As the binding sites for RNAP, NusE, and Rho are masked in this closed state of tmNusG, the respective interactions are supposed to be repressed (11). Compared with ecNusG, tmNusG comprises an additional domain, DII, that is inserted into tmNusG-NTD (11). tmNusG-DII is composed of two subdomains and interacts neither with tmNusG-NTD nor with tmNusG-CTD. An insertion is also present in NusG from *Aquifex aeolicus* (aaNusG) which, however, is significantly smaller than tmNusG-DII and shows no sequential or structural homology to tmNusG-DII (11, 13, 14). Although non-sequence specific binding to nucleic acids has been demonstrated for the additional domain in both tmNusG and aaNusG, the functions of these domains remain unclear (13, 15).

NusE/S10 is an important interaction partner of NusG. In *E. coli,* NusE (ecNusE) is a monomeric, globular protein with a ribosome binding loop. This loop has no function in antitermination but is important for its insertion into the 30S subunit, where it folds into a two-stranded β-sheet (16). In solution, this loop is unstructured and causes severe protein instability. Even a deletion variant, in which this loop (aa 46-67) is replaced by a single Ser residue (ecNusEΔ), is only stable in complex with ecNusB (16). During antitermination the ecNusE:ecNusB complex binds a highly conserved RNA sequence (*boxA*) (8, 17). While NusE appears to be the active transcription factor, the function of NusB is to recruit NusE to RNA, as NusB's affinity to *boxA*-RNA is significantly stronger than that of NusE (16, 18).

So far, the intramolecular domain interaction in tmNusG is unique among NusG proteins. Here we used nuclear magnetic resonance (NMR) spectroscopy and targeted amino acid substitutions to demonstrate that tmNusG is indeed autoinhibited and to identify the major determinants of the tight domain interaction. Furthermore, NMR and fluorescence spectroscopic studies were carried out to elucidate possible roles for tmNusG-DII.

MATERIAL AND METHODS

Cloning

The codon optimized gene encoding tmNusEΔ (residues 1, Gly, 3-45, Ser, 68-102) fused to the codons corresponding to an N-terminal His_6 tag and a Tobacco Etch Virus (TEV) protease cleavage site was obtained from GenScript (Piscataway, NJ, USA) and cloned into pET29b (Novagen, Madison, WI, USA) *via NdeI* and *BamHI* restriction sites.

The plasmid pETM11-tmNusGΔ^{R275A,R279A} coding for tmNusGΔ^{R275A,R279A} (11) was used as template for site-directed mutagenesis to generate tmNusGΔ^{R275A,R279A,F336A} with the following primers: 5'-gta aac gta act ata gcc gga cgt gaa act cc-3', 5'- gg agt ttc acg tcc ggc tat agt tac gtt tac -3'. Point mutations were introduced according to the QuikChange kit (Agilent Technologies, Santa Clara, CA, USA) and confirmed by sequencing (GATC Biotech, Köln, Germany). *tmNusG-DII* (encoding residues 42-233 of tmNusG) was cloned into pETGB1a (provided by Gunter Stier, EMBL Heidelberg, Germany) *via NcoI* and *BamHI* restriction sites using pET22b-tmNusG (provided by Markus Wahl, FU Berlin, Germany) as template and the primers 5'-cat gcc atg gct gaa gag gta gtt ttg gac-3' and 5'-cgg gat ccc tac ggg aag agt ttt ctt ctt g-3'. The resulting protein carried a His_6 tag followed by the 56 amino acid immunoglobulin binding domain B1 of streptococcal protein G (GB1) and a TEV protease cleavage site at its N-terminus.

The codon optimized *tmNusB* gene was obtained from GenScript (Piscataway, NJ, USA) and cloned into pET16b (Novagen, Madison, WI, USA) using *NcoI* and *BamHI* restriction sites.

The codon optimized gene encoding tmRho was obtained from GenScript (Piscataway, NJ, USA) and cloned into the *E. coli* expression vector pET101/D-TOPO with the Champion™ pET101 Directional TOPO© Expression Kit (Invitrogen, Carlsbad, CA, USA).

Genes encoding the β, β', and ω subunits of *T. maritima* RNAP (tmRNAP) were amplified by polymerase chain reaction (PCR) from genomic *T. maritima* DNA (DSMZ, Braunschweig, Germany) with the following primers:

tmrpoB: 5'-gga att cca tat gaa aga gat ctc ttg cgg tag g-3', 5'-cgg gat cct cag tac ttg tcg ata tct atc tcg-3'

tmrpoC: 5'-gga att cca tat gcc aat gtc ctc ttt caa gag g-3', 5'-acg cgt cga ccg cga gtt ctt ctt cca ctg c-3'

tmrpoZ: 5'-gga att cca tat gga aaa aat tgt gaa gtt cg-3', 5'-ggg gta cct cac ttc acc ttc gga atg-3'

tmrpoB and *tmrpoC* were both cloned into pET29b (Novagen, Madison, WI, USA) using *NdeI* and *BamHI* or *NdeI* and *SalI* restriction sites, respectively. This allowed the production of the β' subunit with a C-terminal nine amino acid linker followed by a His_6 tag. The gene encoding the ω subunit was cloned into the multiple cloning site 2 of pACYCDuet-1 (Novagen, Madison, WI, USA) *via NdeI* and *KpnI* restriction sites.

Gene expression and protein purification

tmNusEΔ. *His_6-tmNusEΔ* was expressed in *E. coli* Rosetta (DE3) pLysS (Novagen, Madison, WI, USA). Cells were grown in lysogeny broth (LB) medium supplemented with 30 µg/ml kanamycin and 34 µg/ml chloramphenicol at 37 °C until an optical density at 600 nm (OD_{600}) of 0.6-0.8 was reached. Gene expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cells were harvested 4 h after induction by centrifugation (9,000 x *g*, 15 min, 4 °C). Cells were resuspended in buffer A (50 mM Tris/HCl, pH 7.5, 300 mM NaCl, 10 mM imidazole) and lysed using a microfluidizer (Microfluidics, Newton, MA, USA). After centrifugation (12,000 x *g*, 30 min, 4 °C) the crude extract was applied to a HisTrap FF column (GE Healthcare, Munich, Germany). Elution was performed with a step gradient with increasing imidazole concentrations (10-500 mM in buffer A). His_6-tmNusEΔ containing fractions were combined and cleaved during overnight dialysis at 4 °C (50 mM Tris/HCl, pH 7.5, 300 mM NaCl; molecular weight cut-off (MWCO): 1,000 Da) by TEV protease. TEV protease and the cleaved off GB1 tag were removed by a second nickel affinity chromatography using the same conditions as above. Pure tmNusEΔ was in the wash fraction as it exhibits weak and non-specific binding to the HisTrap FF column. It was dialyzed against the required buffer, concentrated by ultrafiltration (Vivascience, Hannover, Germany; MWCO: 1,000 Da), frozen in liquid nitrogen, and stored at -80 °C.

tmNusB. *E.coli* BL21(DE3) cells (Novagen, Madison, WI, USA) were transformed with pET16b_tmNusB and grown in LB medium containing 100 µg/ml ampicillin. Overexpression was induced by 1 mM IPTG at an OD_{600} of 0.6-0.8. 4 hours after induction cells were harvested by centrifugation (9000 x *g*, 15 min, 4 °C) and resuspended in buffer B (75 mM potassium phosphate, pH 7.0). After addition of half a protease inhibitor cocktail tablet (cOmplete, EDTA-free, Roche Diagnostics GmbH, Mannheim, Germany), 0.05 g lysozyme (Roche Diagnostics GmbH, Mannheim, Germany), and DNase I (Roche Diagnostics GmbH, Mannheim, Germany), the cell suspension was stirred on ice for 30 min. Cells were lysed with a microfluidizer (Microfluidics, Newton, MA, USA) and the lysate centrifuged (12,000 x *g*, 30 min, 4 °C). The supernatant was incubated at 80 °C for 30 min and centrifuged (12,000 x *g*, 30 min, 4 °C). Subsequently, nucleic acids were precipitated by addition of 0.6 % (v/v) polyethylenimine and 15 min stirring on ice. After centrifugation (12,000 x *g*, 30 min, 4 °C) $(NH_4)_2SO_4$ was added to the supernatant to a final concentration of 1.1 M, and the protein solution was applied to a HiTrap HP Butyl column (GE Healthcare, Munich, Germany). After washing with buffer B supplemented with 1 M KCl and 1 M $(NH_4)_2SO_4$, elution was performed using a step gradient with buffer B. tmNusB containing fractions were combined and dialyzed over night against buffer C (50 mM Tris/HCl, pH 7.5, 50 mM NaCl). The protein solution was applied to a HiTrap Heparin HP column (GE Healthcare, Munich, Germany) and elution was performed by a step gradient with increasing NaCl concentrations (50 mM-1 M in buffer C). Fractions containing pure tmNusB were combined, dialyzed against the desired buffer, concentrated by ultrafiltration (Vivascience, Hannover, Germany; MWCO: 5,000 Da), frozen in liquid nitrogen, and stored at -80 °C.

tmNusG. *E. coli* BL21 (DE3) cells (Novagen, Madison, WI, USA) containing pET22b-tmNusG were grown in LB medium in the presence of 100 µg/ml ampicillin at 37 °C. At an OD_{600} of 0.6-0.8

expression was induced with 1 mM IPTG and cells were harvested 3 h after induction by centrifugation (9000 x *g*, 15 min, 4 °C). Cells were resuspended in buffer D (50 mM Tris/HCl, pH 7.5, 100 mM NaCl), lysed with a microfluidizer (Microfluidics, Newton, MA, USA) and the lysate was heated to 85 °C for 20 min. After centrifugation (12,000 x *g*, 30 min, 4 °C), the supernatant was applied to a HiTrap Heparin HP column (GE Healthcare, Munich, Germany). Elution was performed with a step gradient with increasing NaCl concentrations (100 mM-1 M in buffer D). Fractions containing pure tmNusG were combined, dialyzed against the required buffer, concentrated by ultrafiltration (MWCO = 10,000 Da), frozen in liquid nitrogen, and stored at -80 °C.

tmNusG-NTD, tmNusG-CTD, tmNusGΔ. The proteins were produced and purified as described (11). The protocol for tmNusGΔ was also used for all its variants.

tmNusG-DII. The gene encoding tmNusG-DII fused to a His_6-GB1 tag was expressed in *E. coli* Rosetta (DE3) (Novagen, Madison, WI, USA) harboring the plasmid pETGB1a-tmNusG-DII. Gene expression, cell lysis and purification were carried out as described for tmNusEΔ, with an MWCO of 3.5 or 5 kDa for dialysis or ultrafiltration, respectively.

tmRho. *E. coli* BL21 (DE3) cells (Novagen, Madison, WI, USA) were transformed with pET101/D-TOPO_tmRho and grown at 37 °C in the presence of 100 µg/ml ampicillin. After induction with 0.4 mM IPTG at an OD_{600} of 0.6-0.8 cells were grown for 3 h, harvested by centrifugation (9,000 x *g*, 15 min, 4 °C) and resuspended in buffer E (100 mM potassium phosphate, pH 7.6, 100 mM KCl). The crude extract was prepared as described for tmNusEΔ and then heated to 90 °C for 15 min in a water bath. Subsequently, nucleic acids were precipitated by addition of 0.2 % (v/v) polyethylenimine and 15 min stirring on ice. After centrifugation (12,000 x *g*, 30 min, 4 °C) $(NH_4)_2SO_4$ was added to the supernatant to a final concentration of 1.1 M, and the protein solution was applied to a HiTrap HP Butyl column (GE Healthcare, Munich, Germany). After washing with buffer E supplemented with 1 M $(NH_4)_2SO_4$, elution was performed using a step gradient with buffer E containing 10 % (v/v) 2-propanole. tmRho containing fractions were combined, dialyzed over night against buffer C, and concentrated by ultrafiltration (MWCO = 10,000 Da). The protein was applied to a Superdex 200 column (Tricorn 10/300; GE Healthcare, Munich, Germany), and the fractions containing pure tmRho were combined. tmRho dialysis and storage were performed as above (MWCO = 10,000 Da). Concentrations of tmRho always refer to the hexamer.

β subunit of tmRNAP. The gene coding for the β subunit of tmRNAP was expressed in *E. coli* Rosetta (DE3) pLysS (Novagen, Madison, WI, USA). Cells were grown at 37 °C supplemented with 30 µg/ml kanamycin and 34 µg/ml chloramphenicol. At an OD_{600} of ~ 0.8 expression was induced by addition of 0.1 mM IPTG. Cells were harvested 3 h after induction by centrifugation (9,000 x *g*, 15 min), resuspended in 50 mM Tris/HCl (pH 7.3), 50 mM NaCl, and lysed as described for tmNusEΔ. After centrifugation (30 min, 4 °C, 12,000 x *g*) the pellet was dissolved in 1 mM EDTA, pH 8.0, supplemented with 1 mg/ml deoxycholic acid sodium salt, 20 mM DTT, and 0.2 mg/ml lysozyme, and again centrifuged (30 min, 4 °C, 12,000 x *g*). The pellet was washed three times with the same

solution, three times with 50 mM Tris/HCl, pH 8.0, 50 mM NaCl, 10 mM EDTA, 5 mM DTT, and once with H_2O. Finally, the pellet was resuspended in 50 mM Tris/HCl, pH 7.3, 8 M urea, 500 mM NaCl and stirred for 1 h at room temperature. Urea was removed by dialysis against 50 mM Tris/HCl, pH 7.3, 5 % (v/v) glycerol, 500 mM NaCl, 0.5 mM EDTA, 1 mM DTT at 4 °C for 3 h and the same buffer without NaCl overnight. The dialysate was centrifuged (30 min, 4 °C, 12,000 x *g*) and the supernatant was applied to a Q Sepharose FF column (GE Healthcare, Munich, Germany). After washing with buffer F (50 mM Tris/HCl, pH 7.5, 5 % (v/v) glycerol, 0.5 mM EDTA, 1 mM DTT) elution was performed using a constant NaCl gradient (0-1 M NaCl) in buffer F. Fractions containing pure tmβ were combined, dialyzed against the required buffer, concentrated by ultrafiltration (MWCO = 10,000 Da), and stored at -80 °C after freezing in liquid nitrogen.

The complex of β' and ω subunits of tmRNAP. The genes of β' and ω subunits of tmRNAP (tmβ', tmω) were coexpressed from individual plasmids in *E. coli* BL21 (DE3) cells (Novagen, Madison, WI, USA) to prevent tmβ' from degradation in LB medium in the presence of 30 µg/ml kanamycin and 34 µg/ml chloramphenicol. At an OD_{600} of 0.6-0.8 expression was induced with 0.1 mM IPTG and cells were harvested (9000 x *g*, 15 min, 4 °C) after 4 h. Cells were resuspended in buffer G (50 mM Tris/HCl, pH 7.5, 500 mM NaCl, 10 mM imidazole) and the crude extract was prepared as described for tmNusEΔ. The supernatant was applied to a HisTrap HP column (GE Healthcare, Munich, Germany). The column was washed with buffer G, and the tmβ'ω complex was eluted using a step gradient with increasing imidazole concentrations (10-500 mM imidazole in 50 mM Tris/HCl, pH 7,5, 500 mM NaCl). Fractions containing tmβ'ω were combined and treated as tmβ above.

Isotopic labeling. ^{15}N- and ^{15}N-, ^{13}C-labeled proteins were obtained from *E. coli* cells grown in M9 minimal medium (19, 20) with $(^{15}NH_4)_2SO_4$ (Campro Scientific, Berlin, Germany) and ^{13}C D-glucose (Spectra Stable Isotopes, Columbia, MD, USA), respectively, as only nitrogen and carbon source. Expression and purification were as described for proteins produced in LB medium.

NMR spectroscopy

NMR measurements were performed on Bruker *Avance* 600 MHz, 700 MHz, and 800 MHz spectrometers, the latter two equipped with cryogenically cooled probes. All spectra were recorded at 323 K and the initial sample volume was always 550 µL, if not stated otherwise. Standard heteronuclear double and triple resonance experiments were conducted for resonance assignment at 323 K (tmNusEΔ: backbone and side chain; tmNusG-DII: backbone) (21, 22). ^{15}N- and ^{13}C-edited 3D nuclear Overhauser-effect spectroscopy (NOESY) experiments were recorded with mixing times of 120 ms for obtaining distance restraints.

For hydrogen/deuterium (H/D) exchange measurements the ^{15}N-labeled proteins were in 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5, 50 mM NaCl. After lyophilization proteins were dissolved in D_2O (99.98 %), and the decay of signal intensities was observed in a series of [^1H,^{15}N]-heteronuclear single quantum coherence (HSQC) spectra. Exchange rates were determined by fitting the signal decay to a monoexponential curve, and the protection factors were

calculated by dividing the experimental exchange rates by the intrinsic exchange rates calculated from the amino acid sequence and experimental conditions with tabulated parameters (23, 24).

^{15}N longitudinal and transverse relaxation rates of tmNusGΔ^{R275A,R279A,F336A} were determined using standard methods at 600.2 MHz ^{1}H frequency and a calibrated temperature of 323 K (buffer: 25 mM HEPES, pH 7.5, 50 mM NaCl) (25). The relaxation rates (R_1, R_2) were fitted to monoexponential decays using the program CURVE FIT (A. G. Palmer, Dept. of Biochemistry and Molecular Biophysics, Columbia University, USA). The rotation correlation times were calculated with TENSOR2 (26).

For the interaction studies of tmNusG variants and tmNusEΔ proteins were in 25 mM HEPES, pH 7.5, 50 mM NaCl. To evaluate titration experiments we calculated the normalized chemical shift changes $\Delta\delta_{norm}$ according to equation (1).

$$\Delta\delta_{norm} = \sqrt{(\Delta\delta^{1}\mathrm{H})^2 + [0.1\cdot(\Delta\delta^{15}\mathrm{N})]^2} \qquad \text{(eq. 1)}$$

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

The dissociation constant (K_D) was calculated from [^{1}H,^{15}N]-HSQC titrations by analyzing the chemical shift changes and fitting a two-state model with equation 2 to the chemical shift change of residues showing fast exchange in the chemical shift timescale.

$$\Delta\nu_{norm} = \Delta\nu_{End}\cdot\frac{[P]_0\cdot r + [P]_0 + K_D - \sqrt{(K_D + [P]_0 + [P]_0\cdot r)^2 - 4\cdot[P]_0^2\cdot r}}{2\cdot[P]_0} \qquad \text{(eq. 2)}$$

where $\Delta\nu_{norm}$ is the normalized resonance frequency difference in Hz, $\Delta\nu_{End}$ the normalized resonance frequency difference between free and fully bound labeled protein in Hz, K_D the dissociation constant, r the ligand:labeled protein ratio and $[P]_0$ the total concentration of labeled protein. K_D and $\Delta\nu_{End}$ were used as fitting parameters. The reduction of $[P]_0$ due to dilution was accounted for during fitting.

Interaction studies between ^{15}N-tmNusG variants and tmRho, tmRNAP subunits, or *E. coli* RNAP (ecRNAP) were carried out with 20-30 μM labeled protein and the signal intensity in one-dimensional (1D) [^{1}H,^{15}N] HSQC spectra after addition of the binding partner was analyzed. Intensities were normalized by concentration and number of scans. As pulse lengths changed less than 1 % upon addition of the potential binding partner, the influence of these changes on signal intensity was neglected. Binding to tmRho was measured in 100 mM potassium phosphate (pH 7.6), 100 mM KCl at 298 K, binding to the tmRNAP subunits in 10 mM potassium phosphate, pH 7.5, 50 mM NaCl, 5 % (v/v) glycerol, 0.5 mM EDTA at 323 K, and binding to ecRNAP in 10 mM potassium phosphate, pH 7.5, 50 mM NaCl, 5 % (v/v) glycerol, 0.5 mM EDTA, 2 mM DTT at 298 K.
Backbone assignments of tmNusG-CTD, tmNusGΔ, tmNusGΔ^{R279A} tmNusGΔ^{R275A,R279A} were taken from a previous study (11) and used to assign tmNusGΔ^{R275A,R279A,F336A}. Spectra of tmNusG variants were recorded with proteins in 25 mM HEPES, pH 7.5, 50 mM NaCl.

Structure calculation of tmNusEΔ

tmNusEΔ was in 25 mM HEPES, pH 7.5, 50 mM NaCl and all spectra were recorded at 323 K. 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. Experimental NOESY spectra were validated semi-quantitatively against back-calculated spectra to confirm the assignment and to avoid bias of upper distance restraints by spin diffusion. Dihedral restraints were taken from analysis of chemical shifts by the TALOS software package (27). Hydrogen bonds were included for backbone amide protons in regular secondary structure, when the amide proton does 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 with a three-step simulated annealing protocol with floating assignment of prochiral groups including a conformational database potential (28). The 20 structures (out of a total of 120 structures) showing the lowest values of the target function excluding the database potential were further analyzed with X-PLOR, PROCHECK and PyMOL (28-30).

Molecular Modeling

In order to test the possibility of an intramolecular interdomain interaction of ecNusG and aaNusG analogous to the closed conformation in tmNusG, we explored closed states of ecNusG and aaNusG by molecular dynamics simulations. Structural models of ecNusG and aaNusG with domain interaction were constructed by fitting the structures of the individual ecNusG or aaNusG domains to the tmNusG structure. During the simulation using a standard XPLOR protocol (XPLOR-NIH 2.1.2) for solution structure calculation, backbone coordinates were held fixed for all residues except the linker region between NTD and CTD (ecNusG: residues 116-127, aaNusG: residues 183-192) and the flexible region (residues 45-64) to reduce steric clashes between NTD and CTD in the initial structure. No further restraints were applied. This approach does not include attractive forces, but is used only for testing possible conformations that do not sacrifice typical bond geometries.

Programs

Graphical representations of protein structures as well as superimpositions were created with PyMOL (30). Sequence alignments were calculated with ClustalW (31) and edited with GeneDoc (http://www.psc.edu/biomed/genedoc).

Fluorescence anisotropy

Fluorescence anisotropy titrations were performed on a Fluorolog-Tau-3 spectrofluorometer (HORIBA Jobin Yvon GmbH, Unterhaching, Germany). 6-carboxyfluorescein (6-FAM) labeled nucleic acids (table S1) were titrated with tmNusG-NTD, tmNusG-DII or tmNusG-CTD at 25 °C. Double stranded (ds) DNA was generated by mixing the single stranded (ss) oligonucleotides in equimolar

concentrations, heating them to 95 °C for 1 min, followed by incubation for 10 min at 60 °C and cooling them to room temperature. Titrations were carried out in 1.3 ml 25 mM HEPES, pH 7.5, 50 mM NaCl for DNA or 25 mM Tris/HCl, pH 7.7, 50 mM NaCl for RNA in a 10 x 4 mm quartz cuvette (Hellma, Müllheim, Germany) using 5 or 50 nM of nucleic acids. The sample was excited at 492 nm, and emission was recorded at 515-517 nm. The slit widths were set to 3 nm and 4 nm for excitation and emission, respectively, when 50 nM nucleic acid were used, and to 7 and 8 nm for 5 nM nucleic acid. After sample equilibration, each data point was collected six times with an integration time of 0.8 s. The anisotropic data was fitted to a two-state binding model (equations 2 + 3) for determination of the K_D values using GraFit 5.0 (Erithacus Software).

$$A = \frac{A_S + \frac{[\text{complex}]}{[\text{S}]_0} \cdot \left(R \cdot A_{complex} - A_S\right)}{1 - \frac{[\text{complex}]}{[\text{S}]_0} + R \cdot \frac{[\text{complex}]}{[\text{S}]_0}} \qquad \text{(eq. 2)}$$

with

$$[\text{complex}] = \frac{K_D + [\text{P}]_0 + [\text{S}]_0 - \sqrt{\left(K_D + [\text{P}]_0 + [\text{S}]_0\right)^2 - 4 \cdot [\text{S}]_0 \cdot [\text{P}]_0}}{2} \qquad \text{(eq. 3)}$$

where A is the measured anisotropy, A_S the anisotropy of free nucleic acid, $A_{complex}$ the anisotropy of the complex, [complex] the concentration of the complex, $[S]_0$ and $[P]_0$ the total nucleic acid and protein concentrations, respectively, K_D the dissociation constant, and R the ratio of the fluorescence intensities of fully bound and free substrate. Each titration was measured three times.

RESULTS AND DISCUSSION

Solution structure of tmNusEΔ

ecNusE interacts with ecNusG-CTD (6). To test whether a similar interaction also occurs in *T. maritima*, we first solved the solution structure of tmNusE. Initial attempts to study the full length protein, however, failed due to its structural instability after purification, a situation already known from ecNusE (16, 32). Substitution of the ribosomal binding loop (residues 46-67) by a single Ser residue, tmNusEΔ, resulted in samples suitable for NMR spectroscopic studies, even in the absence of tmNusB. The [^{1}H,^{15}N]-HSQC spectra of tmNusEΔ showed the large chemical shift dispersion typical for globular, folded proteins, even up to 80 °C (Supplementary Figures S1A-C). This is in stark contrast to ecNusEΔ, which exists in a structured state only in the NusEΔ:NusB complex or when it is part of the ribosome (16, 33).

With multidimensional heteronuclear NMR spectroscopy, resonances were assigned and the solution structure of tmNusEΔ was determined based on 1341 distance and 110 dihedral restraints derived from multiple NMR experiments (Table 1, Figure 1A). Free tmNusEΔ is composed of a four-stranded, antiparallel β-sheet, flanked by two roughly antiparallel α-helices on one side. The structure superimposes well with that of ecNusEΔ in the ecNusEΔ:ecNusB complex (Figure 1B, root mean square deviation for backbone atoms: 1.0 Å). Characteristic structural features of ecNusEΔ, e.g. the cis-conformation of Pro39 in the ecNusEΔ:ecNusB interface, are also present in tmNusEΔ, suggesting that these characteristics do not result from ecNusEΔ:ecNusB complex formation. Titration of ^{15}N-tmNusEΔ with tmNusB indicated protein interaction (Supplementary Figure S1D). However, due to exchange processes close to or in the intermediate range of the NMR timescale, ^{15}N-tmNusEΔ signals disappear upon tmNusB addition. Although finally the signals of the complex reappear, tracking of signal shifts during the titration and resonance assignment of the complex was not possible. ^{15}N-tmNusEΔ signals with similar chemical shifts in the free and complex form are most likely from residues that are not directly involved in tmNusB binding. Comparison of tmNusEΔ with ecNusEΔ in the ecNusEΔ:ecNusB complex shows that these residues are located opposite of the ecNusB binding surface in ecNusEΔ, suggesting a similar type of interaction of the *E. coli* and the *T. maritima* proteins. These data and the high structural similarity to ecNusEΔ may indicate that tmNusEΔ fulfills the same functions as its *E. coli* homolog.

Table 1: Experimental constraints for structure calculation and statistics of tmNusEΔ.

Experimental derived restraints		
distance restraints		
	NOE	1265
	intraresidual	467
	sequential	321
	medium range	182
	long range	295
	hydrogen bonds	37
dihedral restraints		110
restraint violation		
average distance restraint violation (Å)	0.0029 +/- 0.0002	
distance restraint violation > 0.1 Å	0	
average dihedral restraint violation (°)	0.05 +/- 0.02	
dihedral restraint violation > 1°	0	
deviation from ideal geometry		
bond length (Å)	0.00043 +/- 0.00001	
bond angle (°)	0.08 +/- 0.002	
coordinate precision[a,b]		
backbone heavy atoms (Å)	0.44	
all heavy atoms (Å)	0.94	
Ramachandran plot statistics[c] (%)	93.8/ 5.1/ 0.1/ 1.0	

[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 Gly3-Val78
[c] Ramachandran plot statistics are determined by PROCHECK (29) and noted by most favored/additionally allowed/generously allowed/disallowed.

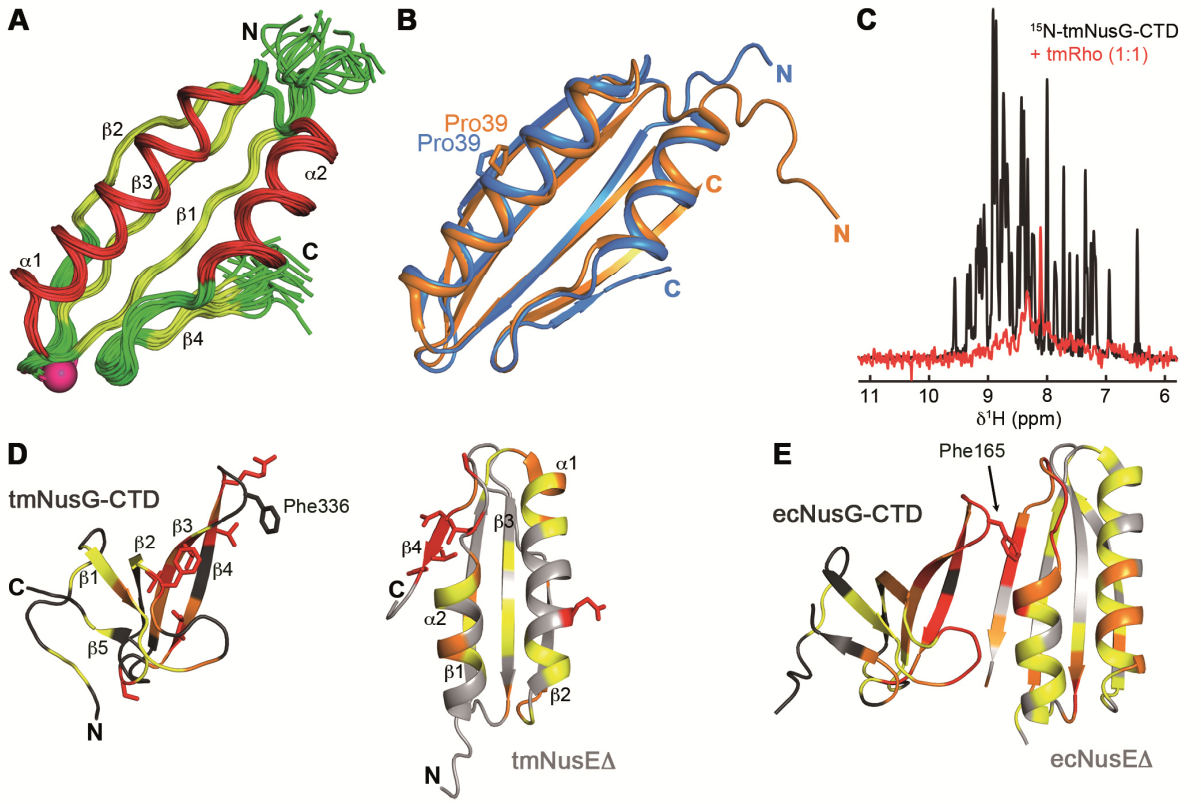


Figure 1: tmNusEΔ and tmRho are interaction partners of tmNusG-CTD. (A) Solution structure of tmNusEΔ. Superposition of 20 accepted structures of tmNusEΔ in ribbon representation. α-helices, red; β-strands, yellow; loops, green. Secondary structure elements and termini are labeled. The Ser replacing the ribosome binding loop is shown as pink sphere. **(B)** Superposition of tmNusEΔ, blue, and ecNusEΔ in the ecNusEΔ:ecNusB complex, orange (Protein Data Bank (PDB) ID: 3D3B). Selected amino acids are shown as sticks and labeled. **(C)** 1D [^{1}H,^{15}N]-HSQC spectra of 338 μM ^{15}N-tmNusG-CTD, black, and of 27 μM ^{15}N-tmNusG-CTD in the presence of tmRho in equimolar concentrations, red. **(D)** Interaction of tmNusEΔ with tmNusG-CTD. tmNusG-CTD (dark grey; PDB ID: 2LQ8) and tmNusEΔ (light grey) are in ribbon representation. Normalized chemical shift changes of the [^{1}H,^{15}N]-HSQC titrations of ^{15}N-tmNusEΔ with tmNusG-CTD and ^{15}N- tmNusG-CTD with tmNusEΔ are mapped on the structures. $\Delta\delta_{norm}$ > 0.2 ppm, red; 0.2 ppm > $\Delta\delta_{norm}$ > 0.1 ppm, orange; 0.1 ppm > $\Delta\delta_{norm}$ > 0.04 ppm, yellow. Highly affected amino acids are in sticks representation. **(E)** Structure of the ecNusEΔ:ecNusG-CTD complex (PDB ID: 2KVQ). Representation as in **(D)**. ecNusG-CTD, dark grey; ecNusEΔ, light grey. Normalized chemical shift changes of [^{1}H,^{15}N]-HSQC titrations are taken from Ref. (6).

tmNusEΔ and tmRho are targets of tmNusG-CTD

As NusE and Rho are both interaction partners of NusG-CTD in *E. coli*, we asked whether tmNusG-CTD had the analogous targets in *T. maritima* (6). First, 1D [^{1}H,^{15}N]-HSQC spectra of ^{15}N-tmNusG-CTD in the absence and in the presence of tmRho were recorded (Figure 1C). ^{15}N-tmNusG-CTD signals decreased significantly upon addition of tmRho, indicating interaction of these proteins as complex formation dramatically increases the molecular mass (MM) of tmNusG-CTD ($MM_{tmNusG-CTD}$: 7 kDa, MM_{tmRho}: 290 kDa), which, in turn, leads to faster magnetization relaxation and finally results in significant line broadening. Only some signals corresponding to amino acids in random coil areas remain visible. Weak or no binding to tmRho was detected for tmNusG-NTD and tmNusG-DII,

respectively (Supplementary Figure S2). The weak tmNusG-NTD:tmRho interaction is probably unspecific and might be attributed generally to the hydrophobic area of tmNusG-NTD responsible for RNAP binding. This confirms that tmNusG interacts specifically with tmRho *via* its CTD, just like ecNusG, suggesting that NusG:Rho binding is conserved in bacteria.

To test if tmNusG-CTD binds to tmNusEΔ we performed [^{1}H,^{15}N]-HSQC NMR titrations in which either ^{15}N-labeled tmNusEΔ was titrated with unlabeled tmNusG-CTD or *vice versa* (Supplementary Figures S3A,B). In both cases signals of the ^{15}N-labeled protein shifted significantly or disappeared upon stepwise addition of the unlabeled partner due to tmNusG-CTD:tmNusEΔ complex formation. The normalized changes of the chemical shifts ($\Delta\delta_{norm}$) were plotted against the amino acid sequences and mapped on the structures to identify the binding surfaces (Supplementary Figures S3C,D and Figure 1D). In tmNusEΔ strongly affected residues are predominantly found at the C-terminus of β-strand β4 as well as in the preceding loop. Binding of tmNusEΔ had an effect on two regions of tmNusG-CTD in particular that involve β-strands β3 and β4 as well as the loop between β1 and β2. The binding interface between tmNusEΔ and tmNusG-CTD is highly similar to that of the ecNusEΔ:ecNusG-CTD complex (Figure 1E) (6). In ecNusG-CTD, Phe165 is a key residue for NusE binding (6), and the corresponding residue in tmNusG-CTD, Phe336, is also located in the loop between β3 and β4 (Figures 1D,E). This suggests that Phe336 is also involved in tmNusEΔ binding, although direct evidence could not be found as we could not assign this residue in the [^{1}H,^{15}N]-HSQC spectra. Since overlapping ecRho and ecNusEΔ binding sites have been proposed for ecNusG-CTD, this might also be true for tmNusG-CTD as suggested by the structural similarity of ecNusG-CTD and tmNusG-CTD (6, 11).

From the chemical shift perturbations a K_D-value of around 13 μM was estimated for the tmNusEΔ:tmNusG-CTD interaction (Supplementary Figures S3E,F), indicating tighter binding than in the *E. coli* system (50 μM) (6). Moreover, owing to the high protein concentrations required for NMR experiments, the K_D-values in the low micromolar range rather represent an upper limit. Although the titrations were carried out at 50 °C, a temperature close to the growth conditions of *T. maritima* (55-90 °C), the K_D might still be different at the optimal growth temperature of 80 °C (34). Neither for tmNusG-NTD nor for tmNusG-DII an interaction with tmNusEΔ was detected (Supplementary Figure S4). The high similarity of the binding interfaces of NusG-CTD and NusE in *T. maritima* and *E. coli* in combination with the fact that *T. maritima* and *E. coli* are evolutionary distant (35) suggests a conserved binding mode for NusG-CTD and NusE in bacteria, an interaction important for both transcription:translation coupling and antitermination.

Domain interaction renders tmNusG autoinhibited

Since tmNusG-CTD has interaction partners analogous to those of ecNusG-CTD, we suggest that it also carries out analogous functions. In the closed state, however, the binding sites for tmRho and tmNusE on tmNusG-CTD and that for RNAP on tmNusG-NTD are masked, implying an autoinhibitory regulation for tmNusG (11). To test whether full length tmNusG was able to bind tmNusEΔ or tmRho, we recorded 1D or two-dimensional (2D) [^{1}H,^{15}N]-HSQC spectra of ^{15}N-tmNusG in the absence and presence of these factors (Figures 2A,B). However, neither addition of tmNusEΔ nor addition of tmRho

changed the corresponding spectrum, indicating that no complex formation occurs. Repeating the experiments with ^{15}N-tmNusGΔ, a variant in which tmNusG-DII is replaced by the linker of ecNusG (11), led to identical results (Figures 2C,D). These data indicate that the closed conformation of tmNusG is indeed a silent state. The open and closed conformation are in dynamic equilibrium with the majority of tmNusG being in the closed form. Neither tmNusEΔ nor tmRho can shift this equilbrium towards the open state.

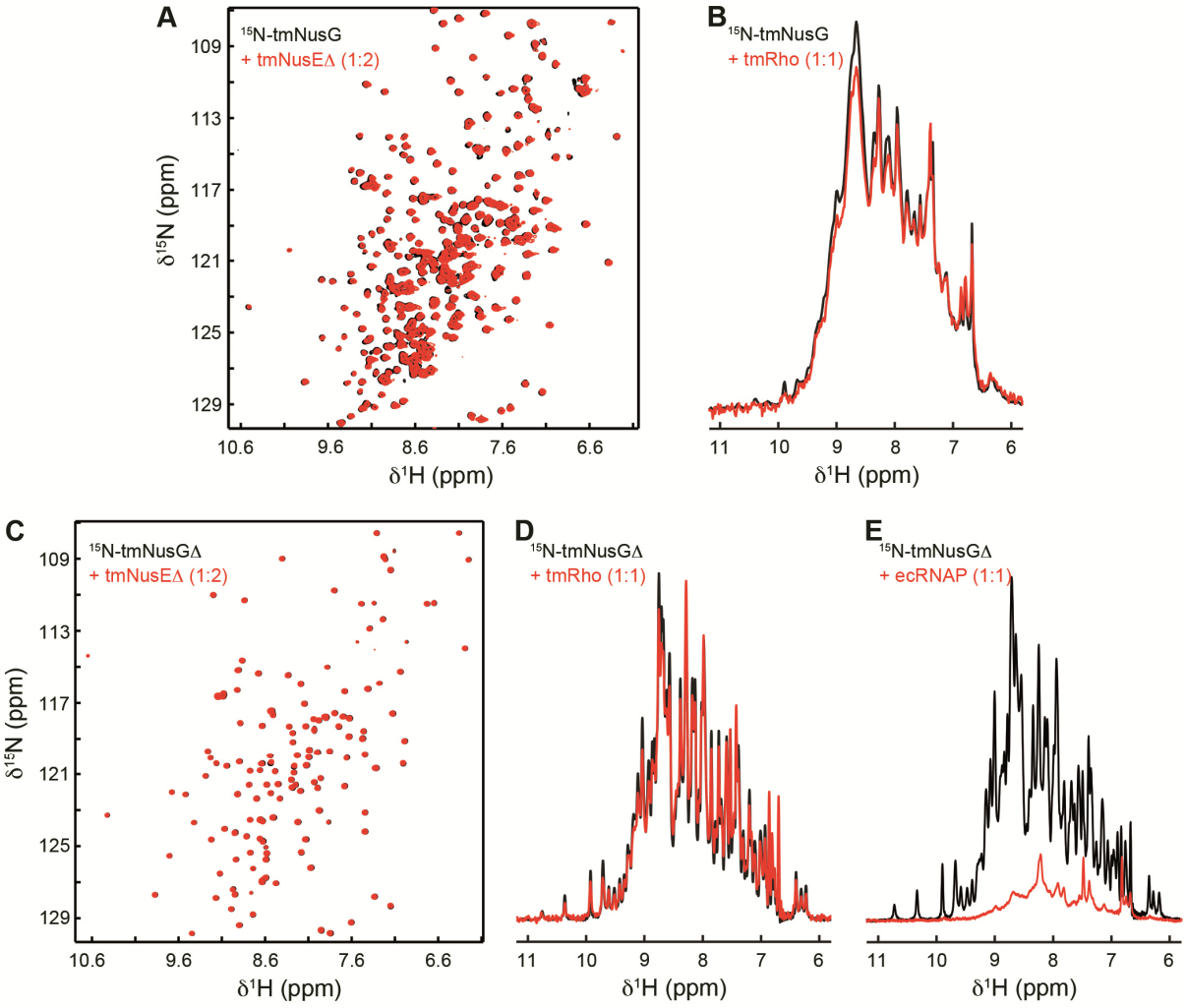


Figure 2: tmNusG and tmNusGΔ are autoinhibited. (A) 2D [^{1}H,^{15}N]-HSQC spectra of 150 μM ^{15}N-tmNusG, black, and of 89 μM ^{15}N-tmNusG in the presence of tmNusEΔ in a twofold molar excess, red. **(B)** 1D [^{1}H,^{15}N]-HSQC spectra of 278 μM ^{15}N-tmNusG, black, and of 30 μM ^{15}N-tmNusG in the presence of tmRho in equimolar concentrations, red. **(C)** 2D [^{1}H,^{15}N]-HSQC spectra of 100 μM ^{15}N-tmNusGΔ, black, and of 74 μM ^{15}N-tmNusGΔ in the presence of tmNusEΔ in a twofold molar excess, red. **(D)** 1D [^{1}H,^{15}N]-HSQC spectra of 348 μM ^{15}N-tmNusGΔ, black, and of 27 μM ^{15}N-tmNusGΔ in the presence of tmRho in equimolar concentrations, red. **(E)** 1D [^{1}H,^{15}N]-HSQC spectra of 348 μM ^{15}N-tmNusGΔ, black, and of 25 μM ^{15}N-tmNusGΔ in the presence of ecRNAP in equimolar concentrations, red.

Next we tested whether ^{15}N-tmNusGΔ interacts with RNAP. Since RNAP from *T. maritima* (tmRNAP) was not available, we used RNAP from *E. coli* (ecRNAP; Figure 2E). In the 1D [^{1}H,^{15}N]-HSQC spectrum ^{15}N-tmNusGΔ signals decreased significantly upon ecRNAP addition, suggesting complex formation. This implies that ecRNAP stabilizes the open conformation by interacting with tmNusG-

NTD. Although we cannot exclude that the autoinhibited form interacts non-specifically with ecRNAP, RNAP may be the signal to promote NusG domain dissociation.

Determinants of the domain interaction in tmNusGΔ

The intramolecular interaction in tmNusG strongly increases the stability of tmNusG-CTD (ΔΔG = 10.3 kJ/mol at 50 °C) and is so far unique for NusG proteins (11). Two interdomain salt bridges, Arg275:Asp314 and Arg279:Glu313, were suggested to contribute to the tmNusG-NTD:tmNusG-CTD interaction. However, their elimination results only in partial domain opening, indicating that additional interactions are responsible for the stabilization of the autoinhibited state (11). Another important determinant of the domain interaction may be Phe336 in tmNusG-CTD as it points into a hydrophobic pocket of the RNAP binding site on tmNusG-NTD (Figure 3A) (11). Due to the similarity between the NusG-CTD:NusEΔ binding sites in *E. coli* and *T. maritima* (Figure 1D,E) and the fact that EcNusEΔ and EcRho interaction surfaces on EcNusG-CTD overlap (6), Phe336 in tmNusG-CTD is probably involved in the interaction of tmNusG-CTD with tmRho and tmNusE.

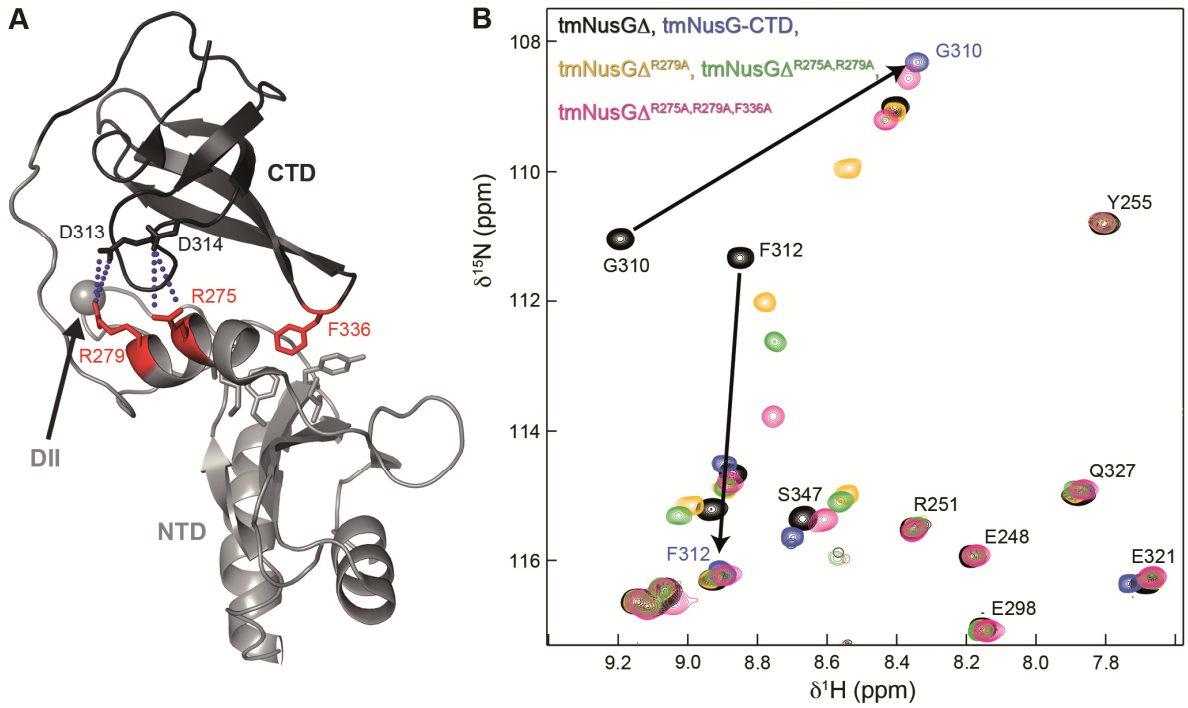


Figure 3: Determinants of the domain interactions in tmNusGΔ. (A) Cartoon representation of tmNusGΔ (PDB ID: 2LQ8). tmNusG-NTD, light grey; tmNusG-CTD, dark grey. Amino acids involved in domain interaction are shown as sticks with residues which were exchanged by alanines colored in red. The grey sphere marks the position where tmNusG-DII is integrated into tmNusG-NTD. Salt bridges are indicated by blue dots. **(B)** Section of the superposition of 2D [^{1}H,^{15}N]-HSQC spectra of ^{15}N-tmNusGΔ (100 μM), black, ^{15}N-tmNusG-CTD (150 μM), cyan, ^{15}N-tmNusGΔ^{R279A} (100 μM), yellow, ^{15}N-tmNusGΔ^{R275A,R279A} (100 μM), green, and ^{15}N-tmNusGΔ^{R275A,R279A,F336A} (420 μM), pink. Arrows indicate how tmNusG-CTD signals of tmNusGΔ shift towards the signals of isolated tmNusG-CTD upon successive amino acid exchanges.

To quantify the contribution of the salt bridges and Phe336 to the domain interaction, we stepwise exchanged Arg275, Arg279, and Phe336 by Ala in tmNusGΔ and measured H/D exchange. The lyophilized, ^{1}H,^{15}N-labeled proteins were dissolved in D_2O and H/D exchange was monitored at 323 K

via the decay of signal intensities in a series of 2D [^{1}H,^{15}N]-HSQC spectra. Based on the amide proton exchange rates the protection factor (PF) was calculated, which is a measure for the stabilization of a conformation compared to the unfolded state (Supplementary Figure S5A). In tmNusGΔ the PFs are significantly higher than those of the isolated tmNusG-CTD, corresponding to the increased stability of the CTD in tmNusGΔ (11). Successive disruption of both salt bridges and the Phe336Ala exchange gradually lowered the PFs, finally resulting in PF values similar to those of the isolated tmNusG-CTD (Supplementary Figure S5A). This is exemplified by the PF of Gly317: $2.4x10^{8}$ in tmNusGΔ, $2.8x10^{7}$ in tmNusGΔ^{R279A}, $7.4x10^{6}$ in tmNusGΔ^{R275A,R279A}, $3.6x10^{6}$ in tmNusGΔ^{R275A,R279A,F336A}, $5.2x10^{6}$ in tmNusG-CTD (11). The decrease in PFs correlates with a weakening of the domain interaction. Thus, both tmNusGΔ^{R275A,R279A} and tmNusGΔ^{R275A,R279A,F336A} were not stabilized as compared to isolated tmNusG-CTD. Weakening of the domain interaction was confirmed by 2D [^{1}H,^{15}N]-HSQC spectra of the ^{15}N-labeled tmNusGΔ variants (Figure 3B and Supplementary Figure S5B). With every amino acid exchange the CTD signals gradually shifted from the position typical for tmNusGΔ towards the resonances of isolated tmNusG-CTD. Hence, the tmNusGΔ variants appear to reflect intermediate stages in domain opening of tmNusG, and confirming Phe336 as important factor for domain interaction (see signal shift of Phe312 or Glu339). In [^{1}H,^{15}N]-HSQC spectra of ^{15}N-labeled wild type NusG proteins from other bacteria signals of the isolated CTD and of the CTD in full length protein superimpose perfectly, demonstrating the absence of a closed conformation (ecNusG (3), *Thermus thermophilus* NusG (ttNusG) (36), *Mycobacterium tuberculosis* NusG (mtNusG) (37)). The small chemical shift differences between the spectra of tmNusGΔ^{R275A,R279A,F336A} and isolated tmNusG-CTD thus suggest still an observable contribution of the closed conformation to the NMR parameters.

To confirm these results we conducted spin relaxation experiments to analyze the relative movements of the NTD and the CTD in tmNusGΔ^{R275A,R279A,F336A}, which, in turn, characterizes the overall tumbling of the protein. Uniform rotational tumbling of a multidomain protein corresponds to tight domain interaction, whereas a multidomain protein with non-interacting domains requires an individual description of the rotational tumbling for each domain. Analysis of the ratio of the transverse relaxation rate R_2 and the longitudinal relaxation rate R_1 offers an elegant method to detect interdomain movements on a timescale faster than the overall molecular tumbling (38). This ratio shows a uniform distribution for tmNusGΔ, confirming the association of tmNusG-NTD and tmNusG-CTD (11). The R_2/R_1 distribution in tmNusGΔ^{R275A,R279A,F336A} (Supplementary Figure S5C) exhibits a slightly bimodal characteristic, reflecting a contribution of individual rotational behavior of the two domains to the overall tumbling. The effective rotation correlation times are 15.5 ns and 13.4 ns for tmNusG-NTD and tmNusG-CTD, respectively, suggesting that tmNusGΔ^{R275A,R279A,F336A} is not in a fully closed conformation. As the differences in the apparent rotational correlation times of the domains are not as big as in ecNusG (12), NTD and CTD in tmNusGΔ^{R275A,R279A,F336A} still affect each other during rotation, and the open and the closed conformation are in an equilibrium on the time scale of molecular rotation. This is consistent with the small chemical shift differences of isolated tmNusG-CTD as compared to tmNusGΔ^{R275A,R279A,F336A}. Thus, Arg275, Arg279 and Phe336 are major determinants of the autoinhibited state in tmNusG, although further interactions are involved.

Domain opening of tmNusG allows binding to tmNusEΔ and tmRho.

In contrast to tmNusGΔ isolated tmNusG-CTD is able to bind tmRho and tmNusEΔ. As RNAP is the potential trigger to promote domain separation, we used the tmNusGΔ variants with weakened domain interaction to demonstrate that tmNusG can execute the same functions as ecNusG as soon as the autoinhibited state has been released. 1D or 2D [^{1}H,^{15}N]-HSQC spectra of ^{15}N-tmNusGΔ^{R279A} recorded in the absence or presence of tmNusEΔ or tmRho clearly show no or only weak binding, respectively. This is in accordance with the H/D exchange experiments since domain association is only slightly reduced (Figures 4A,B). In contrast, significant chemical shift changes occurred when tmNusEΔ was titrated to ^{15}N-tmNusGΔ^{R275A,R279A}, and signal intensity of ^{15}N-tmNusGΔ^{R275A,R279A} decreased considerably upon addition of tmRho (Figures 4C,D), indicating domain opening. Although the additional elimination of Phe336 further decreased domain interaction (Figure 3B), tmNusGΔ^{R275A,R279A,F336A} binds neither to tmNusEΔ nor to tmRho (Supplementary Figure S6), confirming that Phe336 is essential for these interactions, just like the corresponding Phe165 in ecNusG (6).

Autoinhibition of tmNusG probably parallels its thermostability. While intermolecular interactions between NTD and CTD of bacterial NusG proteins have been reported, the tight intramolecular domain interaction in tmNusG is unique (11, 12, 14). As demonstrated, two salt bridges and Phe336 are important contributors to the domain interaction. Phe336 is highly conserved in bacterial NusG proteins (Supplementary Figure S7), probably since it is essential for NusE and Rho binding. However, the presence of Phe336 is sufficient to cause partial autoinhibition only in tmNusG. Therefore, Phe336 and its binding pocket in tmNusG-NTD are optimized with respect to tighter binding compared to other NusGs. The amino acid combinations allowing the formation of the two salt bridges are not conserved. Only in ttNusG an Arg and an Asp residue are present at the positions corresponding to Arg275 and Asp314 in tmNusG, but no NTD:CTD interaction was observed (36). tmNusG has a longer linker than several other NusGs. The linker of ttNusG is three amino acids shorter, that of ecNusG five and that of aaNusG even seven amino acids. Molecular modeling showed that the linker of ecNusG and even aaNusG are sufficient to allow a closed state similar to tmNusG. However, in either case the linker must adopt an unlikely, nearly extended conformation. Compared to tmNusG, mtNusG contains a significantly longer linker, but did not show any domain interaction (37) suggesting that linker length is not the pivotal factor for intramolecular domain interaction. Although transient intra- or intermolecular NTD:CTD interactions might occur in all NusG proteins, only tmNusG has developed additional features to stabilize the autoinhibited state. However, in contrast to the ecNusG paralog RfaH, autoinhibition in tmNusG appears to merely support thermostability and has no regulatory role (11, 39, 40).

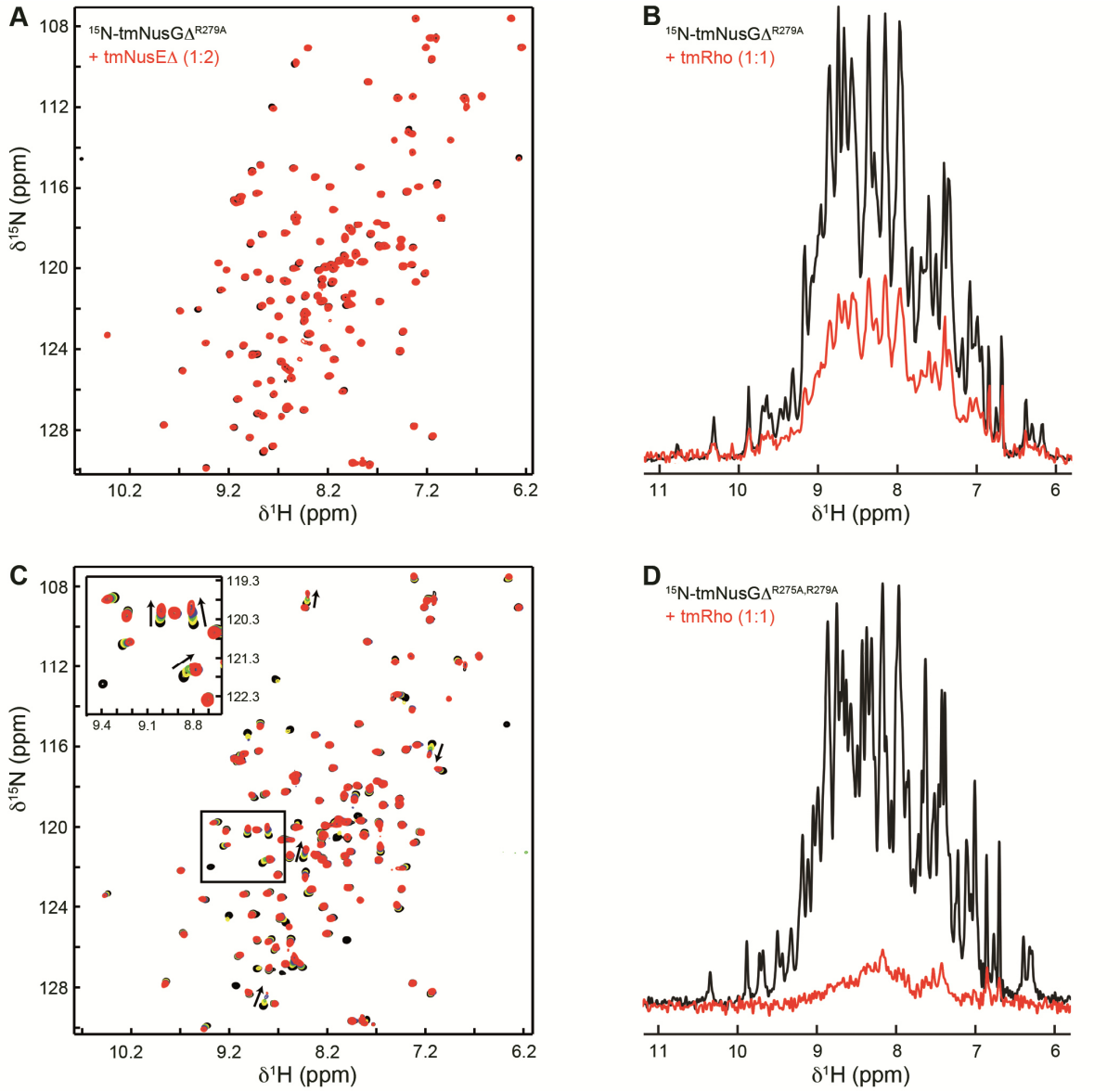


Figure 4: Interaction of tmNusGΔ variants with tmNusEΔ and tmRho. (A) 2D [^{1}H,^{15}N]-HSQC spectra of 150 µM ^{15}N-tmNusGΔ^{R279A}, black, and of 89 µM ^{15}N-tmNusGΔ^{R279A} in the presence of tmNusEΔ in a two-fold molar excess, red. **(B)** 1D [^{1}H,^{15}N]-HSQC spectra of 172 µM ^{15}N-tmNusGΔ^{R279A}, black, and of 30 µM ^{15}N-tmNusGΔ^{R279A} in the presence of tmRho in equimolar concentrations, red. **(C)** 2D [^{1}H,^{15}N]-HSQC spectra of the titration of ^{15}N-tmNusGΔ^{R275A,R279A} with tmNusEΔ. tmNusEΔ (stock concentration: 438 µM) was added to 150 µM ^{15}N-tmNusGΔ^{R275A,R279A} (molar ratios 1:0, black; 1:0.5, yellow; 1:1, green; 1:2, blue; 1:5, red). The insert shows a blow-up of the boxed region. Arrows indicate changes of the chemical shifts during the titration. **(D)** 1D [^{1}H,^{15}N]-HSQC spectra of 71 µM ^{15}N-tmNusGΔ^{R275A,R279A}, black, and of 26 µM ^{15}N-tmNusGΔ^{R275A,R279A} in the presence of tmRho in equimolar concentrations, red.

tmNusG-DII binds to RNAP

Aside from non-specific binding to nucleic acids (15), functions of tmNusG-DII are unknown. Since tmNusGΔ interacts with RNAP (Figure 2E) we analyzed if also tmNusG-DII binds to RNAP to uncover possible roles of this domain. As for tmNusGΔ we recorded 1D [^{1}H,^{15}N]-HSQC spectra of ^{15}N-tmNusG-DII in the absence and presence of ecRNAP (Figure 5A). ^{15}N-tmNusG-DII signals decreased

significantly upon addition of ecRNAP and residual signal intensity could only be found in regions typical for unstructured parts, indicating complex formation. As NusG-NTD binds to the β and β' subunits (4, 5), tmNusG-DII will probably also interact with one of these.

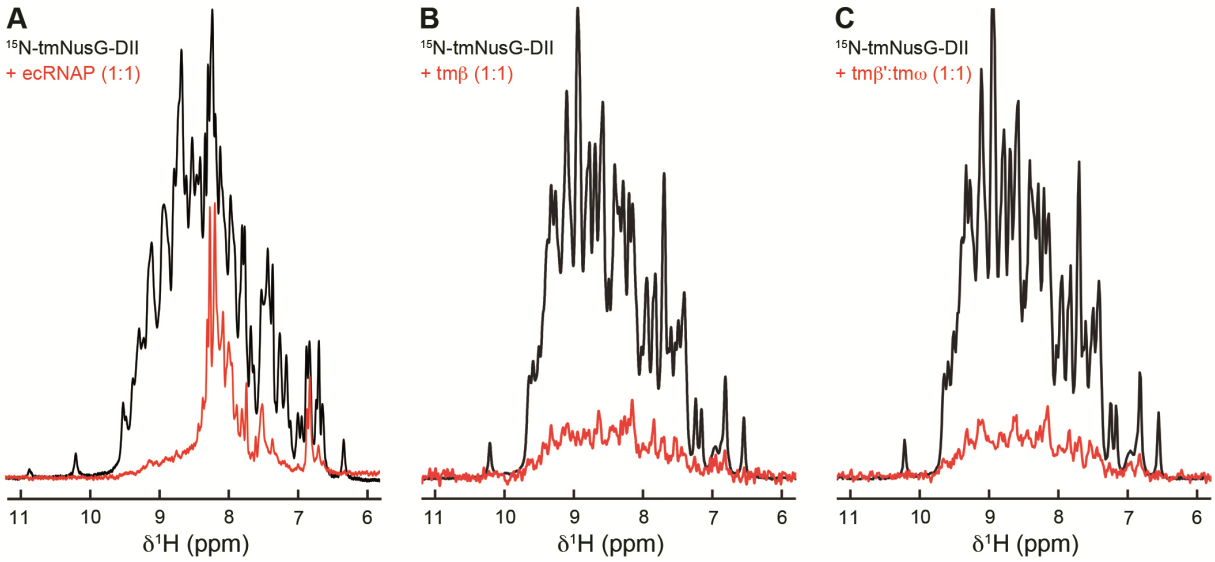


Figure 5: Binding of tmNusG-DII to ecRNAP and tmRNAP subunits. 1D [^{1}H,^{15}N]-HSQC spectra of 375 μM ^{15}N-tmNusG-DII, black, and of ^{15}N-tmNusG-DII in the presence, red, of **(A)** ecRNAP (25 μM each), **(B)** tmβ (30 μM each) and **(C)** tmβ':tmω (each 30 mM). Differences between the spectra of free ^{15}N-tmNusG-DII in **(A)** and **(B,C)** are due to different temperatures in the experiments.

We used a method previously established for the *E. coli* system to identify the interacting RNAP subunit (41). tmRNAP subunit genes coding for the β and β' subunit (tmβ, tmβ') were expressed separately. While tmβ could be purified individually, tmβ' was only stable in complex with the ω subunit (tmω). The integrity of tmβ and tmβ':tmω was checked by testing their ability to bind to known interaction partners with ecRNAP as reference (Supplementary Figure S8). The 1D [^{1}H,^{15}N]-HSQC spectra revealed an interaction of ^{15}N-tmNusG-NTD with both tmβ and tmβ':tmω whereas ^{15}N-tmNusG-CTD did not bind either subunit.

The interaction of tmNusG-DII with tmRNAP subunits was tested using the same approach. Addition of tmβ and tmβ':tmω to ^{15}N-tmNusG-DII led to a drastic signal decrease (Figures 5B,C), suggesting that tmNusG-DII binds to both the tmβ subunit and the tmβ':tmω complex.

tmNusG-DII binds preferably dsDNA

The ability of tmNusG-DII to bind non-specifically to nucleic acids has been demonstrated qualitatively by gel shift assays and electron microscopy using long oligonucleotides (0.7-5.2 kb) (15). These data suggested a preference for dsDNA and RNA over ssDNA. We used fluorescence anisotropy measurements to further elucidate these interactions by titrating 6-FAM labeled ssDNA, dsDNA and ssRNA of about 20 nucleotides (nt) in length with tmNusG-DII (Figures 6A,B and Table 2). The domain had a similar low micromolar affinity for ssDNA as for ssRNA, but exhibited a clear preference for dsDNA with a K_D value of 40 nM. Different sequences for each type of nucleic acid resulted in similar K_D values, suggesting that the interaction is sequence-independent.

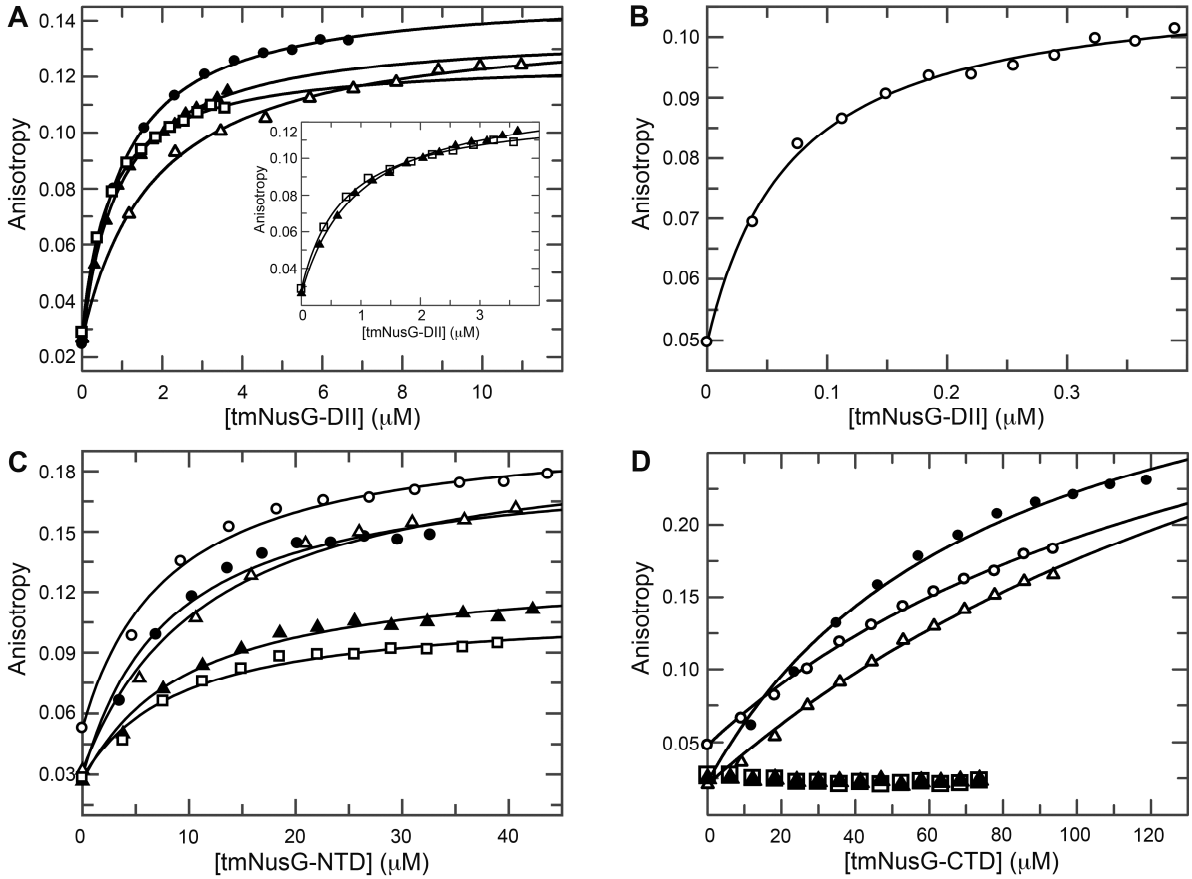


Figure 6: Fluorescence anisotropy titrations of nucleic acids with tmNusG domains. Titrations of **(A)** ssDNA 1 (filled circles), ssDNA 2 (open triangles), ssRNA 1 (filled triangles), and ssRNA 2 (open squares) and **(B)** dsDNA (open circles) with tmNusG-DII. The insert in **(A)** shows a blow-up of the titration with ssRNAs 1 and 2. **(C)** Titration of ssDNA 1, ssDNA 2, ssRNA 1, ssRNA 2 and dsDNA with tmNusG-NTD (symbols as in (A,B)). **(D)** Titration of ssDNA 1, ssDNA 2, ssRNA 1, ssRNA 2 and dsDNA with tmNusG-CTD (symbols as in **(A,B)**). Solid lines show the best fit to eq. 2. Nucleic acids were labeled with 6-FAM.

As non-specific nucleic acid binding has also been reported for a tmNusG construct lacking the DII domain, similar to the tmNusGΔ variant, we repeated the titrations with isolated tmNusG-NTD and tmNusG-CTD (Figures 6C,D) (15). While tmNusG-CTD showed only very weak binding to ssDNA and dsDNA and no binding to ssRNA, tmNusG-NTD interacted with all nucleic acids with a similar affinity of 3-4 µM. These data are consistent with the finding that *Bacillus subtilis* NusG, ttNusG and *E. coli* RfaH contact the nontemplate DNA strand when bound to the RNAP in the TEC (42-44). This seems to be a feature conserved in all NusG proteins. The main nucleic acid binding ability of tmNusG, however, can be attributed to tmNusG-DII, which preferentially binds dsDNA.

To determine the nucleic acid binding site of tmNusG-DII, we conducted a [^{1}H,^{15}N]-HSQC titration with ^{15}N-tmNusG-DII and ssDNA. Chemical shift changes as well as disappearing signals could be observed, confirming complex formation (Figure 7A). The normalized chemical shift changes of affected residues were mapped on the structure of tmNusG-DII, revealing that especially the part of tmNusG-DII that comprises subdomain 2 is involved in ssDNA binding. Although the affected residues do not form a completely continuous patch, the electrostatic surface potential reveals that the

determined binding site superimposes with a positively charged area, suggesting that we indeed identified the nucleic acid binding site.

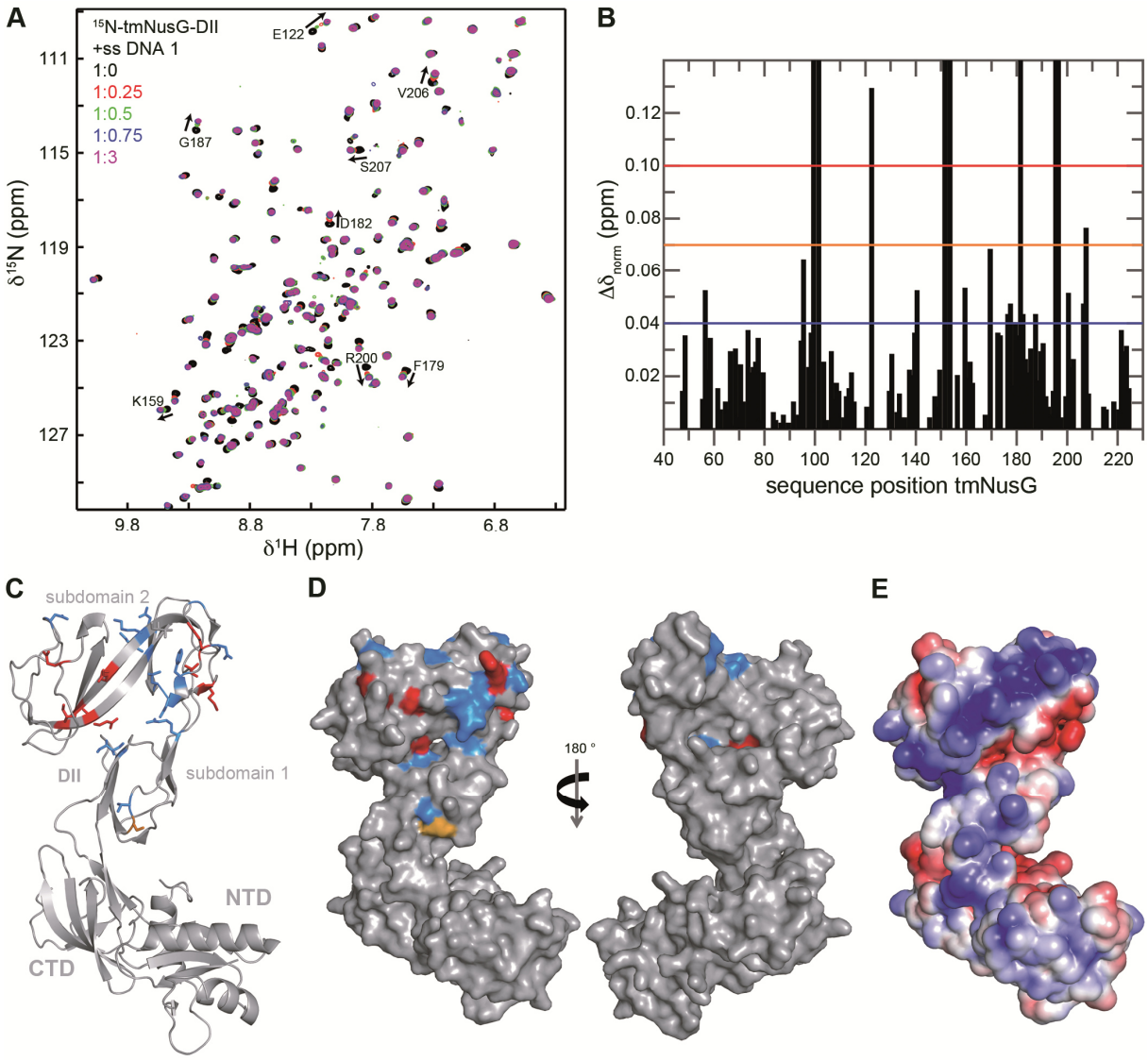


Figure 7: The DNA binding site of tmNusG-DII. (A) [^{1}H,^{15}N]-HSQC spectra of the titration of ^{15}N-tmNusG-DII with ssDNA 1. ssDNA 1 (stock concentration: 1 mM) was added to 150 µM ^{15}N-tmNusG-DII (molar ratios 1:0, black; 1:0.25, red; 1:0.5, green; 1:0.75, blue: 1:3, magenta). Selected Signals are assigned, arrows indicate changes of the chemical shifts. **(B)** Normalized chemical shift changes derived from the HSQC titration in **(A)** versus tmNusG-DII sequence position. The significance levels are indicated by horizontal lines. $\Delta\delta_{norm}$ = 0.04 ppm, blue; $\Delta\delta_{norm}$ = 0.07 ppm, orange; $\Delta\delta_{norm}$ = 0.1 ppm, red. Gaps represent prolines and not assigned amino acids. **(C, D)** Mapping of the normalized chemical shift changes on the structure of tmNusG (grey, PDB ID: 2XHC) in ribbon **(C)** and surface **(D)** representation. $\Delta\delta_{norm}$ > 0.1 ppm, red; 0.1 ppm > $\Delta\delta_{norm}$ > 0.07 ppm, orange; 0.07 ppm > $\Delta\delta_{norm}$ > 0.04 ppm, blue. **(E)** Electrostatic surface potential of tmNusG calculated with the program APBS (47), colored from -3 kT/e^{-} (red) to +3 kT/e^{-} (blue).

Conclusions: tmNusG-DII recruits tmNusG to the TEC and stabilizes the tmNusG:TEC complex

The interaction of NTD and CTD in tmNusG masks the binding sites for tmNusE, tmRho and tmRNAP, preventing these interactions and rendering tmNusG silent. We showed that domain opening is essential for tmNusG to accomplish the functions known from ecNusG. The closed and the open state

are in dynamic equilibrium with 98 % being in the closed conformation, even at temperatures close to the optimal growth conditions of *T. maritima* (11). Neither tmNusEΔ nor tmRho are able to shift the equilibrium towards the open state.

We suggest that several factors might contribute to release the autoinhibition. As shown by NMR spectroscopy interaction of tmNusG with RNAP might be the driving force to promote domain opening. Under physiological conditions, however, tmNusG binds to the TEC and not to RNAP alone. We demonstrated that tmNusG-NTD interacts non-specifically with nucleic acids, suggesting that tmNusG-NTD, once bound to the TEC, might additionally interact with the non-template strand in the transcription bubble to increase the overall affinity of tmNusG. Similar situations are known from other NusG proteins. *Bacillus subtilis* NusG, for example, recognizes a specific sequence in the non-template strand within the paused transcription bubble in certain operons (42) and also the NTD of RfaH, a paralog of NusG, interacts with a specific DNA element in the non-template strand during its recruitment to the TEC (45). Moreover, tmNusG-DII interacts with the tmβ subunit and the tmβ':tmω complex. This interaction might again increase the overall affinity of tmNusG for the TEC, facilitating its recruitment.

Based on the complex structure of *Pyrococcal* Spt4/5 and the clamp domain (4) we generated a model of tmNusG bound to elongating RNAP from *Thermus thermophilus* (ttRNAP) (Figure 8). Fluorescence spectroscopic titrations of nucleic acids with tmNusG-DII showed a clear preference of this domain for dsDNA. In the model tmNusG-DII is in close proximity to upstream dsDNA and separated from downstream dsDNA by parts of the β' subunit, suggesting that tmNusG-DII interacts with upstream dsDNA. The binding of tmNusG-DII to dsDNA might further stabilize the tmNusG:TEC complex at the high temperatures at which *T. maritima* lives. *A. aeolicus* is a hyperthermophilic organism that grows at temperatures between 67 °C and 95 °C with an optimum at 85 °C (46). Like tmNusG, aaNusG has an additional domain that binds non-specifically ssDNA, dsDNA and RNA, but that is much smaller than tmNusG-DII and exhibits no sequential similarity to tmNusG-DII (11, 13). aaNusG also differs from tmNusG as aaNusG-NTD and aaNusG-CTD do not interact, and aaNusG is not autoinhibited (13, 14). Thus, the additional domain found in NusG proteins in some (hyper)thermophilic organisms might be an adaption to the high temperatures of the natural habitats as it allows the stabilization of the TEC by either interaction with nucleic acids or with RNAP or both.

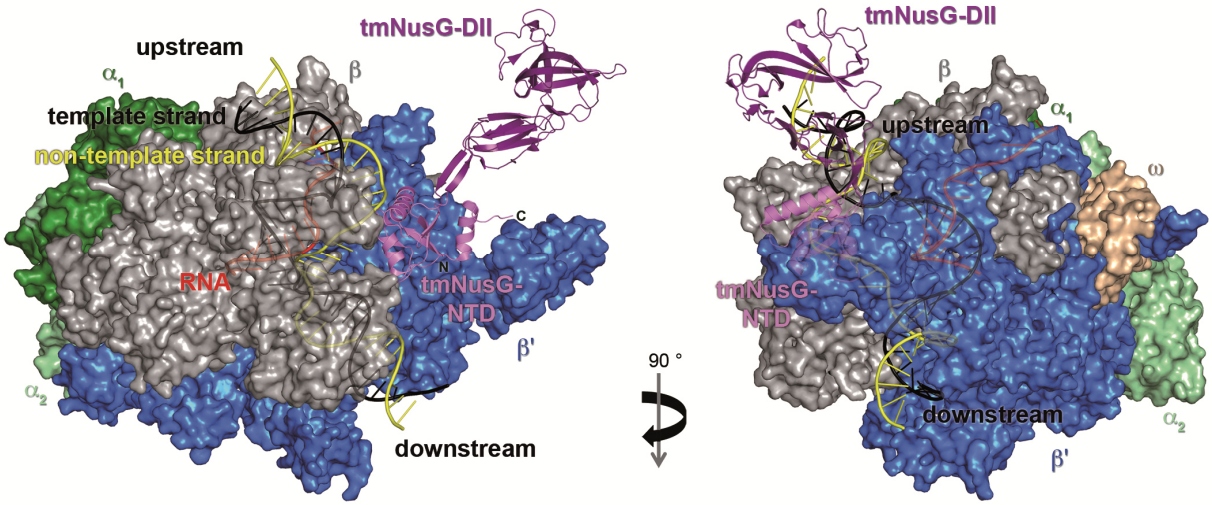


Figure 8: Model of tmNusG bound to elongating ttRNAP. Surface representation of ttRNAP with nucleic acids shown as ribbons (PDB ID: 2O5I; α1, dark green; α2, light green; β, grey; β', blue; ω, light orange; DNA template strand, black; DNA non-template strand, yellow; RNA, red). tmNusG (PDB ID: 2HXC; ribbon representation; tmNusG-NTD, violet; tmNusG-DII, purple; tmNusG-CTD, not shown; termini are labeled) was modeled to ttRNAP according to the *Pyrococcus furiosus* Spt4/5 complex bound to the RNAP clamp domain (PDB ID: 3QQC) (4) by superimposition of tmNusG-NTD on Spt5-NTD.

SUPPLEMENTARY DATA

Supplementary data are available at NAR online.

ACCESSION NUMBERS

The structure coordinates and chemical shift assignments of tmNusEΔ have been deposited in the PDB and the Biological Magnetic Resonance Data Bank with accession numbers 2MEW and 19533, respectively.

ACKNOWLEDGEMENT

We thank Ramona Heißmann and Ulrike Persau for excellent technical support and Dr. Max Gottesman, Columbia University, for helpful discussions. We also thank Gunter Stier (EMBL Heidelberg, Germany) and Markus Wahl (FU Berlin, Germany) for kindly providing pETGB1a and pET22b-tmNusG, respectively.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (grants Ro 617/17-1, Ro 617/21-1 to P.R.) and the Ludwig-Schaefer-Scholarship 2015 from the Columbia University Medical Center (to P.R.).

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. 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.

2. NandyMazumdar,M. and Artsimovitch,I. (2015) Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins - shifting shapes and paradigms. *Bioessays,* **37,** 324-334.

3. Mooney,R.A., Schweimer,K., Rösch,P., Gottesman,M.E. 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.

4. Martinez-Rucobo,F.W., Sainsbury,S., Cheung,A.C. and Cramer,P. (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *Embo J.,* **30,** 1302-1310.

5. 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.

6. 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.

7. 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.

8. Nodwell,J.R. and Greenblatt,J. (1993) Recognition of *boxA* antiterminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. *Cell,* **72,** 261-268.

9. Friedman,D.I., Olson,E.R., Georgopoulos,C., Tilly,K., Herskowitz,I. and Banuett,F. (1984) Interactions of bacteriophage and host macromolecules in the growth of bacteriophage lambda. *Microbiol. Rev.,* **48,** 299-325.

10. Torres,M., Balada,J.M., Zellars,M., Squires,C. and Squires,C.L. (2004) *In vivo* effect of NusB and NusG on rRNA transcription antitermination. *J. Bacteriol.,* **186,** 1304-10.

11. Drögemüller,J., Stegmann,C.M., Mandal,A., Steiner,T., Burmann,B.M., Gottesman,M.E., Wöhrl,B.M., Rösch,P., Wahl,M.C. and Schweimer,K. (2013) An auto-inhibited state in the crystal structure of *Thermotoga maritima* NusG. *Structure*, **21,** 365-375.

12. 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.

13. Steiner,T., Kaiser,J.T., Marinkovic,S., Huber,R. and Wahl,M.C. (2002) Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *Embo J.,* **21,** 4641-4653.

14. 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.

15. Liao,D., Lurz,R., Dobrinski,B. and Dennis,P.P. (1996) A NusG-like protein from *Thermotoga maritima* binds to DNA and RNA. *J. Bacteriol.,* **178,** 4089-4098.

16. 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.

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

18. Greive,S.J., Lins,A.F. and von Hippel,P.H. (2005) Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to *boxA* RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in *Escherichia coli*. *J. Biol. Chem.,* **280,** 36397-36408.

19. Sambrook,J. and Russel,D.W. (1994) Molecular Cloning - A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

21. Bax,A. and Grzesiek,A. (1993) Methodological advances in protein NMR. *Acc. Chem. Res.,* **26,** 131-138.

22. Sattler,M., Schleucher,J. and Griesinger,C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution empolying pulsed field gradients. *Prog. NMR Spectrosc.,* **34,** 93-158.

23. Bai,Y., Milne,J.S., Mayne,L. and Englander,S.W. (1994) Protein stability parameters measured by hydrogen exchange. *Proteins,* **20,** 4-14.

24. Bai,Y., Milne,J.S., Mayne,L. and Englander,S.W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins,* **17,** 75-86.

25. Kay,L.E., Torchia,D.A. and Bax,A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry (N. Y.),* **28,** 8972-8979.

26. Dosset,P., Hus,J.C., Blackledge,M. and Marion,D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR,* **16,** 23-28.

27. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR,* **13,** 289-302.

28. Schwieters,C.D., Kuszewski,J.J., Tjandra,N. and Clore,G.M. (2003) The xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.,* **160,** 66-74.

29. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst.,* **26,** 283-291.

30. Schrödinger L. (2010) The PyMOL molecular graphics system, version 1.3. *Schrödinger, LLC, Mannheim, Germany*.

31. Higgins,D.G., Thompson,J.D. and Gibson,T.J. (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.,* **22,** 4673-4680.

32. Das,R., Loss,S., Li,J., Waugh,D.S., Tarasov,S., Wingfield,P.T., Byrd,R.A. and Altieri,A.S. (2008) Structural biophysics of the NusB:NusE antitermination complex. *J. Mol. Biol.,* **376,** 705-720.

33. Wimberly,B.T., Brodersen,D.E., Clemons,W.M.J., Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature,* **407,** 327-339.

34. Huber,R., Langworthy,T.A., König,H., Thomm,M., Woese,C.R., Sleytr,U.B. and Stetter,K.O. (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90° C. *Arch. Microbiol.,* **144,** 324-333.

35. Woese,C.R. (1987) Bacterial evolution. *Microbiol. Rev.,* **51,** 221-271.

36. Reay,P., Yamasaki,K., Terada,T., Kuramitsu,S., Shirouzu,M. and Yokoyama,S. (2004) Structural and sequence comparisons arising from the solution structure of the transcription elongation factor NusG from *Thermus thermophilus*. *Proteins,* **56,** 40-51.

37. Strauss,M., Schweimer,K., Burmann,B.M., Richter,A., Guttler,S., Wöhrl,B.M. and Rösch,P. (2016) The two domains of *Mycobacterium tuberculosis* NusG protein are dynamically independent. *J. Biomol. Struct. Dyn.,* **34,** 352-361.

38. Horstmann,M., Ehses,P., Schweimer,K., Steinert,M., Kamphausen,T., Fischer,G., Hacker,J., Rösch,P. and Faber,C. (2006) Domain motions of the Mip protein from *Legionella pneumophila*. *Biochemistry,* **45,** 12303-12311.

39. Burmann,B.M., Knauer,S.H., Sevostyana,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.

40. Knauer,S.H., Rösch,P. and Artsimovitch,I. (2012) Transformation: The next level of regulation. *RNA Biol.,* **9,** 1418-1423.

41. Drögemüller,J., Strauss,M., Schweimer,K., Wohrl,B.M., Knauer,S.H. and Rösch,P. (2015) Exploring RNA polymerase regulation by NMR spectroscopy. *Sci. Rep.,* **5,** 10825-10835.

42. Yakhnin,A.V., Murakami,K.S. and 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,** 5299-5308.

43. Sevostyanova,A. and Artsimovitch,I. (2010) Functional analysis of *Thermus thermophilus* transcription factor NusG. *Nucleic Acids Res.,* **38,** 7432-7445.

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

45. 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.

46. Huber,R., Wilharm,T., Huber,D., Trincone,A., Burggraf,S., König,H., Reinhard,R., Rockinger,I., Fricke,H. and Stetter,K.O. (1992) A*quifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Syst. Appl. Microbiol.,* **15,** 340-351.

47. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.,* **98,** 10037-10041.

SUPPLEMENTARY INFORMATION
for
Thermotoga maritima NusG: domain interaction mediates autoinhibition and thermostability

Johanna Drögemüller, Christin Schneider, Kristian Schweimer, Martin Strauß, Birgitta M. Wöhrl, Paul Rösch, and Stefan H. Knauer*

Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany

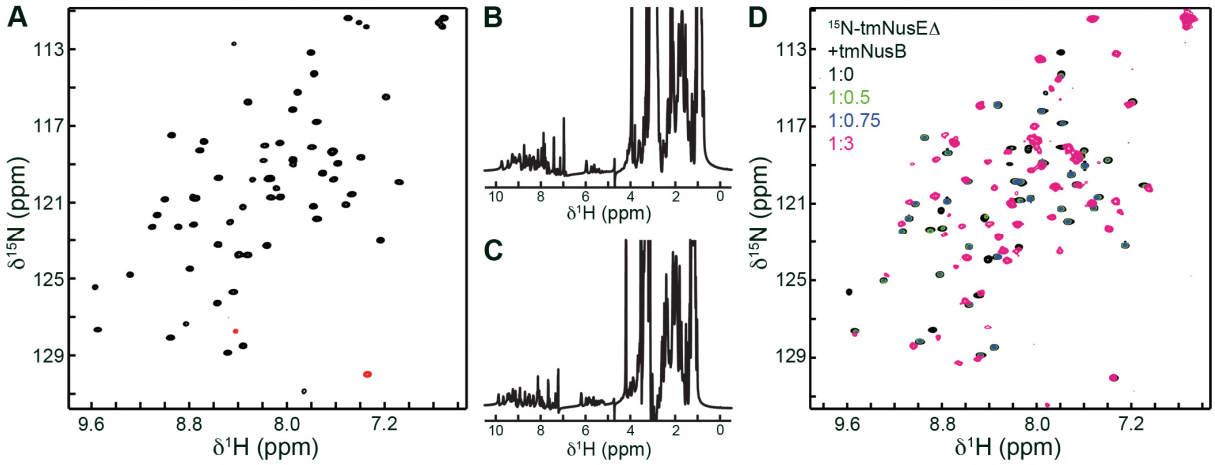
* To whom correspondence should be addressed. Tel: 0049 921 553868; Fax: 0049 921 16490459; Email: stefan.knauer@uni-bayreuth.de

Present Address: Christin Schneider, Lehrstuhl für Biochemie und Molekulare Medizin, Friedrich-Alexander-Universität Erlangen-Nürnberg, Fahrstraße 17, 91054 Erlangen, Germany

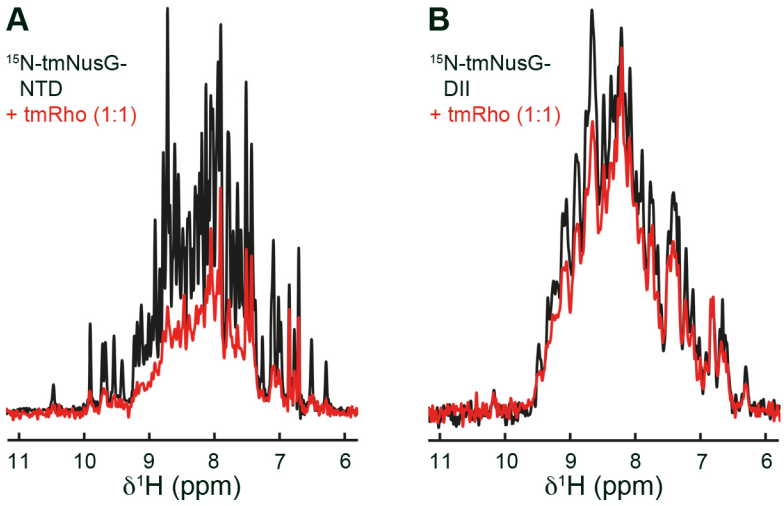
Contents:

Supplementary Table S1. Oligonucleotide sequences used in the fluorescence anisotropy experiments.

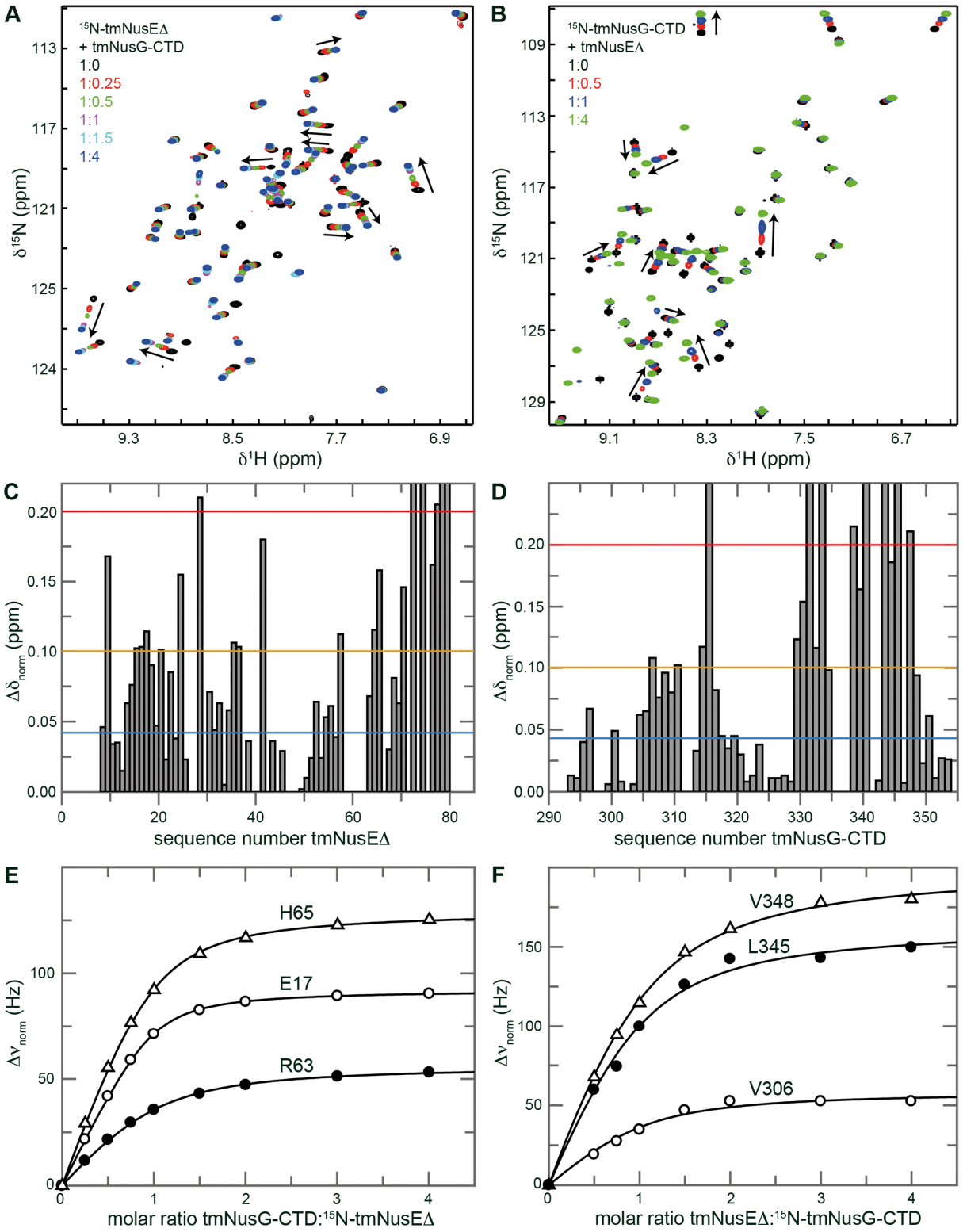
oligonucleotide	sequence	modification
ssDNA 1	5'-cgg atc cat tag ttc acc atg-3'	5' 6-FAM
ssDNA 2	5'-tag att ttg cat aac ttt aaa-3'	5' 6-FAM
dsDNA	5'-tag att ttg cat aac ttt aaa-3' 3'-atc taa aac gta ttg aaa ttt-5'	5' 6-FAM -
RNA 1	3'-ccg auc gcu cuc cug gug auc cuu ucc-5'	3' 6-FAM
RNA 2	3'-cac ugc ucu uua aca auu a-5'	5' 6-FAM



Supplementary Figure S1: NMR spectra of tmNusEΔ. (A) 2D [^{1}H,^{15}N]-HSQC spectrum of ^{15}N-tmNusEΔ at 323 K. **(B, C)** 1D [^{1}H] spectrum of tmNusEΔ at **(B)** 323 K and **(C)** 353 K. **(D)** Interaction of tmNusEΔ with tmNusB. [^{1}H,^{15}N]-HSQC spectra of 150 μM ^{15}N-tmNusEΔ titrated with tmNusB (stock concentration: 544 μM; molar ratios 1:0, black; 1:0.5, green; 1:0.75, blue; 1:3, pink). Proteins were in 25 mM HEPES, pH 7.5, 50 mM NaCl.

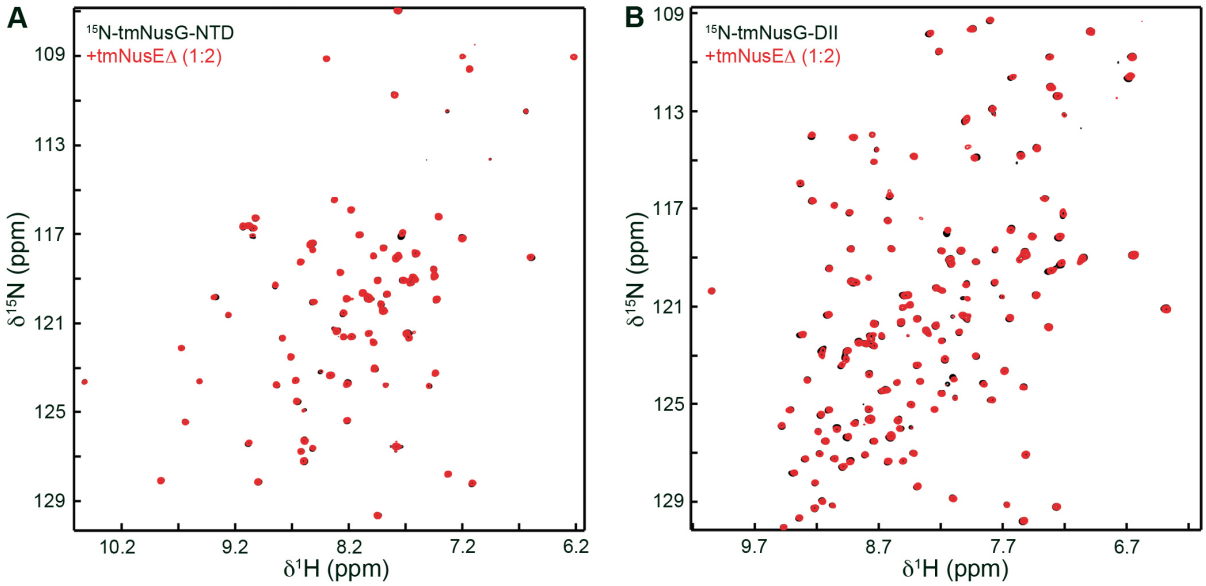


Supplementary Figure S2: Interaction studies of tmNusG-NTD and tmNusG-DII with tmRho. (A) 1D [^{1}H,^{15}N]-HSQC spectra of 94 μM ^{15}N-tmNusG-NTD, black, and 22 μM ^{15}N-tmNusG-NTD in the presence of tmRho in equimolar concentrations, red. **(B)** 1D [^{1}H,^{15}N]-HSQC spectra of 10 μM ^{15}N-tmNusG-DII, black, and 10 μM ^{15}N-tmNusG-DII in the presence of tmRho in equimolar concentrations, red.



Supplementary Figure S3: Interaction between tmNusG-CTD and tmNusEΔ. (A) [^{1}H,^{15}N]-HSQC titration of 150 μM ^{15}N-tmNusEΔ with tmNusG-CTD (stock concentration 574 μM; molar ratios 1:0, black; 1:0.25, red; 1:0.5, green; 1:1, magenta; 1:1.5, cyan; 1:4, blue). Selected Signals are assigned and arrows indicate the chemical shift changes. **(B)** [^{1}H,^{15}N]-HSQC titration of 200 μM ^{15}N-tmNusG-CTD with tmNusEΔ (stock concentration 432 μM; molar ratios 1:0, black; 1:0.5, red; 1:1, blue; 1:4, green). Selected Signals are assigned and arrows indicate the chemical shift changes. **(C, D)** Normalized chemical shift changes of ^{15}N-tmNusEΔ caused by the interaction with tmNusG-CTD **(C)** and of ^{15}N-tmNusG-CTD caused by the interaction with tmNusEΔ **(D)** are mapped against the

sequence position of the labeled protein. Significance levels are indicated by horizontal lines; $\Delta\delta_{norm}$ = 0.04 ppm, blue; $\Delta\delta_{norm}$ = 0.1 ppm, orange; $\Delta\delta_{norm}$ = 0.2 ppm, red. Gaps represent prolines and not assigned amino acids. **(E+F)** Determination of K_D values for tmNusEΔ:tmNusG-CTD binding. Titration curves of selected amino acids derived from the [^{1}H,^{15}N]-HSQC titrations in **(A)** and **(B)**. **(E)** Normalized chemical shift changes of E17 (open circles, K_D = 10 µM), R63 (filled circles, K_D = 37 µM), and H65 (open triangles, K_D = 18 µM) of ^{15}N-tmNusEΔ caused by the interaction with tmNusG-CTD are mapped against the molar ratio of tmNusG-CTD:^{15}N-tmNusEΔ. **(F)** Normalized chemical shift changes of V306 (open circles, K_D = 3 µM), L345 (filled circles, K_D = 9 µM), and V347 (open triangles, K_D = 26 µM) of ^{15}N-tmNusG-CTD caused by the interaction with tmNusEΔ are mapped against the molar ratio of tmNusEΔ:^{15}N-tmNusG-CTD.



Supplementary Figure S4: tmNusG-NTD and tmNusG-DII do not bind to tmNusEΔ. (A) 2D [^{1}H,^{15}N]-HSQC spectra of 150 µM ^{15}N-tmNusG-NTD, black, and of 93 µM ^{15}N-tmNusG-NTD in the presence of a twofold molar excess of tmNusEΔ, red. **(B)** 2D [^{1}H,^{15}N]-HSQC spectra of 150 µM ^{15}N-tmNusG-DII, black, and of 93 µM ^{15}N-tmNusG-DII in the presence of a twofold molar excess of tmNusEΔ, red.

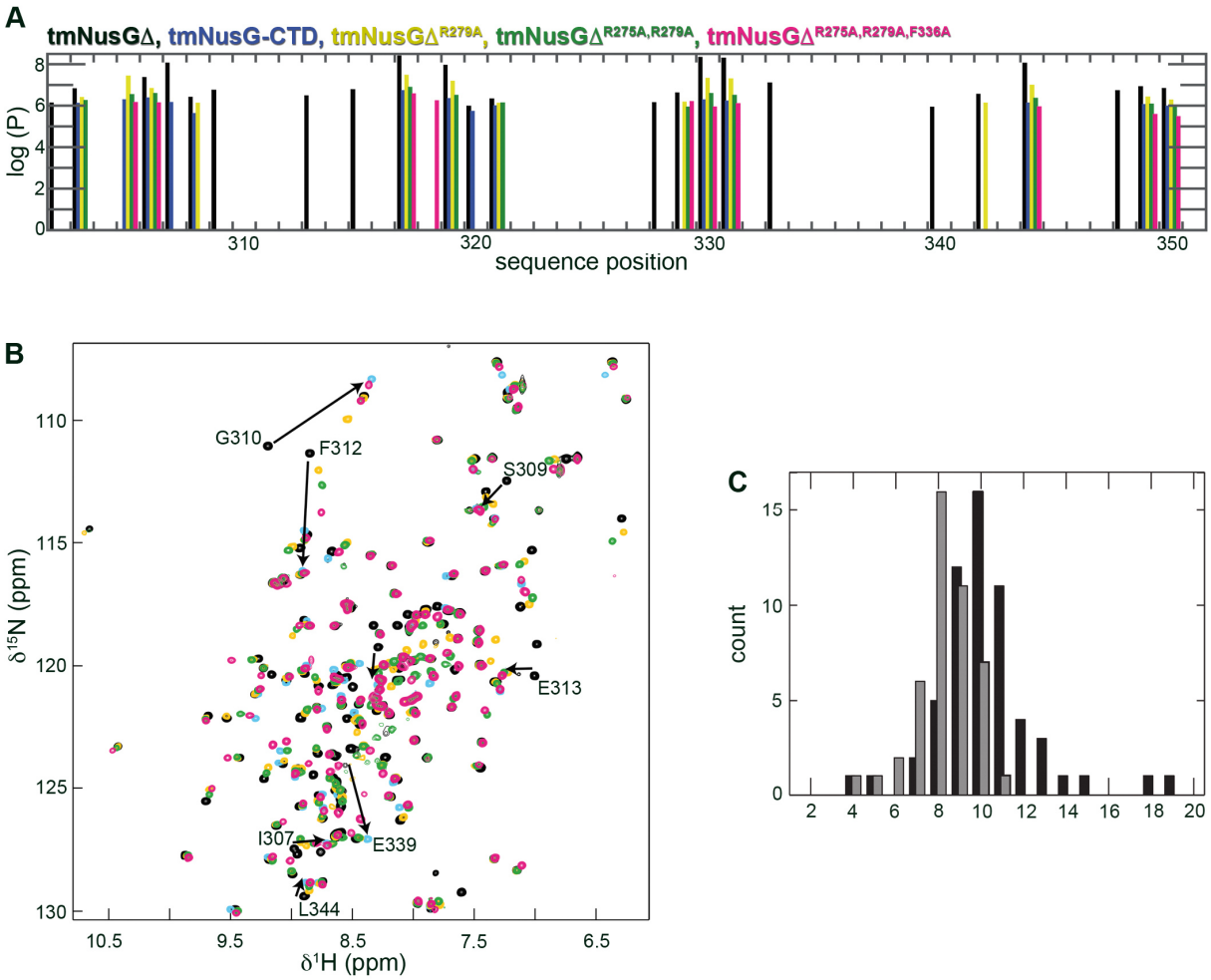


Figure S5: Weakening of the domain interaction in tmNusGΔ variants. (A) The logarithm of the protection factor (PF) of tmNusG-CTD determined by H/D exchange is mapped against its amino acid sequence. tmNusGΔ, black; isolated tmNusG-CTD, blue (1); tmNusGΔ^{R279A}, yellow; tmNusGΔ^{R275A,R279A}, green; tmNusGΔ^{R275A,R279A,F336A}, magenta. **(B)** [^{1}H,^{15}N]-HSQC spectra of ^{15}N-tmNusGΔ (100 μM), black, ^{15}N-tmNusG-CTD (150 μM), cyan, ^{15}N-tmNusGΔ^{R279A} (100 μM), yellow, ^{15}N-tmNusGΔ^{R275A,R279A} (100 μM), green, and ^{15}N-tmNusGΔ^{R275A,R279A,F336A} (420 μM), pink. Arrows indicate how tmNusG-CTD signals of tmNusGΔ shift towards the resonances of isolated tmNusG-CTD upon increased weakening of domain interaction. **(C)** Distribution of R_2/R_1 for the NTD, black, and the CTD, grey, in tmNusGΔ^{R275A,R279A,F336A}.

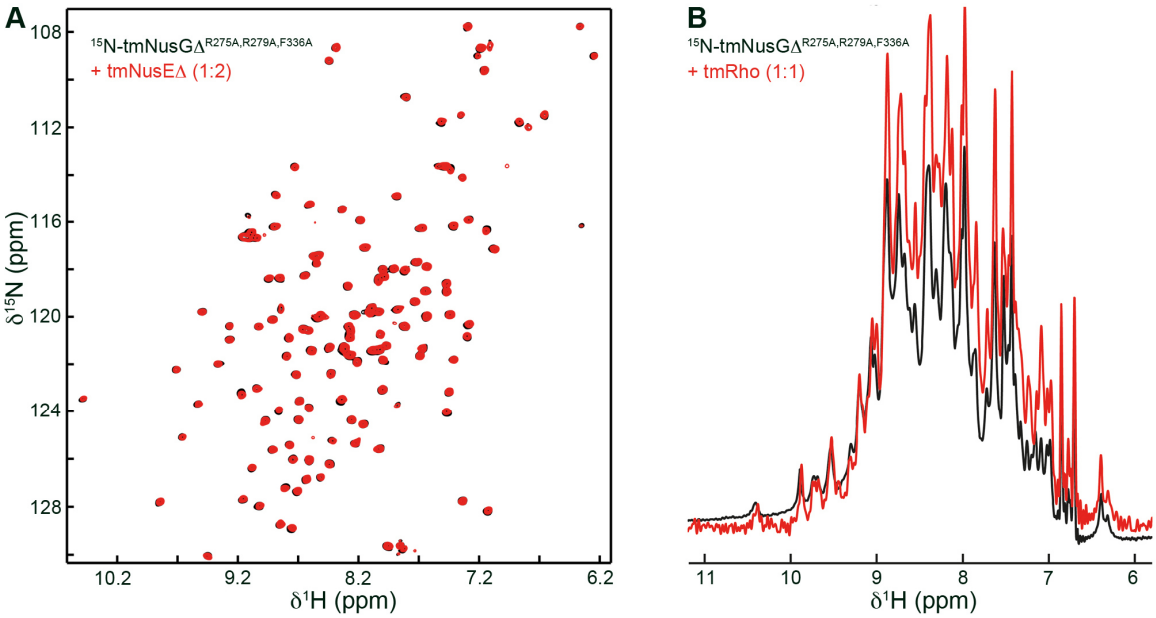


Figure S6: tmNusGΔ^{R275A,R279A,F336A} interacts neither with tmNusEΔ nor with tmRho. (A) 2D [^{1}H,^{15}N]-HSQC spectra of 150 μM ^{15}N-tmNusGΔ^{R275A,R279A,F336A}, black, and of 89 μM ^{15}N-tmNusGΔ^{R275A,R279A,F336A} after addition of tmNusEΔ in a twofold molar excess, red. **(B)** 1D [^{1}H,^{15}N]-HSQC spectra of 420 μM ^{15}N-tmNusGΔ^{R275A,R279A,F336A}, black, and of 30 μM ^{15}N-tmNusGΔ^{R275A,R279A,F336A} in the presence of tmRho in equimolar concentrations, red.

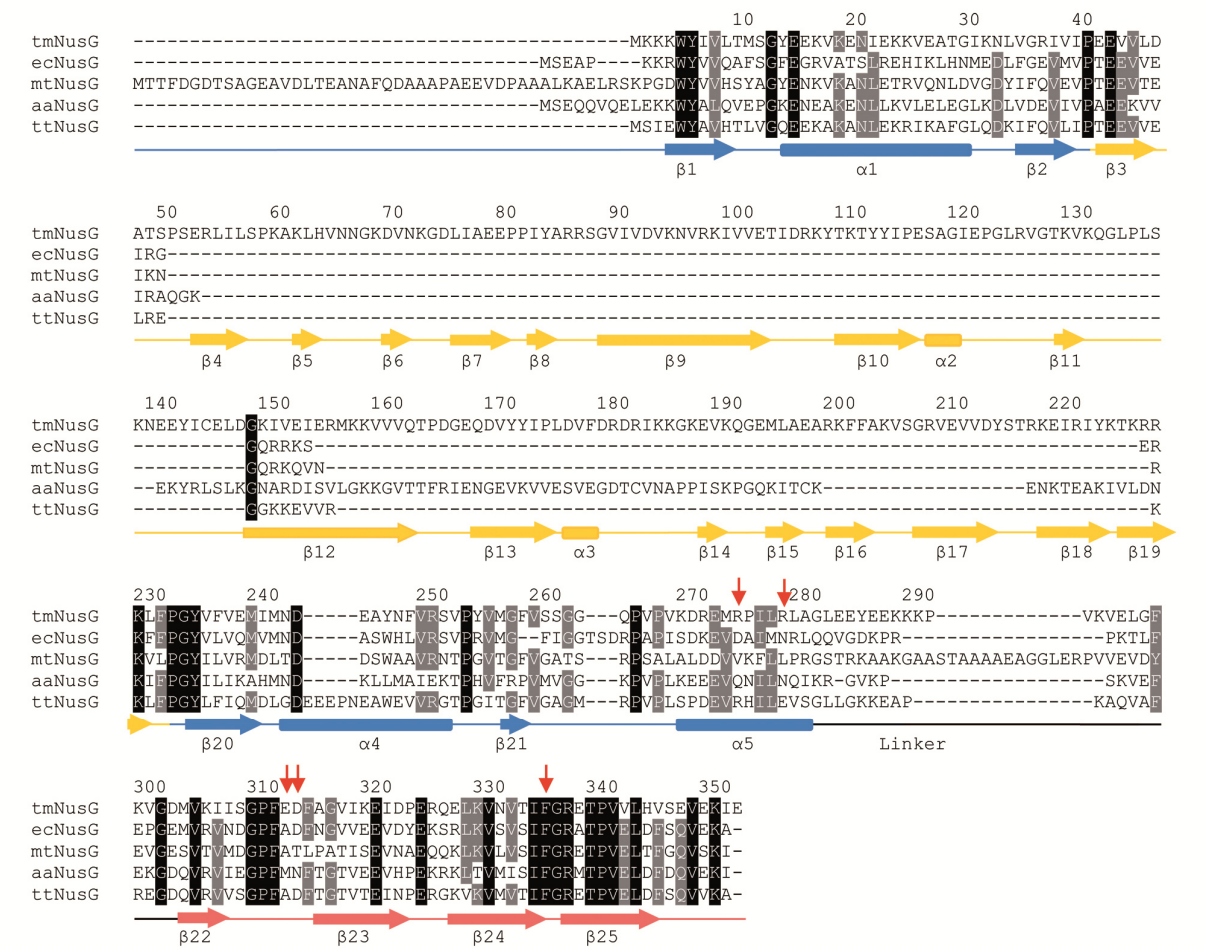


Figure S7: Sequence alignment of bacterial NusGs. Completely conserved residues are marked in black, 70 % conserved residues in grey. Secondary structure elements of tmNusG are indicated below the sequences. NTD, blue; DII, yellow; linker, black; CTD, red; α-helices, cylinders; β-strands, arrows. Numbering is according to tmNusG. Amino acids involved in the tmNusG interdomain salt bridges and Phe336 are marked by red arrows. Corresponding residues in NusG structures from other organisms are at the same positions as in tmNusG.

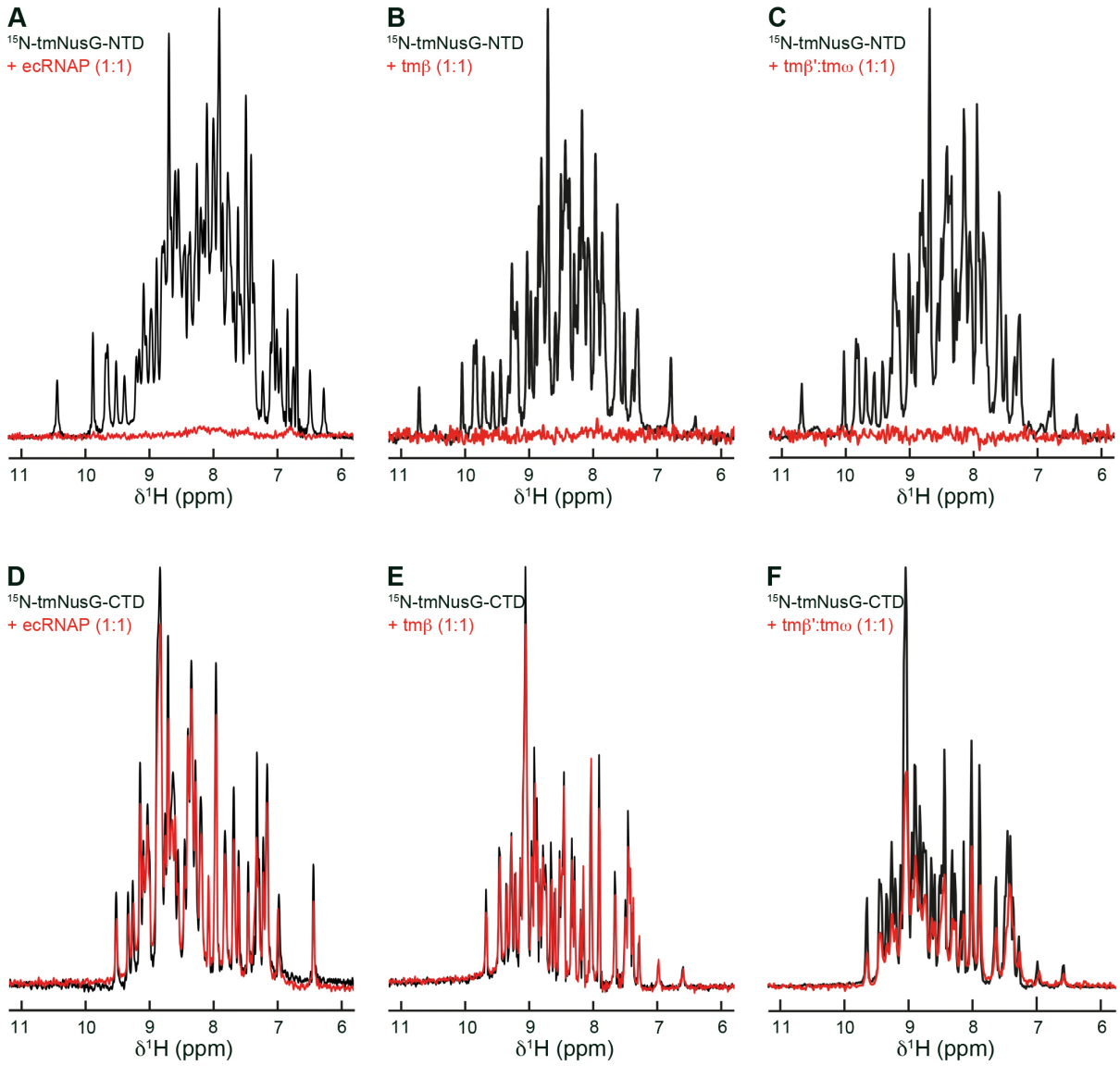


Figure S8: Interaction of tmNusG domains with ecRNAP and subunits of tmRNAP. (A-C) 1D [^{1}H,^{15}N]-HSQC spectra of 108 μM ^{15}N-tmNusG-NTD, black, and of ^{15}N-tmNusG-NTD in the presence of **(A)** ecRNAP (30 μM each), **(B)** tmβ (30 μM each), and **(C)** tmβ':tmω (27 μM each), red. **(D-F)** 1D [^{1}H,^{15}N]-HSQC spectra of 137 μM ^{15}N-tmNusG-CTD, black, and of ^{15}N-tmNusG-CTD in the presence of **(D)** ecRNAP (14 μM each), **(E)** tmβ (25 μM each), and **(F)** tmβ':tmω (27 μM each), red. Experiments with ecRNAP were carried out at 298 K, experiments with tmRNAP subunits at 323 K.

Supplementary References

1. Drögemüller,J., Stegmann,C.M., Mandal,A., Steiner,T., Burmann,B.M., Gottesman,M.E., Wöhrl,B.M., Rösch,P., Wahl,M.C. and Schweimer,K. (2013) An auto-inhibited state in the crystal structure of *Thermotoga maritima* NusG. *Structure,* **21,** 365-375.

8 Einzelarbeit B

Johanna Drögemüller*, Christian M. Stegmann*, Angshuman Mandal, Thomas Steiner, Björn M. Burmann, Max E. Gottesman, Birgitta M. Wöhrl, Paul Rösch, Markus C. Wahl und Kristian Schweimer (2013): An Autoinhibited State in the Structure of *Thermotoga maritima* NusG. *Structure* **21,** 365-375

* beide Autoren haben im gleichen Maße zur Arbeit beigetragen

Structure
Article

An Autoinhibited State in the Structure of *Thermotoga maritima* NusG

Johanna Drögemüller,[1,5] Christian M. Stegmann,[2,5,6] Angshuman Mandal,[2] Thomas Steiner,[3] Björn M. Burmann,[1,7] Max E. Gottesman,[4] Birgitta M. Wöhrl,[1] Paul Rösch,[1] Markus C. Wahl,[2,*] and Kristian Schweimer[1,*]
[1]Lehrstuhl Biopolymere und Forschungszentrum für Biomakromoleküle, Universität Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany
[2]AG Strukturbiochemie, Institut für Chemie und Biochemie, Freie Universität Berlin, Takustr. 6, D-14195 Berlin, Germany
[3]Max-Planck Institut für Biochemie, Abteilung Strukturforschung, Am Klopferspitz 18a, D-82152 Martinsried, Germany
[4]Departments of Microbiology and Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA
[5]These authors contributed equally to this work
[6]Present address: Bayer HealthCare, Global Drug Discovery, 13353 Berlin, Germany
[7]Present address: Biozentrum, University of Basel, Klingelbergstr. 70, 4056 Basel, Switzerland
*Correspondence: mwahl@chemie.fu-berlin.de (M.C.W.), kristian.schweimer@uni-bayreuth.de (K.S.)
http://dx.doi.org/10.1016/j.str.2012.12.015

SUMMARY

NusG is a conserved regulatory protein interacting with RNA polymerase (RNAP) and other proteins to form multicomponent complexes that modulate transcription. The crystal structure of *Thermotoga maritima* NusG (*Tm*NusG) shows a three-domain architecture, comprising well-conserved amino-terminal (NTD) and carboxy-terminal (CTD) domains with an additional, species-specific domain inserted into the NTD. NTD and CTD directly contact each other, occluding a surface of the NTD for binding to RNAP and a surface on the CTD interacting either with transcription termination factor Rho or transcription antitermination factor NusE. NMR spectroscopy confirmed the intramolecular NTD-CTD interaction up to the optimal growth temperature of *Thermotoga maritima*. The domain interaction involves a dynamic equilibrium between open and closed states and contributes significantly to the overall fold stability of the protein. Wild-type *Tm*NusG and deletion variants could not replace endogenous *Escherichia coli* NusG, suggesting that the NTD-CTD interaction of *Tm*NusG represents an autoinhibited state.

INTRODUCTION

Regulatory proteins are frequently composed of several domains, each of which typically sustains interactions to different binding partners (Pawson and Nash, 2000, 2003). Multiple interaction modules can form the basis for assembling and regulating multicomponent complexes. Some of these multidomain proteins are known to be regulated by an autoinhibitory mechanism, in which intramolecular interactions block binding sites for other proteins on the interacting domains (Burmann et al., 2012; Mackereth et al., 2011; Pufall and Graves, 2002). The release of autoinhibition requires large scale conformational changes that in turn critically depend on the specific intramolecular dynamics of the system (Li et al., 2008). Knowledge of thermodynamics and kinetics of autoinhibition release helps to understand the physical principles of affinity control. However, detection of lowly populated high-energy states is necessary, which often represents experimental challenges (Cho et al., 2011).

NusG is a general regulator of bacterial transcription that exerts diverse effects on RNA polymerase (RNAP) in a context-dependent manner. It is the only universally conserved transcription factor in all three domains of life (Werner, 2012). *Escherichia coli* (*Ec*) NusG increases the elongation rate of RNAP in vivo and in vitro (Artsimovitch and Landick, 2000; Burns and Richardson, 1995; Burova et al., 1995) by suppressing transcriptional pauses that involve backtracking of RNAP (Artsimovitch and Landick, 2000; Pasman and von Hippel, 2000). In conjunction with several other Nus-factors, NusG is part of antitermination complexes that resist pausing and termination (Torres et al., 2004; Zhou et al., 2002). These complexes are necessary for bacteriophage λ growth (Li et al., 1992; Sullivan et al., 1992) and for the efficient transcription of ribosomal RNA operons (Squires et al., 1993). At the same time, *Ec*NusG directly interacts with transcription termination factor Rho (Burmann et al., 2010; Pasman and von Hippel, 2000) and enhances Rho-dependent termination (Sullivan and Gottesman, 1992). *Ec*NusG also couples transcription and translation (Burmann et al., 2010).

*Ec*NusG consists of two domains (an amino-terminal domain, NTD; and a carboxy-terminal domain, CTD) connected by a flexible linker (Mooney et al., 2009). In some bacteria, an additional domain (domain II) is inserted in the NTD (Figure S1 available online) (Knowlton et al., 2003; Liao et al., 1996; Steiner et al., 2002) and a similar domain expansion is found in the functional NusG analog, Spt5, of archaea and eukaryotes (Guo et al., 2008; Wenzel et al., 2010). The NusG NTD consists of a four-stranded antiparallel β sheet flanked by two antiparallel helices on one side and by a third, C-terminal helix on the other side. The CTD forms a barrel-type antiparallel β sheet with an embedded KOW motif (Kyrpides et al., 1996), a known RNA binding element.

The multidomain architecture of NusG enables it to act as an adaptor protein that links other proteins bound to these domains in multicomponent complexes. A crystal structure of the NusG-like NTD (NGN) of transcription factor Spt5 from *Pyrococcus furiosus* (*Pf*) in complex with transcription factor Spt4 and the clamp domain of RNAP suggested that a large hydrophobic

Table 1. Crystallographic Data

Data Collection	Full Length	Domain II (Native)	Domain II (SeMet)
Wavelength (Å)	1.05	0.97985	0.97893
Temperature (K)	100	100	100
Space group	$P4_322$	C2	C2
Unit cell parameters			
a, b, c (Å)	60.0, 60.0, 311.5	136.3, 42.2, 88.3	134.9, 42.0, 89.1
α, β, γ (°)	90, 90, 90	90, 106.3, 90	90, 106.3, 90
Resolution (Å)[a]	50.0–2.4	50.0–1.9	50.0–2.0
	(2.44–2.40)	(1.97–1.90)	(2.07–2.00)
Reflections			
Unique	23,437 (1,082)	37,677 (1,846)	29,850 (1,495)
Completeness (%)	99.2 (99.6)	99.9 (72.7)	89.9 (45.7)
Redundancy	4.9 (4.7)	3.8 (3.7)	6.2 (3.0)
Mean I/σ(I)	16.6 (2.0)	13.3 (1.8)	14.6 (2.1)
$R_{sym}(I)$[b]	7.4 (56.7)	9.4 (78.8)	8.6 (39.2)
Phasing			
Resolution (Å)			35.0–2.0
Heavy atom sites			4
SHELXD CC/CC_{weak}[c]			56.5 / 26.9
SHARP overall figure of merit acentric / centric			0.29 / 0.22
Refinement			
Resolution (Å)	50.0–2.4	50.0–1.9	
Reflections			
Number	23,437	35,358	
Completeness (%)	95.4	92.7	
Test set (%)	5	5	
R_{work}[d] (%)	22.0	18.1	
R_{free}[d] (%)	27.4	23.0	
ESU (Å)	0.31	0.25	
Contents of A.U.			
Protein molecules/ residues/atoms	1/341/ 2,763	2/386/ 3,220	
Water oxygens	27	204	
Acetate ions	–	6	
Mean B-factors (Å^2)			
Wilson	56.1	22.5	
Protein	76.0	33.4	
Water	60.1	40.4	
Ramachandran plot[e]			
Preferred (%)	94.1	99.2	
Allowed (%)	4.1	0.8	
Disallowed (%)	1.8	0	
RMSD from target geometry			
Bond lengths (Å)	0.006	0.009	
Bond angles (°)	1.036	1.140	
Chirality (Å)	0.067	0.087	
Dihedral angles (°)	16.01	13.31	
PDB ID	2XHC	2XHA	

[a]Data for the highest resolution shell is in parentheses.
[b]$R_{sym}(I) = \Sigma_{hkl}\Sigma_i|I_i(hkl) - <I(hkl)>| / \Sigma_{hkl}\Sigma_i|I_i(hkl)|$, for *n* independent reflections and *i* observations of a given reflection; where $<I(hkl)>$ = average intensity of the *i* observations.
[c]$CC = [\Sigma wE_oE_c\Sigma w - \Sigma wE_o\Sigma wE_c]/\{[\Sigma wE_o^2\Sigma_w - (\Sigma wE_o)^2]\ [\Sigma wE_c^2\Sigma w - (\Sigma wE_c)^2]\}^{1/2}$, where *w* = weight (see http://shelx.uni-ac.gwdg.de/SHELX/shelx_de.pdf for full definitions).
[d]$R = \Sigma_{hkl}||F_{obs}| - |F_{calc}|| / \Sigma_{hkl}|F_{obs}|$. A.U., asymmetric unit; ESU, estimated overall coordinate error based on maximum likelihood; R_{work}, hkl ∉ T; R_{free}, hkl ∈ T; R_{all}, all reflections; T, test set.
[e]According to Lovell et al. (2003).

patch between the C-terminal helix, and the β sheet of the NusG NTD represents a binding region for the conserved clamp region of RNAP (Martinez-Rucobo et al., 2011). Furthermore, NMR analyses have demonstrated that the CTD of *Ec*NusG forms mutually exclusive complexes with Rho and NusE (equivalent to ribosomal protein S10) (Burmann et al., 2010). These multiple, domain-wise interactions of NusG allow it to integrate the activities of the transcriptional and translational machineries in *E. coli* (Burmann et al., 2010).

Here, we present evidence that in some organisms NusG might be regulated by autoinhibition. The crystal structure of NusG from the marine hyperthermophilic bacterium *Thermotoga maritima* (*Tm*NusG) and the solution structure of a deletion variant thereof reveal a direct interaction between the conserved NTD and CTD that is incompatible with other known protein contacts of these domains. In addition, thermodynamic parameters as well as interconversion dynamics between open and closed states determined by NMR spectroscopy point to an autoregulatory mechanism.

RESULTS AND DISCUSSION

Crystal Structure Analysis of *Tm*NusG

Based on sequence analysis, *Tm*NusG possesses an insertion (domain II) in its NTD not found in most other bacteria (Liao et al., 1996; Steiner et al., 2002). To investigate the fold and domain architecture of *Tm*NusG, we crystallized the full-length recombinant protein produced in *E. coli*. Because molecular replacement with known NusG NTD and CTD structures as the sole search models failed, and well-diffracting crystals of full-length *Tm*NusG were poorly reproducible, we separately produced and crystallized the *Thermotoga*-specific domain II (residues 42–233) and solved its structure by selenomethionine (SeMet) single anomalous dispersion (SAD) at 1.9 Å resolution (Table 1). The structure of domain II together with known NusG NTD and CTD structures allowed us to solve the crystal structure of the full-length protein by molecular replacement at 2.4 Å resolution. Both structures were refined to acceptable R-factors with good stereochemistry (Table 1).

Domain II (*Tm*NusG^{DII}) crystallized in space group C2 with two molecules per asymmetric unit, which are structurally very

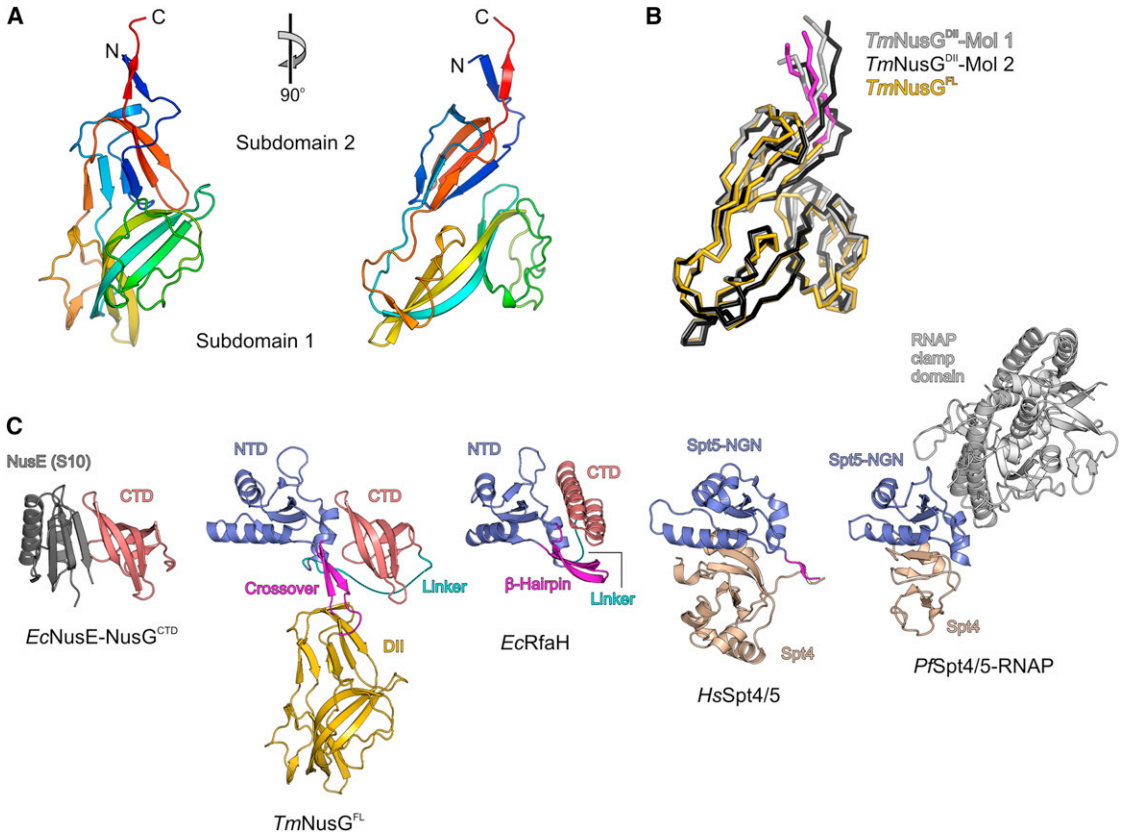


Figure 1. Crystal Structure of *Tm*NusG

(A) Orthogonal ribbon plots of the *Thermotoga*-specific domain II insertion comprising two subdomains. The protein is colored blue to red from N terminus to C terminus to illustrate the chain trace.

(B) Superimposed ribbon plots of the two crystallographically independent molecules of the *Tm*NusG^{DII} crystal structure (light and dark gray) and of the domain II portion of the full-length protein (domain II, gold; crossovers to the NTD, magenta). Minor differences are seen in surface loops and at the domain termini.

(C) Overall structure of *Tm*NusG^{FL} (panel 2) in comparison with an *Ec*NusE-NusG^{CTD} complex (panel 1), *Ec*RfaH (panel 3), *Hs*Spt4-Spt5 complex (panel 4), and *Pf*Spt4-Spt5-RNAP clamp domain complex (panel 5). The NusG CTD (panels 1 and 2) or the equivalents of the NusG NTD (panels 2–5) are shown in the same orientation as the corresponding domains in the structure of *Tm*NusG (panel 2).The orientation of *Tm*NusG^{FL} is the same as for *Tm*NusG^{DII} in the left panel of (A). NTD, blue; domain II, gold; CTD, light red; NusE, dark gray; Spt4, beige; RNAP, gray; NTD-CTD linker, cyan; NTD-domain II crossovers or topologically equivalent regions, magenta.

See also Figure S1.

similar (root mean square deviation [RMSD] 1.37 Å for 187 superimposed Cα atoms). The two molecules differed slightly in some surface loops and at the domain termini (Figures 1A and 1B). All *Tm*NusG residues contained in the domain II construct could be traced for both independent copies of the molecule in the electron density. The regularly structured part of domain II encompasses residues 52–224 and is composed of two subdomains (Figure 1A). The composite N-terminal subdomain (residues 52–80 and 200–224) adopts a β sandwich fold made up of two antiparallel four- and three-stranded β sheets. The continuous C-terminal subdomain (residues 83–197) exhibits a dumbbell-like structure whose central scaffold is formed by a four-stranded antiparallel β sheet. Short, antiparallel peptides (residues 81–82 and 198–199) connect the subdomains. Comparison with entries in the Protein Data Bank (PDB; http://www.pdb.org) showed that the overall fold of domain II is unique with respect to the spatial organization and the connectivity of the secondary structure elements. PDBeFold/SSM (Krissinel and Henrick, 2004) yielded no matches using the default threshold value (lowest acceptable match 70% for query and target). Lowering the threshold value did not yield hits with RMSDs below 4 Å or Q scores above 0.06, confirming the fold.

Crystals of full-length *Tm*NusG (*Tm*NusG^{FL}) belonged to space group $P4_322$ with one molecule per asymmetric unit. The crystal structure of *Tm*NusG^{FL} exhibits a "T"-shape (Figure 1C, panel 2), in which the NTD (residues 3–40 and 234–281) and CTD (residues 299–352) form the bar of the "T" while the stem is formed by domain II that is inserted within the NTD. Both NTD and CTD adopt very similar folds as in other NusG proteins (compared to *Aquifex aeolicus* [*Aa*] NusG, PDB ID 1M1G: RMSD NTD 1.65 Å for 84 superimposed Cα common atoms, CTD 0.77 Å for 53 superimposed Cα atoms). The NTD comprises an antiparallel four-stranded β sheet flanked by two α helices on one side and an additional α helix on the other. The CTD consists of a five-stranded, antiparallel β barrel. The two terminal domains are connected by a long, flexible linker

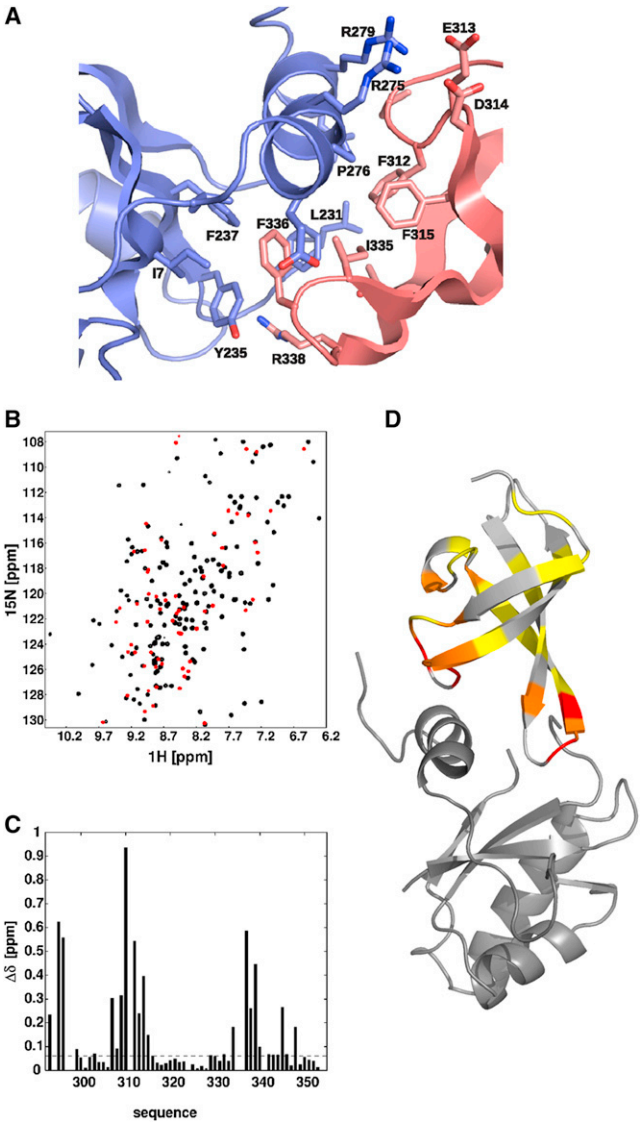


Figure 2. Domain Interaction of *Tm*NusG

(A) Expanded view of the domain interface between NTD (blue) and CTD (light red). Side chains with buried surfaces compared to isolated domains are shown in stick representation.

(B) Superposition of ^{1}H, ^{15}N HSQC spectra of *Tm*NusG$^{NTD-CTD}$ (black) and *Tm*NusG^{CTD} (red).

(C) Normalized chemical shift differences of backbone amide resonances between *Tm*NusG$^{NTD-CTD}$ and *Tm*NusG^{CTD} as a function of sequence.

(D) Mapping of chemical shift differences onto the crystal structure of *Tm*NusG. The inserted domain II is omitted for clarity. Color coding from yellow to red indicates increasing chemical shift differences.

(residues 282–298) whose central part (residues 288–296) lacked well-defined electron density (dashed cyan line in Figure 1C, panel 2).

In *Tm*NusG^{FL}, domain II adopts the same structure as in the two copies of the isolated domain construct (RMSD 1.20/1.66 Å for 184/179 superimposed Cα atoms; Figure 1B). It is connected to the N- and C-terminal portions of the NTD via two peptides (residues 41–51 and 225–233) arranged as two antiparallel β strands (Figure 1C, panel 2). The bulk of domain II lacks direct contacts to either NTD or CTD. Previously, an additional domain has also been found inserted at the corresponding site in the NTD of *Aa*NusG (Knowlton et al., 2003; Steiner et al., 2002). Although the *Aa*NusG insertion exhibits a β sandwich fold reminiscent of the first subdomain of *Tm*NusG domain II, it folds with a different topology. In *Ec*NusG, domain II is replaced by an elongated loop (Burmann et al., 2011; Mooney et al., 2009).

NTD-CTD Contacts in *Tm*NusG Are Reminiscent of Interdomain Interactions in the NusG paralog, RfaH

In the crystal structure of *Tm*NusG, the NTD and CTD directly contact each other, burying 1544 Å^2 of combined surface area at their interface (contribution of NTD and CTD 686 Å^2 and 858 Å^2, respectively). In the NTD, residues from strands β1 (Ile7), β3 (Tyr235 and Phe237), and its preceding loop (Leu231, Phe232, and Pro233), as well as residues from helix α3 (Pro276, Leu280) and its preceding loop (Val269) form a hydrophobic surface patch (Figure 2A). The CTD associates with this surface patch via two loops, between strands β1′ and β2′ (Pro311 and Phe312) and between strands β3′ and β4′ (Ile335 and Phe336), forming extensive, hydrophobic interdomain contacts. In addition, two salt bridges (Arg275-Asp314, and Arg279-Glu313) and a hydrogen bond (between the side chain of Arg338 and backbone of Pro233) are sustained between the domains (Figure 2).

In *E. coli*, the specialized transcription factor RfaH, a paralog of NusG, exhibits a NusG-like NTD and a differently folded, α-helical CTD that are tightly associated (Belogurov et al., 2007). The NTD-CTD interactions in *Tm*NusG and RfaH make use of an equivalent hydrophobic patch on the respective NTDs (Figure 1C, panel 3). In contrast to *Tm*NusG, RfaH-CTD adopts a helical conformation in the closed state. Upon domain opening, RfaH-CTD refolds into the all-β sheet conformation similar to NusG-CTD (Burmann et al., 2012). The domain interaction in RfaH has important physiologic consequences. RfaH requires a specific DNA sequence, *ops*, to unlock its domains and to use the hydrophobic patch on the liberated NTD for binding to RNAP (Belogurov et al., 2007). Thus, the tight domain interaction in RfaH represents an autoinhibited state that restricts the use of the transcription factor to genes bearing an *ops* sequence. These observations suggest that similar autoinhibitory mechanisms may be exploited to regulate the functions of *Tm*NusG. The requirement of a certain factor for opening of *Tm*NusG is unknown. It might also be possible that the affinity to RNAP is sufficient to open *Tm*NusG without an additional factor.

NTD and CTD Contacts Are Mutually Exclusive with NusG-RNAP, NusG-NusE, and NusG-Rho Interactions

The surfaces of the NTD and the CTD involved in intramolecular interaction are congruent with regions of these domains known to bind other protein partners (Figure 1C). In the complex of the *Ec*NusG CTD with NusE (Burmann et al., 2010) the two loops of the CTD that sustain interdomain contacts in *Tm*NusG are involved in the protein-protein interaction (Figure 1C, panel 1). Moreover, the identical CTD surface contacts transcription termination factor Rho (Burmann et al., 2010).

While presently no structure is available of bacterial NusG in complex with RNAP, the high structural similarity of the NusG NTD with the Spt5 NGN domain of archaea and eukaryotes

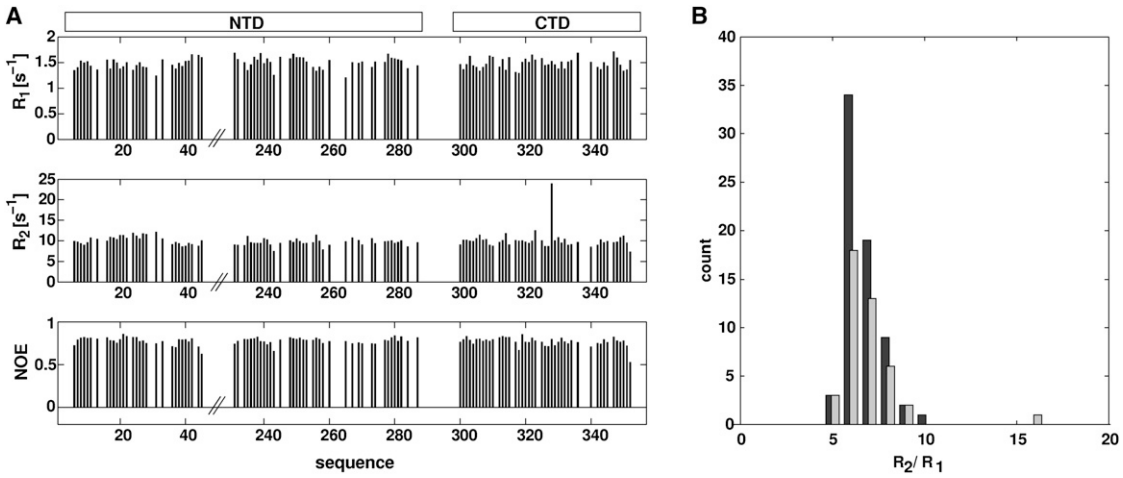


Figure 3. ^{15}N Spin Relaxation Data of *Tm*NusG$^{NTD-CTD}$ at 14.1T Magnetic Field Strength and 323 K

(A) Longitudinal (R_1, top), transverse (R_2, middle) relaxation rates, and [^{1}H]^{15}N steady state NOE (bottom) as a function of amino acid sequence.

(B) Distributions of R_2/R_1 ratios. Data from residues of the NTD are shown in black, data from residues of the CTD are in light gray. The similar distributions of both domains indicate a uniform overall tumbling of the protein.

See also Table S1.

(Figure 1C, panels 4 and 5) allows the prediction of the RNAP-binding site on NusG based on the crystal structure of an archaeal Spt4-Spt5-RNAP clamp domain complex (Martinez-Rucobo et al., 2011). Again, the region of the NusG NTD expected to bind the clamp domain of RNAP partly overlaps with the area of the NTD involved in interdomain contacts in *Tm*NusG (Figure 1C, panel 5). These observations suggest that intramolecular domain interactions observed in *Tm*NusG would interfere with other functional interactions of the NTD and CTD, further supporting the notion that the present conformation of *Tm*NusG represents an autoinhibited state.

NTD-CTD Contacts Are Maintained in Solution

Previous structural analyses of other NusG orthologs have failed to disclose stable NTD-CTD interactions (Burmann et al., 2011; Mooney et al., 2009; Reay et al., 2004; Steiner et al., 2002). Interestingly, however, regions identical to the NTD-CTD contact areas in *Tm*NusG are involved in lattice contacts between neighboring molecules in crystal structures of *Aa*NusG (Knowlton et al., 2003; Steiner et al., 2002) and have been suggested to reflect a functional intramolecular interaction in solution (Knowlton et al., 2003). We therefore asked whether the NTD-CTD interactions observed in the *Tm*NusG crystal structure prevail in solution. To address these questions using NMR spectroscopy, we generated a shorted variant of *Tm*NusG, *Tm*NusG$^{NTD-CTD}$, in which we replaced a region (residues 43–230) encompassing domain II with the corresponding loop of *Ec*NusG (residues 52–61). The replaced region did not exhibit any direct interactions with the NTD or CTD in the crystal structure.

The ^{1}H, ^{15}N HSQC spectrum of *Tm*NusG$^{NTD-CTD}$ showed the characteristic signal dispersion of well-folded globular proteins. Almost all resonances could be assigned by applying standard heteronuclear through-bond correlations. Secondary chemical shifts correlated well with the crystal structure. Expression of the isolated CTD also resulted in a well-folded globular protein whose ^{1}H, ^{15}N correlations were assigned using 3D ^{15}N-edited NOESY and ^{15}N-edited TOCSY spectra. However, the chemical shifts of the isolated CTD displayed remarkable differences in their magnitudes compared to the spectrum of *Tm*NusG$^{NTD-CTD}$ (Figures 2B–2D). Such differences are typically found upon complex formation of protein complexes. Mapping chemical shift changes onto the structure revealed that both loops of the CTD participating in the interaction with the NTD in the crystal structure of *Tm*NusG are strongly affected by the presence of the NTD in solution. *Ec*NusG, in contrast, did not show any chemical shift differences between isolated domains and the full-length protein (Burmann et al., 2011). These observations indicate that NTD and CTD of *Tm*NusG$^{NTD-CTD}$ mutually interact in solution, presumably in a similar manner as observed in the *Tm*NusG crystal structure.

Spin Relaxation Experiments Support Similar NTD-CTD Interactions in Solution as in the Crystal

To further confirm the similar association of NTD and CTD in *Tm*NusG in solution as in the crystal, we conducted spin relaxation experiments, a powerful method to determine the overall tumbling of proteins in solution. Defined domain interactions in a multidomain protein result in a uniform rotational tumbling of the entire protein. In this case, the stochastic rotation can be described in a single frame represented by a unique rotational diffusion tensor for both domains. In the case of a multidomain protein with noninteracting domains, the relative domain movement requires an individual description of the rotational tumbling for each domain. The ratio of the transverse (R_2) and the longitudinal (R_1) relaxation rates is a key parameter for characterizing overall tumbling of a protein because contributions of fast internal dynamics are cancelled out to a large extent. The R_2/R_1 ratio of *Tm*NusG$^{NTD-CTD}$ showed a mono-modal distribution (Figure 3) with an effective isotropic correlation time of 7.4 ns at 50°C, indicating coupling of the two domains on the timescale of the molecular rotation. The situation is different in *Ec*NusG, which displayed different R_2/R_1 distributions for its NTD and

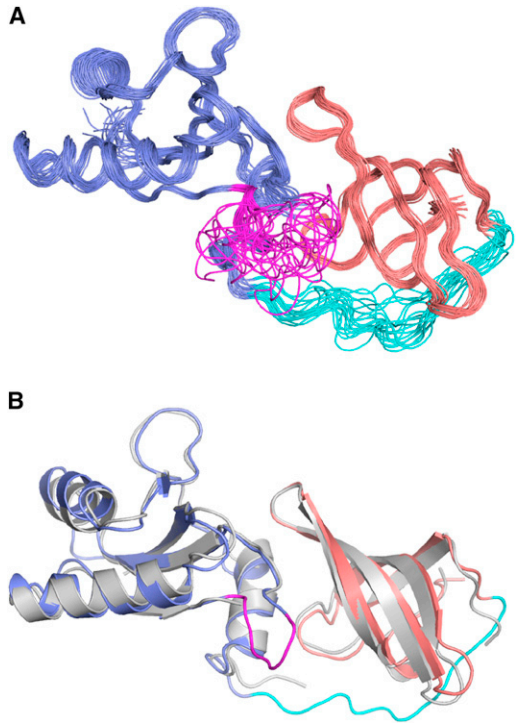


Figure 4. Solution Structure of *Tm*NusG$^{NTD-CTD}$
(A) Superposition of 20 accepted structures.
(B) Comparison of the crystal structure (gray) with the solution structure. The inserted domain II of the crystal structure is omitted for clarity. NTD, blue; CTD, light red; NTD-CTD linker, cyan; NTD-domain II crossover, magenta.

Table 2. Solution Structure Statistics

Experimental Derived Restraints	
Distance restraints	
NOE	1,908
Intraresidual	679
Sequential	480
Medium range	224
Long range	525
Hydrogen bonds	73
Dihedral restraints	302
Residual dipolar couplings	173
Restraint Violation	
Average distance restraint violation (Å)	0.007 ± 0.001
Maximum distance restraint violation (Å)	0.18
Average dihedral restraint violation (°)	0.93 ± 0.2
Maximum dihedral restraint violation (°)	14
Average rdc restraint violation (Hz)	0.59 ± 0.05
Maximum rdc restraint violation (Hz)	3.2
Deviation from Ideal Geometry	
Bond length (Å)	0.00075 ± 0.00005
Bond angle (°)	0.16 ± 0.008
Coordinate Precision[a,b]	
Backbone heavy atoms (Å)	0.49
All heavy atoms (Å)	0.99
Ramachandran Plot Statistics[c] (%)	84.1/12.4/2.5/1.0

[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 Lys5-Pro43, Lys230-Glu287, Phe299-Ile352 (numbering according to full-length *tm*NusG).
[c]Ramachandran plot statistics are determined by PROCHECK and noted by most favored/additionally allowed/generously allowed/disallowed.

CTD, consistent with domain decoupling in that protein (Burmann et al., 2011).

The overall shape of a protein is reflected in the relaxation rates of its nuclear spins due to the orientation of the corresponding inter-nuclear vectors relative to the principal axis of the rotational diffusion tensor. We derived a rotational diffusion tensor of *Tm*NusG$^{NTD-CTD}$ using the ^{15}N relaxation data (Table S1) from *Tm*NusG$^{NTD-CTD}$ and the coordinates of full-length *Tm*NusG. The tumbling of *Tm*NusG$^{NTD-CTD}$ was very well described by a prolate, axial symmetric diffusion tensor with a tumbling time of 7.44 ns and an axial ratio of 1.42. These observations showed that both domains tumble as a rigid unit with an overall elongated shape, exhibiting a stable domain interaction at least on the time scale of the molecular rotation. The derived values perfectly fit the coordinates of NTD and CTD of the *Tm*NusG crystal structure, indicating that *Tm*NusG$^{NTD-CTD}$ in solution adopts an identical overall structure as in the crystal. In particular, NTD and CTD strongly interact in solution.

We directly verified the above conclusions by determining the solution structure of *Tm*NusG$^{NTD-CTD}$, using interdomain NOEs detected in isotope-edited NOESY spectra, and residual dipolar coupling, determined in two liquid crystalline media (pf1 phages [5 mg/ml] and 3% C6E12/Hexanol) (Figure 4\; Table 2). The resulting structural ensemble shows good coordinate precision and reasonable stereochemical properties. The average solution structure superimposes very well with the crystal structure (backbone RMSD of 1.1 Å; Figure 5B).

NTD-CTD Contacts Persist at High Temperatures

The analyses described above demonstrate that in solution *Tm*NusG$^{NTD-CTD}$ exhibits a defined domain interaction at least at 50°C, a low growth temperature for *Thermotoga maritima*. The solution structure closely resembles the domain interaction in the crystal. To test whether this domain interaction also persists at temperatures closer to the optimal growth conditions for *Thermotoga maritima*, a series of HSQC spectra with increasing sample temperatures up to 80°C were recorded. The prominent chemical shift differences between *Tm*NusG$^{NTD-CTD}$ and *Tm*NusG^{CTD} were still observed at 80°C, revealing the presence of intramolecular domain interactions even at this temperature. However, sample stability was significantly reduced under these conditions and the protein precipitated after roughly 2 hours.

Contributions of the Interdomain Interaction to Protein Stability

The optimal growth temperature of 84°C for *Thermotoga maritima* presents particular challenges for the integrity of protein folds. To estimate the contribution of the domain

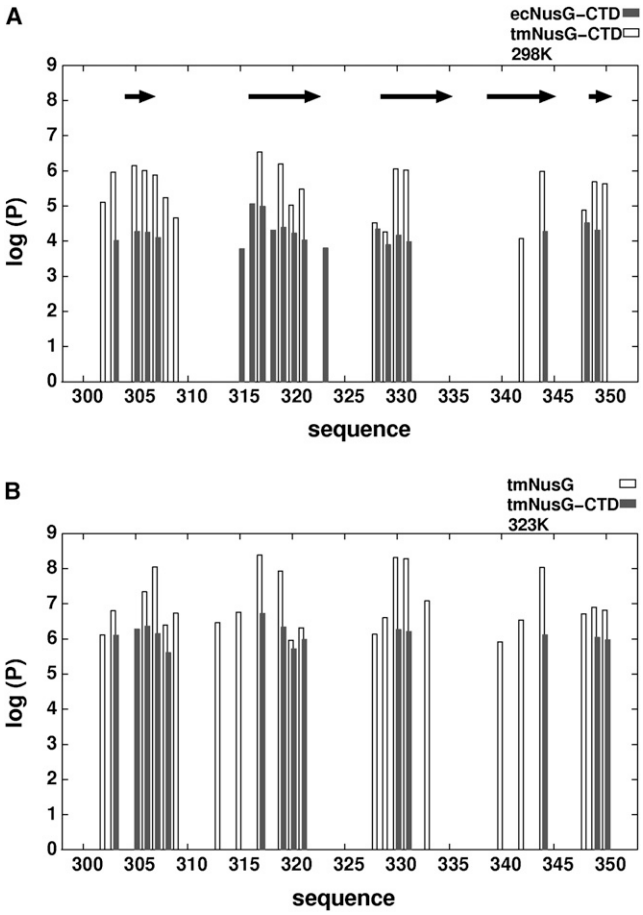


Figure 5. Amide PFs Determined by H/D Exchange

(A) PFs of *Tm*NusG^{CTD} (dark gray) and *Ec*NusG^{CTD} (white) at 298 K as a function of sequence. PFs are expressed as the ratio of the intrinsic exchange rate and the experimentally determined exchange rate. Residues Phe299-Ile352 of *Tm*NusG correspond to residues Phe128-Ala181 of *Ec*NusG. The arrows indicate the location of the β strands.

(B) PFs of the CTD from *Tm*NusG$^{NTD-CTD}$ (white) and isolated *Tm*NusG^{CTD} at 323K.

interaction to protein fold stability, we conducted hydrogen/deuterium (H/D) exchange experiments followed by NMR spectroscopy of *Tm*NusG$^{NTD-CTD}$, an isolated CTD of *Tm*NusG (*Tm*NusG^{CTD}), and an isolated CTD of *Ec*NusG (*Ec*NusG^{CTD}). H/D exchange was monitored via the decay of signal intensities in a series of ^{1}H, ^{15}N HSQC experiments recorded at different time points after dissolving lyophilized proteins in D_2O at 298 K or 323 K.

On average, the exchange rates of *Tm*NusG$^{NTD-CTD}$ and *Tm*NusG^{CTD} were 30 times faster at 323 K than at 298 K. At 323 K, the H/D exchange of *Ec*NusG^{CTD} was too fast to detect, although the protein still showed the well-dispersed ^{1}H, ^{15}N HSQC spectrum of a folded domain. At 298 K, the exchange rates of *Ec*NusG^{CTD} were 65 times faster than for *Tm*NusG^{CTD}. All slowly exchanging amide protons were found in regular secondary structure elements, and the patterns of exchanging amide protons were similar for all three proteins, suggesting similar opening dynamics for all three constructs.

To characterize the stabilization effect due to NTD-CTD interaction, the amide proton exchange was expressed as Protection Factor (PF; Figure 5). From the PF, the free energy of hydrogen exchange was calculated and the highest ΔG_{HX} was assumed to be equal to the free energy of folding. For *Ec*NusG^{CTD}, the maximum PF (Asn145) was 8.8×10^3 which translates into a ΔG_{HX}(298 K) of 28.8 kJ/mol. For *Tm*NusG^{CTD}, the maximum PF (Gly317) was 3.4×10^6 corresponding to a ΔG_{HX}(298 K) of 37.3 kJ/mol. At 323 K, the maximum PF (Gly317) of *Tm*NusG^{CTD} was 5.2×10^6 corresponding to a ΔG_{HX}(323 K) of 41.5 kJ/mol. ΔG_{HX} of *Tm*NusG^{CTD} is lower at 298 K than at 323 K, indicating that the measurements were performed below the optimal temperature for stabilization, which is expected to be close to the optimal growth temperature of *Thermotoga maritima* (84°C). Thus, compared to the mesophilic *Ec*NusG, at 298 K the thermophilic *Tm*NusG^{CTD} is stabilized by about 8.5 kJ/mol.

Comparison of the exchange rates of amide protons of *Tm*NusG$^{NTD-CTD}$ and of *Tm*NusG^{CTD} shows that the protection is increased 40- to 50-fold (PF = 2.4×10^8, ΔG_{HX}(323 K) = 51.8 kJ/mol for Gly317 in *Tm*NusG$^{NTD-CTD}$). This corresponds to an additional stabilization of 10.3 kJ/mol, which is due to the presence of the NTD. This additional stabilization can be attributed to the interaction between NTD and CTD. Converting the portion of the free energy of folding that is caused by the domain interaction into the equilibrium between open and closed states yielded a fraction of 2% of the *Tm*NusG$^{NTD-CTD}$ molecules in the open conformation at 323 K. As a consequence, the closing rate is about 50 times faster than the opening rate.

Dynamics of the Interdomain Interactions

To test the presence of microsecond-millisecond dynamics in the domain interface, ^{15}N relaxation-dispersion experiments were performed with *Tm*NusG$^{NTD-CTD}$. With this method, we did not observe any contribution of chemical exchange to the transverse relaxation rates of nitrogen spins for any of the residues of the CTD that are located at the domain interface and exhibited significant chemical shift differences between *Tm*NusG$^{NTD-CTD}$ and *Tm*NusG^{CTD}. For two residues of the CTD (Asp323 and Glu328), relaxation-dispersion curves showed a contribution of chemical exchange to the transverse relaxation (Figure 6). These residues are located face to face at the end of two β strands, far away from the domain interface, and their observed dynamics reflect local conformational flexibility. Therefore, the exchange rate for domain opening/closing must either be in the fast exchange regime with $k_{ex} \gg 2\pi\Delta\nu_{max}$ ($\Delta\nu_{max}$ being the largest ^{15}N resonance frequency change, ca. 370 Hz for Phe312) or, less likely, in the slow exchange limit for all residues ($k_{ex} < 1\ s^{-1}$). The slow exchange situation would result in observable signals for the open state in a ^{1}H, ^{15}N HSQC spectrum with sufficient signal/noise ratio. In a spectrum with signal/noise > 200:1 for characteristic residues in the domain interface, no signals of the open state could be observed.

The exchange contribution in *Tm*NusG$^{NTD-CTD}$ is determined by the rate for closing ($k_{ex} = k_{open} + k_{close} = (1/50 + 1)\ k_{close}$, due to $K_D = k_{open}/k_{close} = 1{:}50$). In a titration using separate domains, the exchange rate is given by $k_{ex} = k_{off}(1+p_{EL}/p_E)$ with a dissociation constant $K_{D,binary} = k_{on}/k_{off}$, in which p_{EL}, p_E are the molar fractions of bound (p_{EL}) and free (p_E) ^{15}N-labeled

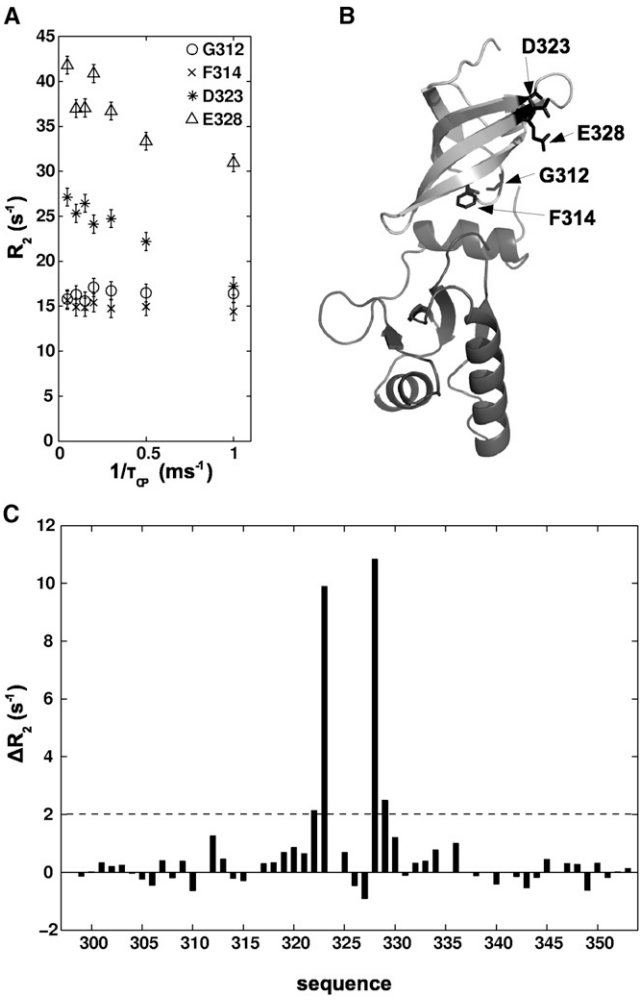


Figure 6. ^{15}N-Relaxation-Dispersion of *Tm*NusG$^{NTD-CTD}$

(A) Selected ^{15}N-relaxation-dispersion curves of residues from the domain interface (Gly312 and Phe314), and Asp323 and Glu328.

(B) Structure of *Tm*NusG$^{NTD-CTD}$. Gly312, Phe314, Asp323, and Glu328 are shown in stick representation.

(C) Differences between transverse relaxation (R_2) at the maximum and minimum applied effective field for residues of the CTD. The residues located at the domain interface show no dependence of the inter-pulse delay during CPMG pulse trains, indicating the lack of dynamics on the microsecond-millisecond scale causing an exchange contributing to the transverse relaxation.

domains. Using a sample composition with known fractions of free and bound (separated) domains, k_{off} can be estimated from the coalescence of signals in the spectra. During the titration of ^{15}N-labeled *Tm*NusG^{CTD} with unlabeled *Tm*NusG^{NTD}, residues with only small (<60 Hz) chemical shift differences between *Tm*NusG$^{NTD-CTD}$ and isolated *Tm*NusG^{CTD} showed continuous chemical shift changes characteristic for exchange behavior in the fast exchange regime (Figure 7). From the chemical shift changes, a $K_{D,binary}$ of 5 μM could be determined. Resonances with chemical shift differences of about 200 Hz either in the proton or nitrogen dimension disappeared beyond $p_{EL}/p_E > 0.2$. This situation corresponds to $k_{off} \approx 1000\ s^{-1}$ and $k_{on} \approx 2 \times 10^8\ M^{-1}s^{-1}$, a value typical for a diffusion-limited association. Under the assumption that the dissociation rate of the binary complex is similar to the opening rate of the full-length protein, a closing rate of 50,000 s^{-1} can be estimated. This rate is faster than the effective rate of association in the binary complex at typical concentrations ([NTD] = 0.1–1 mM), given by k_{on} * [NTD], reflecting the tethered protein domains in the full-length protein.

These results show that *Tm*NusG$^{NTD-CTD}$ exists in a dynamic equilibrium between the major closed conformation and the minor open conformation. The fast opening with a rate of approximately 1,000 s^{-1} is not expected to be a kinetic barrier during formation of protein complexes required for transcriptional processes.

Interdomain Salt Bridges Contribute Little to the Domain Association

The crystal structure of *Tm*NusG identified two salt bridges (Arg275-Asp314 and Arg279-Glu313) between NTD and CTD. An increased number of salt bridges is frequently observed in proteins from thermophilic organisms compared to their mesophilic counterparts (Kumar et al., 2000). Sequence comparison of *Tm*NusG with NusG from other bacteria suggests that these salt bridges are a unique feature of *Tm*NusG; no other NusG ortholog shows amino acid combinations at these positions suitable for analogous ionic interactions.

To test whether these salt bridges are essential for the NTD-CTD interaction in *Tm*NusG, two single point mutants (Arg275Ala and Arg279Ala) and the double mutant (Arg275Ala + Arg279Ala) were studied. ^{1}H, ^{15}N HSQC spectra of *Tm*NusG$^{NTD-CTD}$ (Arg275Ala) and *Tm*NusG$^{NTD-CTD}$ showed no significant chemical shift differences. Only small changes were observed for residues sequentially and spatially close to the mutated site. Specifically, in comparison to the isolated *Tm*NusG^{CTD} the characteristic large chemical shift differences were still present, showing that *Tm*NusG$^{NTD-CTD}$(Arg275Ala) maintains the closed conformation.

Compared to *Tm*NusG$^{NTD-CTD}$, *Tm*NusG$^{NTD-CTD}$(Arg279Ala) and the double mutant *Tm*NusG$^{NTD-CTD}$(Arg275Ala + Arg279Ala), showed larger chemical shifts for resonances of both, NTD and CTD. Especially in the C-terminal helix of the NTD, where the two arginines are located, it was no longer possible to assign the residues unambiguously by HQSC comparisons. Moreover, these mutants exhibited a significantly lower solubility (<100 μM), rendering NMR experiments for more detailed information (^{15}N relaxation, assignment by triple-resonance NMR) impossible. As the signals from the CTD in *Tm*NusG$^{NTD-CTD}$ still showed remarkable differences to those of the isolated CTD, we conclude that a significant fraction of both mutant proteins stayed in the closed conformation. Compared to the parent *Tm*NusG$^{NTD-CTD}$, the large chemical shift changes may have resulted from minor differences in the relative domain orientations, which are due to the lack of the salt bridges.

Together, these results suggest that additional interactions are essential and predominantly responsible for the stable intramolecular domain interaction. This interpretation would also explain that the solubility of *Tm*NusG^{NTD} is much lower than that of *Ec*NusG^{NTD}. Similarly to the NTD of *Ec*RfaH the hydrophobic surface exposed to the solvent is much larger in *Tm*NusG^{NTD}.

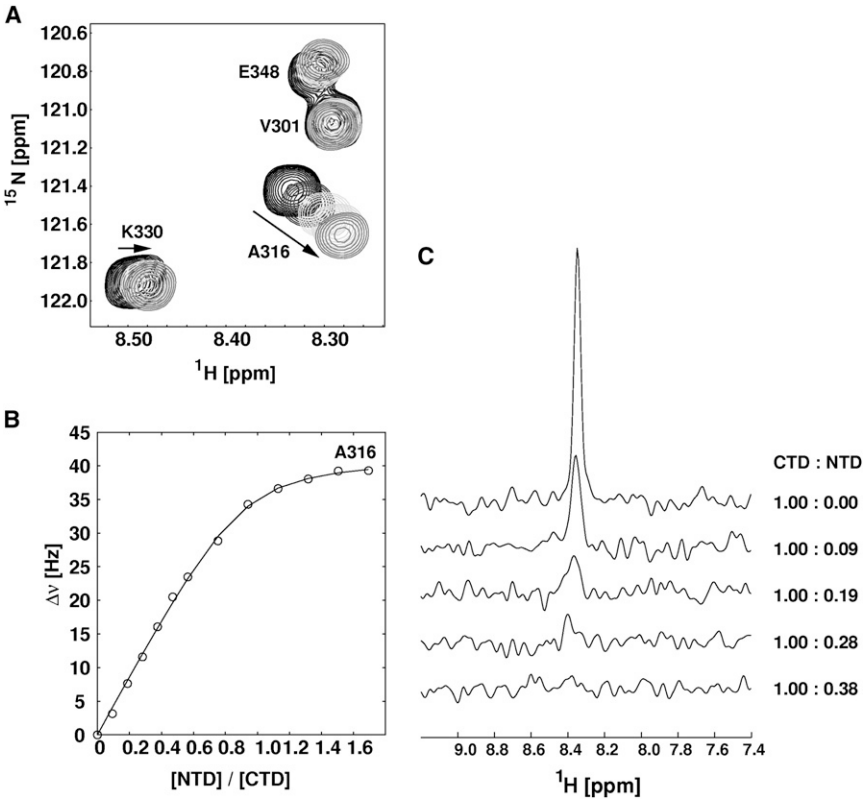


Figure 7. Titration of Isolated ^{15}N-Labeled *Tm*NusG^{CTD} with Unlabeled *Tm*NusG^{NTD}

(A) Expanded region of superimposed ^{1}H, ^{15}N HSQC spectra of *Tm*NusG-CTD with increasing concentrations of *Tm*NusG^{NTD}. The arrows mark the direction of chemical shift changes during titration.

(B) Chemical shift changes (expressed as $\Delta\nu$ = sqrt($(\Delta\nu_{1H})^2$+$(\Delta\nu_{15N})^2$) of Ala316 during the titration. Fitting the titration curve to a two-state binding model resulted in a dissociation constant (K_D) of 5 μM.

C) One-dimensional traces of the ^{1}H, ^{15}N HSQC spectra with given ratios of CTD:NTD showing the disappearing signal of Gly314. The signal disappearing at a CTD:NTD ratio of 0.3 corresponds to a dissociation rate of approximately 1,000 s^{-1}.

*Tm*NusG or *Tm*NusG$^{NTD-CTD}$ Do Not Complement a NusG Deletion in *E. coli*

The genome of *T. maritima* does not encode for another NusG-like protein (suggesting that *T. maritima* lacks an RfaH ortholog) but contains genes for putative NusA, NusB, NusE/S10, and Rho proteins (Nelson et al., 1999). Notably, *Tm*NusA also contains a different domain arrangement compared to *Ec*NusA (Shin et al., 2003; Worbs et al., 2001), in this case lacking two C-terminal double helix-hairpin-helix domains (Bonin et al., 2004; Eisenmann et al., 2005), which in *E. coli* serve an autoinhibitory function (Mah et al., 2000; Schweimer et al., 2011). The conservation of NusG in all bacteria and higher organisms, the lack of an additional NusG-like open reading frame in *T. maritima*, and the observation that also other presumed *T. maritima* transcription factors have different domain compositions compared to *E. coli* orthologs suggest that despite the DII insertion and the NTD-CTD interaction, *Tm*NusG should maintain similar functions as known for *E. coli*.

We assumed that the preferred closed state of *Tm*NusG at 298 K and 323 K might interfere with the functions of this protein as a transcription factor at elevated temperatures. Therefore, we tested whether *Tm*NusG or *Tm*NusG$^{NTD-CTD}$ could rescue an *E. coli nusG* deletion. The corresponding genes were cloned into plasmid pBAD/HisA to allow induction by arabinose. The *E. coli* strain W3102 bearing a chromosomal *nusG::kan* insert and a temperature-sensitive plasmid that expresses *Ec*NusG was used for gene expression (Sullivan and Gottesman, 1992) (Figure 8). At 42°C, the *ecnusG* plasmid is lost and cells are not viable in the absence of an alternative source of active NusG. After the addition of arabinose, full-length *Tm*NusG (or *Tm*NusG$^{NTD-CTD}$, data not shown) expressed from the pBAD/HisA plasmid was not able to support cell growth at 42°C. In agreement with this finding, previous analyses have also shown that *Tm*NusG or a domain II deletion variant of the protein were unable to substitute for *E. coli* NusG (Liao et al., 1996).

In contrast, cells supplemented with an *Aquifex aeolicus nusG* gene on the same plasmid grew with 100% efficiency at 42°C in the presence of arabinose. Even though both NusG proteins originate from hyperthermophilic organisms, *Aa*NusG, unlike *Tm*NusG, does not exhibit a stable NTD-CTD interaction (Knowlton et al., 2003; Steiner et al., 2002). These results indicate that *Tm*NusG is not functional in *E. coli*. This is possibly due to its preference for a stable closed conformation at *E. coli* growth temperatures. Expression of *Tm*NusG in cells producing *Ec*NusG induces a small colony phenotype at 37°C and kills at 42°C (data not shown), suggesting that *Tm*NusG interacts with and inhibits some *E. coli* transcription reaction, possibly Rho-dependent transcription termination.

Conclusions

We have shown with crystal and solution NMR structural analyses that the transcription factor NusG of the hyperthermophilic bacterium *Thermotoga maritima* exhibits a stable interaction between its NTD and CTD. The interaction is subject to fast dynamics on the NMR time scale with ca. 98% of the molecules existing in the closed conformation at 323 K. H/D exchange studies revealed that this interaction contributes significantly to the overall fold stability of the protein. Thus, it may represent a safeguard against unfolding or aggregation at high temperature. However, NusG proteins from other hyperthermophilic organisms, such as *Aquifex aeolicus*, which exhibit a very similar domain architecture as *Tm*NusG, do not show a stable NTD-CTD domain interaction. Therefore, the NTD-CTD interaction does not represent a mandatory adaption to life at high temperatures.

We showed that the closed state of *Tm*NusG is incompatible with other functional interactions of the NTD (with RNAP) and CTD (with NusE or Rho). In contrast to *Tm*NusG, which

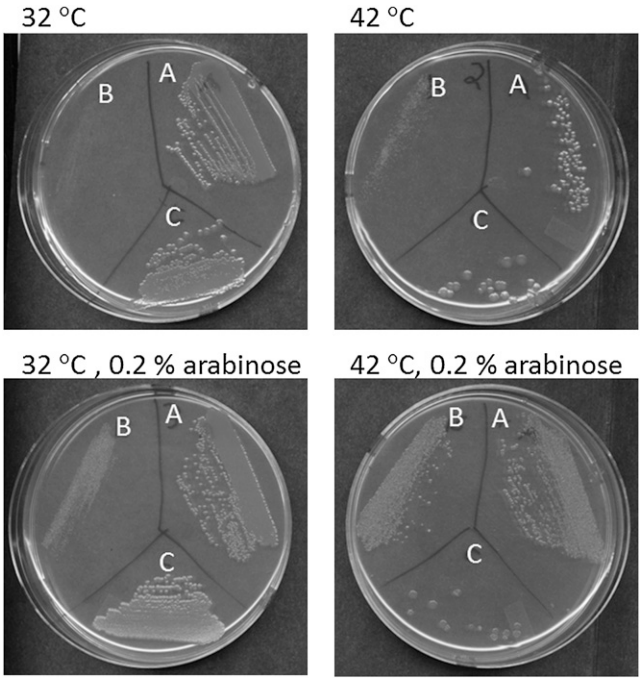


Figure 8. Test for Complementation of *Ec*NusG by *Tm*NusG
(A–C) The growth of the following *E. coli* strains was tested at the growth conditions indicated: (A) 9388, which carries the temperature-sensitive plasmid *ecnusG* and plasmid pBAD-*aanusG*; (B) 9391, which was derived from 9388 at 42°C in the presence of arabinose and has lost the *ecnusG* plasmid; and (C) 11480, which carries the temperature-sensitive plasmid *ecnusG* and plasmid pBAD-*tmnus*G.
See also Figure S2.

predominantly adopts a closed conformation, NusG from *Aquifex aeolicus*, which does not exhibit a stable NTD-CTD interaction, fully complemented a NusG deletion in *E. coli*. Therefore, we conclude that in some organisms, e.g., *Thermotoga maritima*, NusG is subject to autoinhibition via the characterized domain interaction. Autoinhibition may prevent NusG from interacting prematurely with other components of the transcription complex or may preclude nonspecific interactions of NusG with other cellular components. Release from autoinhibition may be achieved by the presentation of NusG binding sites with sufficiently high affinity on preformed transcription complexes. Alternatively, it may require interaction of NusG with a specific component of the transcription complex, such as the participating nucleic acids. An autoinhibited state in the isolated protein that is released in the presence of a specific DNA signal sequence, *ops*, has been described for the NusG paralog RfaH (Belogurov et al., 2007; Burmann et al., 2012). In support of this notion, *Tm*NusG has been found to strongly interact with various nucleic acids (Liao et al., 1996) unlike the mesophilic *Ec*NusG, in which no comparable autoinhibited conformation was detected.

EXPERIMENTAL PROCEDURES

Experimental procedures including details of cloning and of protein production, modification, purification, crystallographic and NMR spectroscopic analysis, structure calculation, and complementation test in *E. coli* are described in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The structure coordinates were deposited in the Protein Data Bank under the accession codes 2XHC (crystal structure of *Tm*NusG^{FL}), 2XHA (crystal structure of *Tm*NusG^{DII}), and 2LQ8 (solution structure of *Tm*NusG$^{NTD-CTD}$). Chemical shift assignments of *Tm*NusG$^{NTD-CTD}$ were deposited in the BioMagResBank under the accession code 18298.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, one table, and two figures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2012.12.015.

ACKNOWLEDGMENTS

We are grateful to Elke Penka and Ramona Heißmann for technical support and Homa Ghalei for help in diffraction data collection. We thank Gleb Bourenkov and Hans Bartunik for support at beamline BW6 (DESY, Hamburg, Germany). This work was supported by grants Ro 617/17-1 (to P.R.) and WA1126/5-1 (to M.C.W.) from the Deutsche Forschungsgemeinschaft.

Received: July 2, 2012
Revised: November 19, 2012
Accepted: December 21, 2012
Published: February 14, 2013

REFERENCES

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

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.

Bonin, I., Mühlberger, R., Bourenkov, G.P., Huber, R., Bacher, A., Richter, G., and Wahl, M.C. (2004). Structural basis for the interaction of Escherichia coli NusA with protein N of phage lambda. Proc. Natl. Acad. Sci. USA *101*, 13762–13767.

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. USA *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.

Cho, J.H., Muralidharan, V., Vila-Perello, M., Raleigh, D.P., Muir, T.W., and Palmer, A.G., 3rd. (2011). Tuning protein autoinhibition by domain destabilization. Nat. Struct. Mol. Biol. *18*, 550–555.

Eisenmann, A., Schwarz, S., Prasch, S., Schweimer, K., and Rösch, P. (2005). The E. coli NusA carboxy-terminal domains are structurally similar and show specific RNAP- and lambdaN interaction. Protein Sci. *14*, 2018–2029.

Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G.A., Teng, M., and Niu, L. (2008). Core structure of the yeast spt4-spt5 complex: a conserved module for regulation of transcription elongation. Structure *16*, 1649–1658.

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.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256–2268.

Kumar, S., Tsai, C.J., and Nussinov, R. (2000). Factors enhancing protein thermostability. Protein Eng. *13*, 179–191.

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.

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

Li, P., Martins, I.R., Amarasinghe, G.K., and Rosen, M.K. (2008). Internal dynamics control activation and activity of the autoinhibited Vav DH domain. Nat. Struct. Mol. Biol. *15*, 613–618.

Liao, D., Lurz, R., Dobrinski, B., and Dennis, P.P. (1996). A NusG-like protein from Thermotoga maritima binds to DNA and RNA. J. Bacteriol. *178*, 4089–4098.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins *50*, 437–450.

Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcárcel, J., and Sattler, M. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature *475*, 408–411.

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

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

Mooney, R.A., Schweimer, K., Rösch, P., Gottesman, M.E., 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.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., et al. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. Nature *399*, 323–329.

Pasman, Z., and von Hippel, P.H. (2000). Regulation of rho-dependent transcription termination by NusG is specific to the Escherichia coli elongation complex. Biochemistry *39*, 5573–5585.

Pawson, T., and Nash, P. (2000). Protein-protein interactions define specificity in signal transduction. Genes Dev. *14*, 1027–1047.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445–452.

Pufall, M.A., and Graves, B.J. (2002). Autoinhibitory domains: modular effectors of cellular regulation. Annu. Rev. Cell Dev. Biol. *18*, 421–462.

Reay, P., Yamasaki, K., Terada, T., Kuramitsu, S., Shirouzu, M., and Yokoyama, S. (2004). Structural and sequence comparisons arising from the solution structure of the transcription elongation factor NusG from *Thermus thermophilus*. Proteins *56*, 40–51.

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

Shin, D.H., Nguyen, H.H., Jancarik, J., Yokota, H., Kim, R., and Kim, S.H. (2003). Crystal structure of NusA from Thermotoga maritima and functional implication of the N-terminal domain. Biochemistry *42*, 13429–13437.

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. USA *90*, 970–974.

Steiner, T., Kaiser, J.T., Marinkoviç, S., Huber, R., and Wahl, M.C. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. EMBO J. *21*, 4641–4653.

Sullivan, S.L., and Gottesman, M.E. (1992). Requirement for E. coli NusG protein in factor-dependent transcription termination. Cell *68*, 989–994.

Sullivan, S.L., Ward, D.F., and Gottesman, M.E. (1992). Effect of Escherichia coli nusG function on lambda N-mediated transcription antitermination. J. Bacteriol. *174*, 1339–1344.

Torres, M., Balada, J.M., Zellars, M., Squires, C., and Squires, C.L. (2004). In vivo effect of NusB and NusG on rRNA transcription antitermination. J. Bacteriol. *186*, 1304–1310.

Wenzel, S., Martins, B.M., Rösch, P., and Wöhrl, B.M. (2010). Crystal structure of the human transcription elongation factor DSIF hSpt4 subunit in complex with the hSpt5 dimerization interface. Biochem. J. *425*, 373–380.

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.

Worbs, M., Bourenkov, G.P., Bartunik, H.D., Huber, R., and Wahl, M.C. (2001). An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. Mol. Cell *7*, 1177–1189.

Zhou, Y., Filter, J.J., Court, D.L., Gottesman, M.E., and Friedman, D.I. (2002). Requirement for NusG for transcription antitermination in vivo by the lambda N protein. J. Bacteriol. *184*, 3416–3418.

Structure, Volume *21*

Supplemental Information

An Autoinhibited State in the Structure of *Thermotoga maritima* NusG

Johanna Drögemüller, Christian M. Stegmann, Angshuman Mandal, Thomas Steiner,, Björn M. Burmann, Max E. Gottesman, Birgitta M. Wöhrl, Paul Rösch, Markus C. Wahl, and Kristian Schweimer

Inventory of Supplemental Information

This table is related to figure 3 (^{15}N spin relaxation data) and summarizes the diffusion tensor analysis to further demonstrate that the intramolecular domain interaction observed in the crystal structure of full length *Tm*NusG is also present in the same fashion in the 'reduced' construct *Tm*NusG$^{NTD-CTD}$ under solution state conditions.

This figure is related to figure 1. The sequence alignment shows that the additional domain of *Tm*NusG is inserted in the same manner as found for *Aquifex aeolicus* NusG.

This figure is related to figure 8. This figures shows the expression of *Tm*NusG in *E. coli* strain 11840 in which no complementation of *Ec*NusG could be observed.

5) References for Supplemental Experimental Procedures

Supplemental Experimental Procedures

Protein production

A DNA construct encoding *Tm*NusG^{FL} was PCR-amplified from *Thermotoga maritima* genomic DNA and cloned *via* NcoI/BamHI restriction sites into the expression vector pET22b(+) (Novagen, Madison, WI, USA). For protein production, the plasmid was transformed into *E. coli* BL21(DE3) cells (Novagen, Madison, WI, USA) and the cells were cultivated at 37 °C in Luria-Bertani (LB) medium in the presence of 100 μg/ml ampicillin. After cell density had reached an OD_{600} of 0.8, expression was induced by addition of 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) and the culture incubated at 30 °C for an additional 3 h. Cells were harvested in 50 mM Tris-HCl, pH 7.0, 20 mM NaCl, 1 mM DTT. Cells were disrupted by sonification and the soluble fraction was heated for 20 min to 90 °C. The supernatant was loaded on a CM Sepharose Fast Flow column (GE Healthcare, Munich, Germany) and eluted in a linear gradient from 20 mM to 300 mM NaCl in 25 mM Tris-HCl, pH 7.0, 1 mM DTT. Pooled fractions were subsequently adjusted to 500 mM ammonium sulfate, loaded onto a Butyl-Sepharose Fast Flow column (GE Healthcare), washed with 20 mM Tris-HCl, pH 7.0, 500 mM ammonium sulfate, 1 mM DTT and eluted in a linear gradient to 20 mM Tris-HCl, pH 7.0, 1 mM DTT. Pooled fractions were buffer exchanged on NAP-25 columns (GE Healthcare) to 10 mM Tris-HCl, pH 7.0, 100 mM NaCl, 1 mM DTT, concentrated to 7 – 12 mg/ml, aliquoted and shock frozen in liquid nitrogen for storage at -80 °C.

A DNA fragment encoding *Tm*NusG^{DII} (residues 42-233) was cloned *via* NcoI and NotI restriction sites into the expression vector pET-M11 (G. Stier, EMBL, Heidelberg, Germany). For production of *Tm*NusG^{DII} bearing an N-terminal TEV-cleavable His_6-tag, the corresponding plasmid was transformed into Rosetta 2 cells (Novagen, Madison, WI, US) and expression was carried out at 16 °C using an auto-inducing protocol (Studier, 2005). Cells were disrupted with a micro-fluidizer (Microfluidics, Newton, MA, US) and the clarified lysate was passed over Ni-NTA agarose (Qiagen, Hilden, Germany). After elution with imidazole and cleavage of the affinity tag with TEV protease, the sample was dialyzed against 20 mM Tris-HCl, pH 8.0, 120 mM NaCl, 1 mM DTT and passed again over the affinity resin to remove uncleaved protein and free tag. The flow-through was subjected to size exclusion chromatography (Superdex 75) in 20 mM Tris-HCl, pH 8.0, 120 mM NaCl, 1 mM DTT. SeMet-derivatized protein was produced in a defined medium (Studier, 2005) and purified in the same fashion as the native protein. Pooled fractions were concentrated to 10-20 mg/ml by ultra-filtration, aliquoted and flash-frozen in liquid nitrogen for storage at −70 °C.

*Tm*NusG$^{NTD\text{-}CTD}$ (*Tm*NusG(1-42) + *Ec*NusG(52-61) + *Tm*NusG(231-354)) was cloned *via* NotI and NcoI restriction sites into pET-M11. *E. coli* strain BL21(DE3) harboring the recombinant plasmid was

grown at 37 °C in LB medium (Luria-Bertani) or M9-medium (for isotope labeling) containing kanamycin (30 µg/ml) until an OD_{600} of 0.8 was reached and the cells were induced to a final concentration of 1 mM IPTG. Cells were harvested 4 h after induction, resuspended in 4 times the pellet weight of lysis buffer (50 mM TRIS, 300 mM NaCl, pH 7.5, ½ Protease inhibitor tablet (Complete, EDTA-free, Roche Diagnostics GmbH, Mannheim, Germany)), and lysed as above. The lysate was centrifuged for 30 min at 12,000 g and the supernatant was applied to a HisTrapHP column (GE Healthcare, Munich, Germany) using a step gradient with increasing imidazole concentrations (10 – 500 mM). The eluted *Tm*$NusG_{NTD-CTD}$ fusion protein was cleaved by TEV-protease during dialysis against 50 mM TRIS, 300 mM NaCl, pH 7.5 at 4 °C overnight. The cleaved protein was reapplied to a HisTrapHP column (GE Healthcare, Munich, Germany) and uncleaved protein and TEV-protease were eluted with 50 mM TRIS, 300 mM NaCl, 500 mM Imidazol, pH 7.5. Cleaved *Tm*$NusG_{NTD-CTD}$ (in the flow-through) was dialyzed against buffer as used for NMR measurements (25 mM HEPES, 150 mM NaCl, pH 7.5) and concentrated by ultrafiltration (Vivascience, MWCO 5,000 Da).

The gene encoding *Tm*$NusG^{CTD}$ was cloned *via* NcoI and BamHI restriction sites into the *E. coli* expression vector pET-GB1a (G. Stier, EMBL, Heidelberg, Germany) containing an aminoterminal GB1-fusion tag (streptococcal immunoglobulin-binding domain of protein G), an aminoterminal hexa-histidine tag, and a TEV cleavage site between GB1 and *Tm*$NusG^{CTD}$. The gene for *Tm*$NusG^{NTD}$ containing an aminoterminal hexa-histidine tag and a TEV cleavage site was cloned *via* TOPO-cloning (ChampionTM pET Directional TOPO Expression Kit Invitrogen, Karlsruhe, Germany) into the *E. coli* expression vector pET101/D-TOPO. The single mutants (Arg275Ala, Arg279Ala) and the double mutant (Arg275Ala+Arg279Ala) of *Tm*$NusG^{NTD-CTD}$ were created by site-directed PCR mutagenesis. *Tm*$NusG^{CTD}$, *Tm*$NusG^{NTD}$ and all mutants were purified as described for *Tm*$NusG^{NTD-CTD}$.

Crystallographic analyses

Proteins were crystallized by the sitting drop vapor diffusion method at 20 °C. Crystals of *Tm*$NusG^{FL}$ were obtained with a reservoir containing 12 % (w/v) PEG 4000, 200 mM lithium sulfate, 100 mM citric acid/sodium citrate, pH 5.4 – 6.2. Crystals of native *Tm*$NusG^{DII}$ grew in 0.1 M Tris, pH 8.8, 0.2 M sodium acetate, 20 % (w/v) PEG 3350. Crystals of selenomethione-derivatized *Tm*$NusG^{DII}$ were grown in 0.2 M LiCl, 20 % (w/v) PEG 3350. For data collection at cryogenic temperatures, crystals were transferred into paratone N and flash-cooled in liquid nitrogen. All diffraction data were processed with the HKL package (Otwinowski and Minor,).

Attempts to solve the structure of full-length *Tm*NusG by molecular replacement using known structures of NTD and CTD from other organisms failed, most likely because the domains account for

less than one third and for about one sixth of the total *Tm*NusG molecular mass, respectively. As these domains were eventually found to be arranged differently in *Tm*NusG compared to other NusG molecules, they could also not be combined to a single search model. Because well diffracting crystals of full-length *Tm*NusG were poorly reproducible, we resorted to solving the *Tm*NusG^{DII} portion (ca. 50 % of *Tm*NusG) first by anomalous scattering techniques and use the resulting model subsequently for solving the structure of *Tm*NusG^{FL} by molecular replacement.

SAD phasing of SeMet-*Tm*NusG^{DII} data was performed with the SHELX suite of programs (Sheldrick, 2008). SHELXD located four selenium atoms per asymmetric unit. Density modification with SHELXE unequivocally identified the correct hand of the heavy atom sub-structure (pseudo-free correlation coefficients 82 % vs. 55 % for correct vs. incorrect hand). Automated model building with ARP/wARP (Langer, et al., 2008) resulted in a model comprising more than 90 % of the expected protein residues. The FOM-weighted mean phase error relative to the final refined model was 23.4 °. Refinement was conducted using the PHENIX package (Adams, et al., 2004). 5 % of the data were set aside as a test set for calculation of the free R-factor. Ordered solvent was included as implemented in PHENIX. The full-length structure was solved by molecular replacement with Phaser (McCoy, et al., 2007) using the structure of *Tm*NusG^{DII} and of the NTD and CTD of *Aa*NusG (Steiner, et al., 2002) as search models. The *Tm*NusG^{FL} structure was refined in the same way as *Tm*NusG^{DII}.

NMR spectroscopy

All NMR experiments were conducted on Bruker Avance 600 MHz, 700 MHz, and 800 MHz spectrometers, the latter two equipped with cryogenically cooled probes. Standard triple resonance experiments (Bax and Grzesiek, 1993; Sattler, et al., 1999) were conducted for backbone and sidechain resonance assignments at 323 K. ^{15}N- and ^{13}C-edited 3D NOESY experiments were recorded with mixing times of 100 ms at 323 K.

Residual dipolar couplings were determined by IPAP experiments (Ottiger, et al., 1998) using samples with ^{15}N-labeled protein dissolved in C12E6/hexanol/water (Rückert and Otting, 2000)(3 % w/v C12E6, molar ratio C12/E6:hexanol 0.64) at 295 K and with 5 mg/ml Pf1 phages (Hansen, et al., 1998) (Hyglos, Bernried, Germany) at 323K.

For the characterization of overall and internal motion, ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates together with the $\{^1H\}^{15}N$ steady state NOE were recorded using standard methods (Kay, et al., 1989) at 600.2 MHz ^{1}H frequency and calibrated temperatures of 298 and 323 K. Relaxation delays of R_1 and R_2 relaxation experiments were fitted to a mono-exponential decay using

the program curve fit (A.G. Palmer, Dept. of Biochemistry and Molecular Biophysics, Columbia University, USA). Diffusion tensor analysis was done with the TENSOR2 package (Dosset, et al., 2000) with the crystal structure coordinates. Relaxation dispersion experiments were performed at 14.1 T using the constant time approach with a 80 ms constant time period (Mulder, et al., 2001; Tollinger, et al., 2001). The CPMG interval τ_{CP} (= t-180°-t) was varied between 0.5 and 10 ms.

For H/D experiments, ^{15}N NMR samples were dialyzed against NMR buffer and then lyophilized. The lyophilized protein samples were dissolved in D_2O (99.98 %), and the decay of signal intensities was observed in series of ^{1}H,^{15}N HSQC experiments. The exchange rate was determined by fitting the signal decay to a monoexponential curve. The Protection Factor (Bai, et al., 1994) was calculated by dividing the experimental exchange rates by the intrinsic exchange rates calculated from the amino acid sequence, and experimental conditions using tabulated parameters (Bai, et al., 1993).

Binding of *Tm*NusG^{NTD} (concentration of stock solution 150 µM) to *Tm*NusG^{CTD} (initial concentration 200 µM) was studied by observing chemical shift changes in a series ^{1}H, ^{15}N HSQC spectra of *Tm*NusG^{CTD} with increasing amounts of *Tm*NusG^{NTD}. The dissociation constant was determined by fitting a two-state model to chemical shift changes for residues showing exchange behavior in the fast exchange region of the chemical shift time scale. The dissociation rate of the binary complex was estimated from the signal disappearing of residues showing intermediate exchange behavior, the calculated concentrations of all species using the previous determined K_D and the known sample composition, and the chemical shift difference between free and bound state, taken from the comparison of chemical shift data of *Tm*NusG$^{NTD-CTD}$ and *Tm*NusG^{CTD}, assuming that the binary complex resembles the situation of linked domains in TmNusG$^{NTD-CTD}$.

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 does not show a water exchange cross peak in the ^{15}N-edited NOESY spectrum.

The structure calculations were performed with the program XPLOR-NIH 1.2.1 (Schwieters, et al., 2003) using a three-step simulated annealing protocol with floating assignment of prochiral groups

including a conformational database potential. The 20 structures showing the lowest values of the target function excluding the database potential were further analyzed with X-PLOR (Schwieters, et al., 2003), pymol, and PROCHECK 3.5.4 (Laskowski, et al., 1993).

Complementation tests in E. coli

For complementation tests in *E. coli*, the genes encoding *Tm*NusG^{FL}, *Tm*NusG$^{NTD-CTD}$ and *Aa*NusG were cloned into pBAD/HisA (Invitrogen, Life Technologies GmbH, Darmstadt, Germany) *via* BsmBI/XhoI restriction enzyme sites. Strains are derivatives of W3102 bearing a temperature-sensitive *ecnusG* plasmid carrying a *tetR* marker (Sullivan and Gottesman, 1992). Transformation of the parental strain with plasmids expressing *pBAD-aanusG* or *pBAD-tmnusG* was performed at 32 ºC on LB plus ampicillin (100 µg/ml) plates. Strain 9388 carries plasmid *pBAD-aanusG*. Strain 9391 was derived from 9388 by selection at 42 ºC on LB plus 0.2 % arabinose. Strain 9391 has lost the *ecnusG* plasmid and is tetracyclin-sensitive. Strain 11480 carries plasmid *pBAD-tmnusG*.

Supplemental Table 1

Rotational diffusion tensor analysis

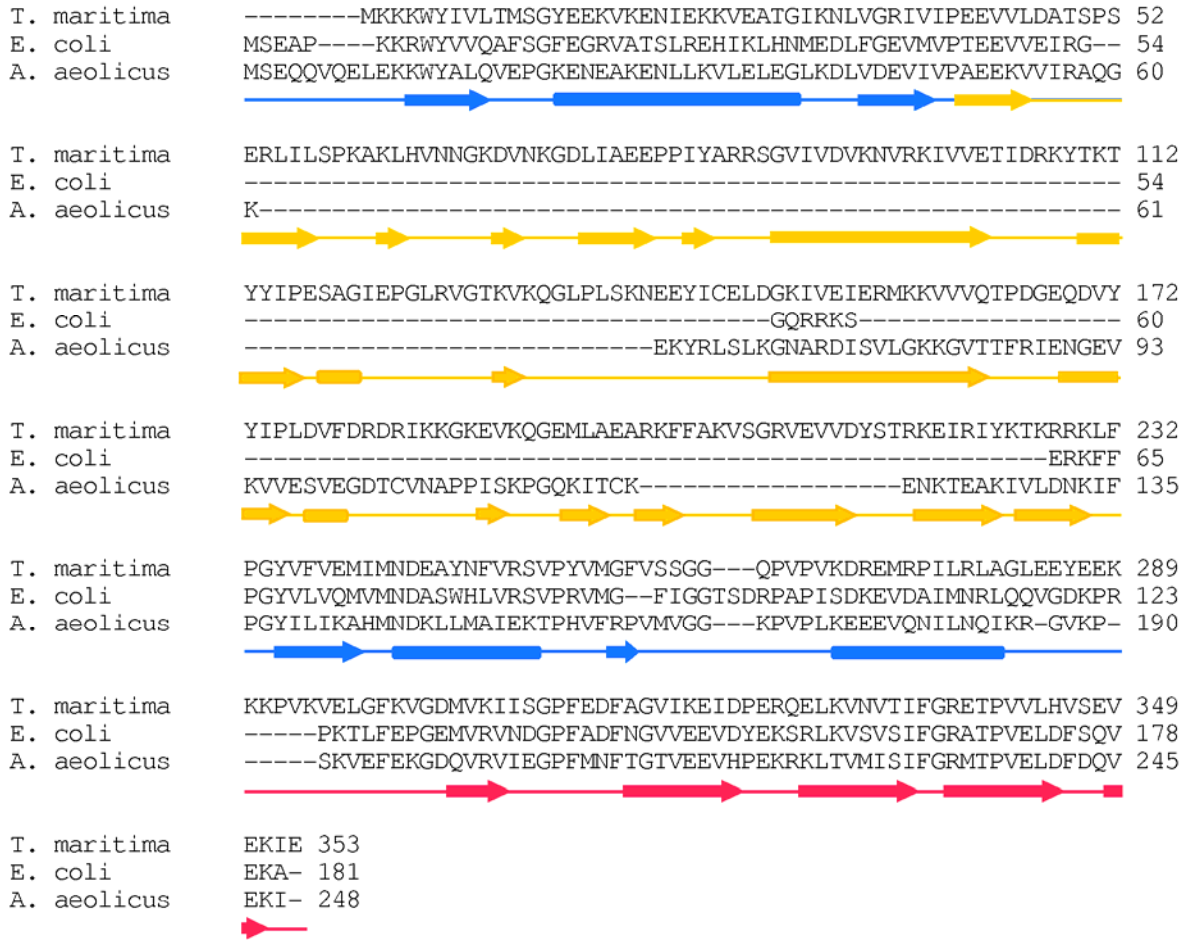
isotropic		**axialsymmetric (prolate)**		**axialsymmetric (oblate)**		**asymmetric**	
		$D_\perp$ (10^8 s^{-1})[a]	0.1967	$D_\perp$ (10^8 s^{-1})[a]	0.2333	D_x (10^8 s^{-1})	0.1933
		$D_\parallel$ (10^8 s^{-1})[a]	0.2787	$D_\parallel$ (10^8 s^{-1})[a]	0.2207	D_y (10^8 s^{-1})	0.1996
						D_z (10^8 s^{-1})	0.2801
t_c (ns)	7.38	t_c (ns)	7.44	t_c (ns)	7.27	t_c (ns)	7.43
χ[b]	$1.39 \cdot 10^2$	χ^2	$6.94 \cdot 10^1$	χ^2	$1.25 \cdot 10^2$	χ^2	$6.88 \cdot 10^1$

[a] $D_\parallel = D_z$, $D_\perp = D_x = D_y$ for the axialsymmetric model

[b] $\chi^2 = \Sigma(T_{1i,exp} - T_{1i,calc})^2/\sigma(T_{1i}) + \Sigma(T_{2i,exp} - T_{2i,calc})^2/\sigma(T_{2i})$

Supplemental Figure 1

Figure S1 Alignment of NusG sequences from *T. maritima*, *E. coli* and *A. aeolicus*. The secondary structure elements corresponding to the *Tm*NusG crystal structure are indicated below the sequences, color coded by domains: NTD – blue, domain II – yellow, CTD – light red.



Supplemental Figure 2

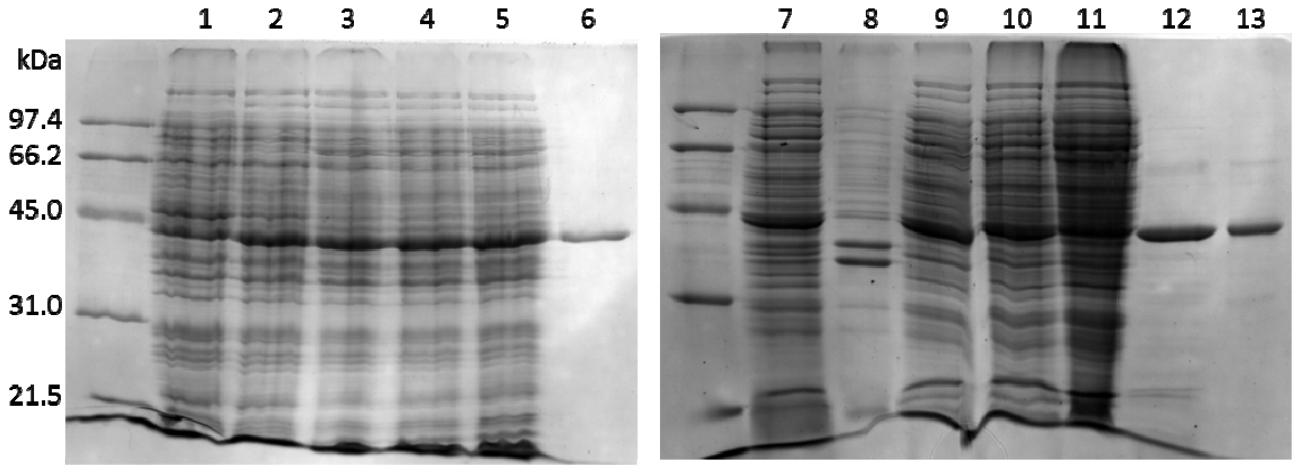


Figure S2: Expression of *Tm*NusG by *E. coli* strain 11840 at 32 °C seen on 15 % SDS-PAGE gels, stained with Coomassie Brilliant Blue.

Lane 1: Cell lysate of 11840 before induction with 0.2 % arabinose, lane 2-5: cell lysates 1-4 h after induction, lane 6: purified *Tm*NusG. Lane 7-9: total cell extract, pellet and supernatant after cell disruption, lane 10-12: total cell extract, pellet and supernatant after heat treatment at 80 °C for 20 min and lane 13: purified *Tm*NusG. The protein expressed in strain 11840 was verified by mass spectrometry.

Supplemental References

Adams, P.D., Gopal, K., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Pai, R.K., Read, R.J., Romo, T.D., *et al.* (2004). Recent developments in the PHENIX software for automated crystallographic structure determination. J. Synchrotron Radiat. 11, 53-55.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. (1994). Protein stability parameters measured by hydrogen exchange. Proteins 20, 4-14.

Bai, Y., Milne, J.S., Mayne, L., and Englander, S.W. (1993). Primary structure effects on peptide group hydrogen exchange. Proteins 17, 75-86.

Bax, A., and Grzesiek, A. (1993). Methodological advances in protein NMR. Acc. Chem. Res. 26, 131-138.

Dosset, P., Hus, J.C., Blackledge, M., and Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. J. Biomol. NMR 16, 23-28.

Hansen, M.R., Mueller, L., and Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat. Struct. Biol. 5, 1065-1074.

Kay, L.E., Torchia, D.A., and Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry (N. Y.) 28, 8972-8979.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nat. Protoc. 3, 1171-1179.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283-291.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. 40, 658-674.

Mulder, F.A., Skrynnikov, N.R., Hon, B., Dahlquist, F.W., and Kay, L.E. (2001). Measurement of slow (micros-ms) time scale dynamics in protein side chains by (15)N relaxation dispersion NMR spectroscopy: application to Asn and Gln residues in a cavity mutant of T4 lysozyme. J. Am. Chem. Soc. 123, 967-975.

Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J. Magn. Reson. 131, 373-378.

Otwinowski, Z., and Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology 276, 307-326.

Rückert, M., and Otting, G. (2000). Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. J. Am. Chem. Soc. 122, 7793-7797.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution empolying pulsed field gradients. Prog. NMR Spectrosc. 34, 93-158.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003). The Xplor-NIH NMR molecular structure determination package. J. Magn. Reson. 160, 66-74.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallogr. A 64, 112-122.

Steiner, T., Kaiser, J.T., Marinkovic, S., Huber, R., and Wahl, M.C. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. EMBO J. 21, 4641-4653.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 41, 207-234.

Sullivan, S.L., and Gottesman, M.E. (1992). Requirement for E. coli NusG protein in factor-dependent transcription termination. Cell 68, 989-994.

Tollinger, M., Skrynnikov, N.R., Mulder, F.A., Forman-Kay, J.D., and Kay, L.E. (2001). Slow dynamics in folded and unfolded states of an SH3 domain. J. Am. Chem. Soc. 123, 11341-11352.

9 Einzelarbeit C

Johanna Drögemüller*, Martin Strauß*, Kristian Schweimer, Birgitta M. Wöhrl, Stefan H. Knauer und Paul Rösch (2015): Exploring RNA Polymerase Regulation by NMR Spectroscopy. *Scientific Reports* **5**: 10825

* beide Autoren haben im gleichen Maße zur Arbeit beigetragen

SCIENTIFIC REPORTS

OPEN

Exploring RNA polymerase regulation by NMR spectroscopy

Johanna Drögemüller*, Martin Strauß*, Kristian Schweimer, Birgitta M. Wöhrl, Stefan H. Knauer & Paul Rösch

Received: 13 February 2015
Accepted: 22 April 2015
Published: 04 June 2015

RNA synthesis is a central process in all organisms, with RNA polymerase (RNAP) as the key enzyme. Multisubunit RNAPs are evolutionary related and are tightly regulated by a multitude of transcription factors. Although *Escherichia coli* RNAP has been studied extensively, only little information is available about its dynamics and transient interactions. This information, however, are crucial for the complete understanding of transcription regulation in atomic detail. To study RNAP by NMR spectroscopy we developed a highly efficient procedure for the assembly of active RNAP from separately expressed subunits that allows specific labeling of the individual constituents. We recorded [^{1}H,^{13}C] correlation spectra of isoleucine, leucine, and valine methyl groups of complete RNAP and the separately labeled β′ subunit within reconstituted RNAP. We further produced all RNAP subunits individually, established experiments to determine which RNAP subunit a certain regulator binds to, and identified the β subunit to bind NusE.

The synthesis of RNA is a central process in cells that is carried out by DNA-dependent RNA polymerases (RNAPs). All cellular genomes are transcribed by multisubunit RNAPs that are evolutionary related. In spite of their differences in size and complexity, RNAPs share overall architecture, active-site organization, mechanism of catalysis, and the principles of interactions with nucleic acids[1].

In bacteria, the RNAP core enzyme consists of five subunits, 2xα, β, β', and ω, with different structural and functional roles[2,3]. The C-terminal domains (CTD) of the α subunits (αCTD) are target for many regulatory proteins and are thus key factors for the regulation of transcription[4,5]. Dimerization of the N-terminal domains (NTD) of the α subunits initiates the RNAP assembly process[6]. Next, the β subunit attaches to the α dimer, followed by recruitment of the β' and the ω subunit[6,7]. While the β and β' subunits constitute the active center of RNAP, the ω subunit plays a structural rather than a functional role as it is supposed to bind to the N- and C-termini of the β' subunit to support its proper folding as well as the assembly of β'ω with the $\alpha_2\beta$ complex[7,8]. The σ factor binds to RNAP at the initiation of transcription to form holo RNAP. σ is essential for the recognition and melting of promoter regions, and it leaves RNAP in later stages of transcription[9,10].

Initiation, elongation, and termination of transcription are highly regulated by transcription factors that bind to the transcription elongation complex (TEC) and modify the RNAP[11]. NusG, for example, enhances the transcription rate and suppresses pausing[12]. It interacts with the RNAP β' clamp helices (β'CH) and the RNAP β gate loop (βGL)[13,14]. In contrast to NusG, NusA modifies RNAP to induce pausing and to modulate intrinsic as well as Rho-dependent termination of transcription (reviewed in [15,16]). NusA, NusG, NusB, and NusE can combine with the TEC and certain RNA sequences to form an antitermination complex which is able to read through termination signals, a process that is essential for efficient transcription of ribosomal DNA or the DNA of lambdoid phages[17]. While NusG-NTD mediates RNAP binding, NusG-CTD interacts with NusE in the NusE:NusB complex[18,19]. As NusE, also known as ribosomal protein S10, can be part of the 30S subunit of the ribosome[20], NusG physically links RNAP and the ribosome, thus coupling transcription and translation[18]. Moreover, NusE may also directly interact with RNAP[21].

Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.H.K. (email: stefan.knauer@uni-bayreuth.de)

Numerous crystallographic studies on prokaryotic and eukaryotic RNAPs have elucidated the structural basis of RNAP architecture and gave insights into its function (reviewed in[22]). However, RNAP regulation is heavily dependent on intra- and intermolecular dynamics as well as transient interactions with regulators, which are difficult to study in atomic detail by X-ray crystallography or electron microscopy.

Although nuclear magnetic resonance (NMR) spectroscopy of supramolecular complexes is aggravated by ^{1}H-^{1}H and ^{1}H-^{13}C dipolar interactions that lead to fast relaxation of the magnetization and therefore loss of signal intensity, deuteration[23], application of more sophisticated pulse sequences like transverse relaxation optimized spectroscopy (TROSY), and use of [^{1}H,^{13}C] methyl group probes result in improvements of spectral quality so that proteins up to 670 kDa have been studied successfully[24–26].

Encouraged by these results, we improved the assembly of *E. coli* RNAP from its separately expressed subunits and started to study this reconstituted RNAP by NMR spectroscopy. We use [^{1}H,^{13}C] correlation spectra of isoleucine, leucine, and valine methyl groups in complete RNAP and in the β' subunit of reassembled RNAP to study transcription regulator interactions with RNAP, and we propose to extend this method to other RNAP subunits and RNAPs of other organisms.

Results and Discussion

***In vitro* RNAP assembly, purification, and biochemical characterization.** Bacterial RNAP without ω subunit, but containing σ factor, can be reconstituted from individually expressed and separately purified protein subunits[27–29]. Analysis of elongating RNAP requires, however, inclusion of the ω subunit and omission of the σ factor. Hence, we combined the cell pellets containing the individually expressed subunits α, β, β', or ω, respectively, in lysis buffer with 8 M urea. After cell lysis the lysate was stirred for one hour and subsequently urea was removed by stepwise dialysis. The assembled core RNAP was purified by Ni^{2+} affinity chromatography, and RNAP eluted from a size exclusion chromatography (SEC) column in peaks at 47.5 ml and 54.8 ml (Fig. 1a), corresponding to molecular masses of 980 kDa and 507 kDa, respectively. Analysis of the peak fractions by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) clearly showed that both peaks contained all RNAP subunits, although the 980 kDa fractions included a high amount of impurities (Fig. 1b). The calculated molecular mass of RNAP of 390 kDa suggests that the protein from the second peak is correctly reconstituted RNAP free of major contaminants. As a reference, we used RNAP assembled *in vivo* ($RNAP^{native}$), where the genes of the subunits were located on a single plasmid. Indeed, $RNAP^{native}$ eluted from the SEC column in a main peak coinciding with the 507 kDa peak (Fig. 1a). An activity assay testing the ability of RNAP to elongate an RNA primer showed that protein from the 507 kDa peak and $RNAP^{native}$ were both functionally identical (Fig. 1c). Therefore, we refer to active reassembled RNAP as $RNAP^{active}$ in contrast to inactive reassembled RNAP ($RNAP^{inactive}$) from the 980 kDa peak.

The far-UV circular dichroism (CD) spectra of $RNAP^{native}$ and $RNAP^{active}$ are very similar (Fig. 1d), with the typical characteristics of a folded protein. In contrast, the spectrum of $RNAP^{inactive}$ is of lower intensity with less distinct minima, in particular the minimum characteristic for α-helical elements at 208 nm is nearly absent, indicating that $RNAP^{active}$ and $RNAP^{native}$ are folded similarly, whereas $RNAP^{inactive}$ is at least partially unfolded or misfolded.

$RNAP^{active}$ was reapplied onto a SEC column to analyze if it was in equilibrium with $RNAP^{inactive}$. The enzyme eluted in a single peak at the same volume as before, indicating that the protein is stable on the time scale of these experiments (Fig. 1a). Additionally, we could increase the yield of correctly assembled RNAP by de- and renaturation of $RNAP^{inactive}$. Subsequent SEC again yielded peaks at 46.3 and 55.5 ml corresponding to the two RNAP states (Fig. 1a). Hence, at least a portion of the misassembled RNAP could be reconstituted into $RNAP^{active}$.

Overall, the yield was 30–60 mg of $RNAP^{active}$ per liter of bacterial cultures producing α, β, β', and ω, the purity exceeding 95%, similar to the published protocols for RNAP assembly lacking ω. Although the ω subunit of RNAP, encoded by the *rpoZ* gene, is neither essential for cell viability nor for RNAP function, the activity of RNAP lacking σ increases when reassembled in the presence of ω[29–31]. In *rpoZ* deletion strains RNAP copurifies with GroEL and loses its activity upon GroEL removal. However, activity can be regained by denaturation and renaturation of RNAP in the presence of ω[31]. ω was suggested to have important functions in folding of the β' subunit, in preventing β' from aggregation as well as in promoting the assembly of $\alpha_2\beta$ with β'ω[7]. Thus, its presence during reconstitution might reduce the amount of misfolded or misassembled RNAP.

Overall, this assembly and purification strategy allows efficient production of complete, pure, and active core RNAP from separately expressed subunits. In contrast to earlier protocols, purification of one or all individual subunits prior to RNAP assembly is unnecessary, the ω subunit is part of the assembled RNAP, and the presence of the initiation factor σ is not required, so that purified RNAP can be used directly in an elongation context. Finally, by using SEC as final purification step we selectively purify active RNAP and exclude all misassembled and inactive variants, a step that was omitted in most previous protocols.

Purification of individual RNAP subunits and analysis of their secondary structure. We expressed and purified all RNAP subunits separately (α, β, β', and ω) with high yield and purity of $> 95\%$, allowing structural analyses (Supplementary Fig. 1). Additionally, the ββ' complex was assembled from individually expressed subunits and purified according to the protocol used for the assembly of

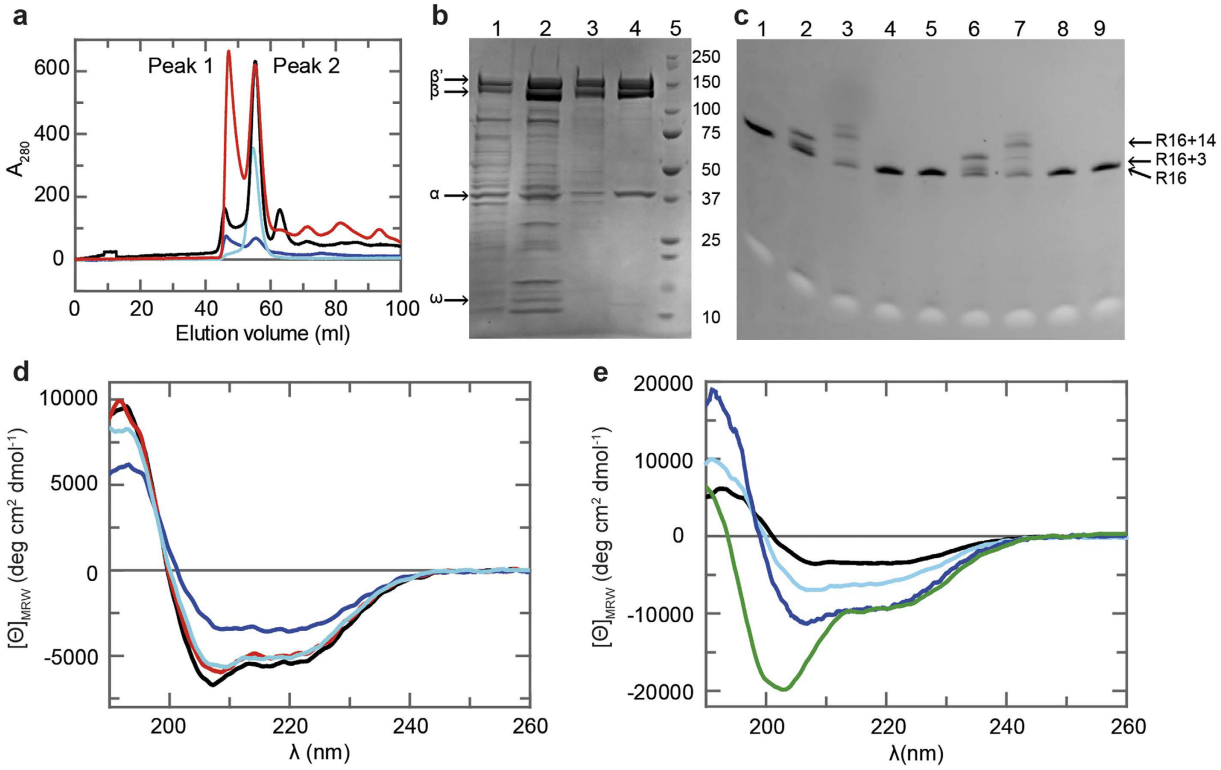


Figure 1. Purification of *in vitro* assembled RNAP. (**a**) Gel filtration chromatograms from an S200 column. Red: combined fractions after Ni^{2+} affinity chromatography; cyan: $RNAP^{active}$; blue: $RNAP^{inactive}$ after de- and renaturation; black: $RNAP^{native}$ (**b**) 4-20% gradient SDS-polyacrylamide gel (Roti-Page, Carl Roth, Karlsruhe, Germany) of aliquots taken during RNAP purification after staining with Coomassie Blue. In lanes 1-4 2 μg protein were applied. Soluble fraction of the assembled RNAP after dialysis (lane 1); combined fractions after Ni^{2+} affinity chromatography (lane 2); SEC peak 1 (lane 3); SEC peak 2 (lane 4); Precision Plus Protein Standard (BioRad, Munich, Germany, lane 5). (**c**) RNAP activity assay, 20% SDS-polyacrylamide gel. 3 pmol RNA were loaded in each lane. Either ATP and CTP or ATP, CTP and GTP were added allowing extension of a 16mer RNA (R16) by 3 or 14 nt, respectively. The arrows indicate the migration positions of R16 and the elongated RNAs. R16, untreated (lane 1); $RNAP^{native}$, elongation by 3 nt (lane 2) or 14 nt (lane 3); $RNAP^{inactive}$, elongation by 3 nt (lane 4) or 14 nt (lane 5); $RNAP^{active}$, elongation by 3 nt (lane 6) or 14 nt (lane 7); control reaction without RNAP, elongation by 3 nt (lane 8) or 14 nt (lane 9). (**d**) Far-UV CD-spectra of 0.6 μM $RNAP^{native}$, black; 0.6 μM $RNAP^{inactive}$, blue; 0.5 μM $RNAP^{active}$, red; 0.6 μM ββ' complex, cyan. (**e**) Far-UV CD-spectra of the separately expressed and purified RNAP subunits. 2.5 μM α, blue; 0.6 μM β, cyan; 1.1 μM β', black; 10 μM ω, green.

RNAP. All proteins were soluble, and although β was isolated from inclusion bodies it showed no tendency to precipitate up to concentrations of 120 μM after refolding. In contrast to previous publications, our protocol yielded soluble β'[28,32].

The far-UV CD spectra of α, β, and β' show the typical characteristics of structured proteins (Fig. 1e), and although the CD spectrum of the ω subunit exhibits the least distinct features, ω does not appear to be completely unfolded. Indeed, ω possesses a structured NTD, followed by an unstructured C-terminus[7] which is in agreement with the [^{1}H,^{15}N]-heteronuclear single quantum coherence (HSQC) spectrum of ^{15}N-labeled ω that shows very low signal dispersion (Supplementary Fig. 2), indicating that the isolated ω is only very poorly folded and might adopt its final structure only upon binding to β' or the complete RNAP. Subunits β and β' represent the largest part of RNAP and the CD spectrum of the ββ' complex is indeed nearly identical to that of $RNAP^{native}$ (Fig. 1d), suggesting that the isolated ββ' complex is assembled as it is in $RNAP^{native}$.

NusG-NTD interacts with β and β' while NusA-NTD binds to β and NusA-AR2 to α. As no activity assay can be conducted for the individual RNAP subunits, their integrity was checked by testing their ability to interact with transcription factors NusG and NusA whose RNAP binding sites are known. NusG consists of two domains that are flexibly connected[19]. It enhances RNAP processivity and reduces pausing by binding to RNAP *via* its NTD[12]. Thus, we first asked which RNAP subunit is the target site

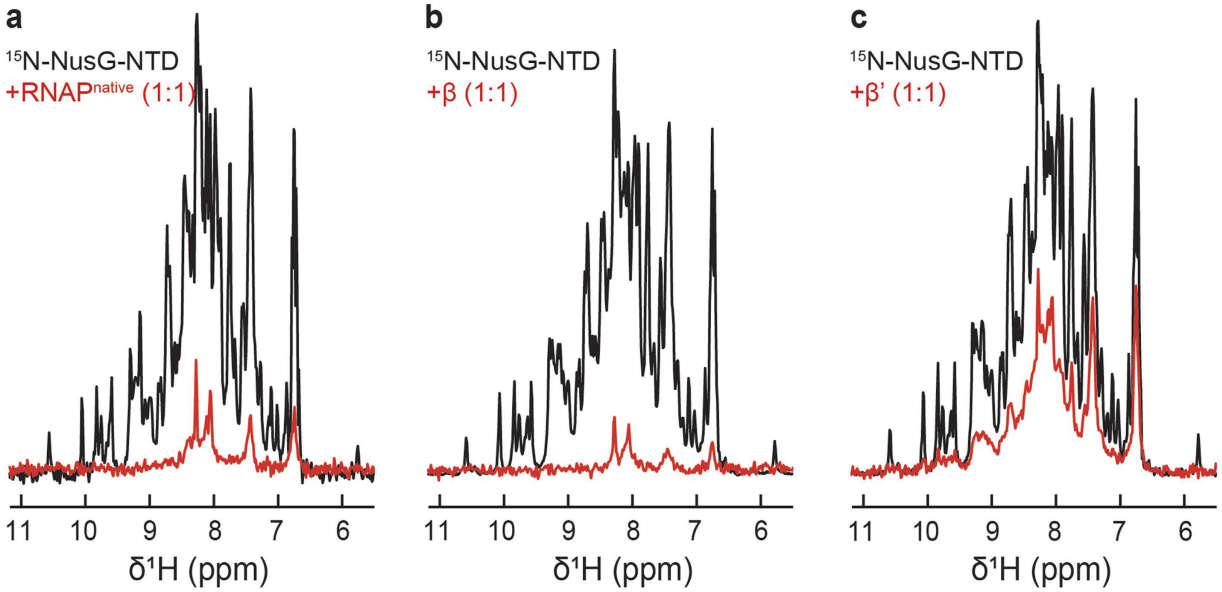


Figure 2. NusG-NTD interaction with RNAP, β, and β'. 1D [^{1}H,^{15}N]-HSQC spectra of the amide region of 30 μM ^{15}N-NusG-NTD in the absence, black, and in the presence of equimolar concentrations, red, of (**a**) RNAP^{native}, (**b**) β subunit, or (**c**) β' subunit.

for NusG-NTD. Upon addition of RNAP^{native}, the signals of ^{15}N-NusG-NTD in the one dimensional (1D) [^{1}H,^{15}N]-HSQC spectrum disappeared, except for a few signals in the random coil area as the resonances of ^{15}N-NusG-NTD are broadened significantly by the dramatic increase in the rotation correlation time due to the formation of the NusG-NTD:RNAP^{native} complex (Fig. 2a). Similarly, addition of isolated β or β' to ^{15}N-NusG-NTD lead to the loss of ^{15}N-NusG-NTD signal intensity (Fig. 2b,c). In contrast, the spectrum remained unaltered upon addition of α or ω (Supplementary Fig. 3), clearly demonstrating that NusG-NTD interacts only with β and β'. When β' was added, the loss of signal intensity was not as dramatic as it was upon addition of RNAP^{native} or β. This can be attributed to inaccuracies in concentration or to a lower affinity of NusG-NTD to β' as compared to complete RNAP or β. Our results are in good agreement with the known binding sites of NusG-NTD, i.e. the β'CH and the βGL (Fig. 3)[13,14].

During transcription, NusA decreases the elongation rate of RNAP, induces pausing, modulates intrinsic and Rho-dependent termination, and is part of the antitermination complex (reviewed in[15,16]). *E. coli* NusA consists of six domains, an NTD, three RNA binding domains (S1, KH1, KH2) that together form the SKK domain, and two C-terminal acidic repeat domains, AR1 and AR2[33,34]. While the interaction partner of NusA-AR2 is the αCTD of RNAP, NusA-NTD binds to the β flap-tip helix[5,35,36].

The NusA-NTD interaction with RNAP and its subunits was probed as with NusG-NTD. The disappearance of ^{15}N-NusA-NTD signals in the presence of RNAP^{native} confirms complex formation (Fig. 4a). However, addition of either β or β' led to an only slight decrease of ^{15}N-NusA-NTD signals, even in the presence of a twofold molar excess of the RNAP subunit (Fig. 4b,c), the effect being slightly more pronounced for the β subunit. In contrast, the signal decrease was more severe when the ββ' complex was added (Fig. 4d). To address the question whether this effect was due to a higher binding affinity or because of the increase in the molecular mass, we determined the observed amide proton transverse relaxation rate R_2 (R_2^{obs}) of free NusA-NTD and of NusA-NTD after addition of β, β', or ββ' in equimolar amounts by spin-echo experiments. R_2^{obs} of NusA-NTD increased in the presence of the individual subunits and the ββ' complex (R_2^{obs}: NusA-NTD, 50 s^{-1}; NusA-NTD+β, 130 s^{-1}; NusA-NTD+β', 90 s^{-1}; NusA-NTD+ββ', 190 s^{-1}). Assuming that R_2^{obs} is population-averaged, the fraction of unbound NusA-NTD was calculated according to equation (3). While the actual R_2 of NusA-NTD corresponds to its R_2^{obs} value, the R_2 values of NusA-NTD completely bound to β, β' or ββ' were estimated based on the proportionality of R_2 and the molecular mass. When β or ββ' were present, approximately 80% of NusA-NTD molecules were unbound, indicating the same affinity of NusA-NTD for β and ββ'. Around 90% of NusA-NTD molecules were free upon addition of β'. Samples containing β' were turbid, suggesting the presence of oligomers with a higher molecular mass, i.e. the fraction of unbound NusA-NTD might be even higher than the estimated value. A small effect of the β' subunit on NusA-NTD binding, however, cannot be excluded. As no interaction was observed between NusA-NTD and the α or the ω subunit (Supplementary Fig. 4), these results agree with previous findings that NusA-NTD interacts with the β flap region (Fig. 3)[35,36].

We probed NusA-AR2:RNAP interaction with the same approach. The signal intensity of ^{15}N-NusA-AR2 was reduced to background levels in the presence of RNAP^{native} (Fig. 5a). The two dimensional (2D)

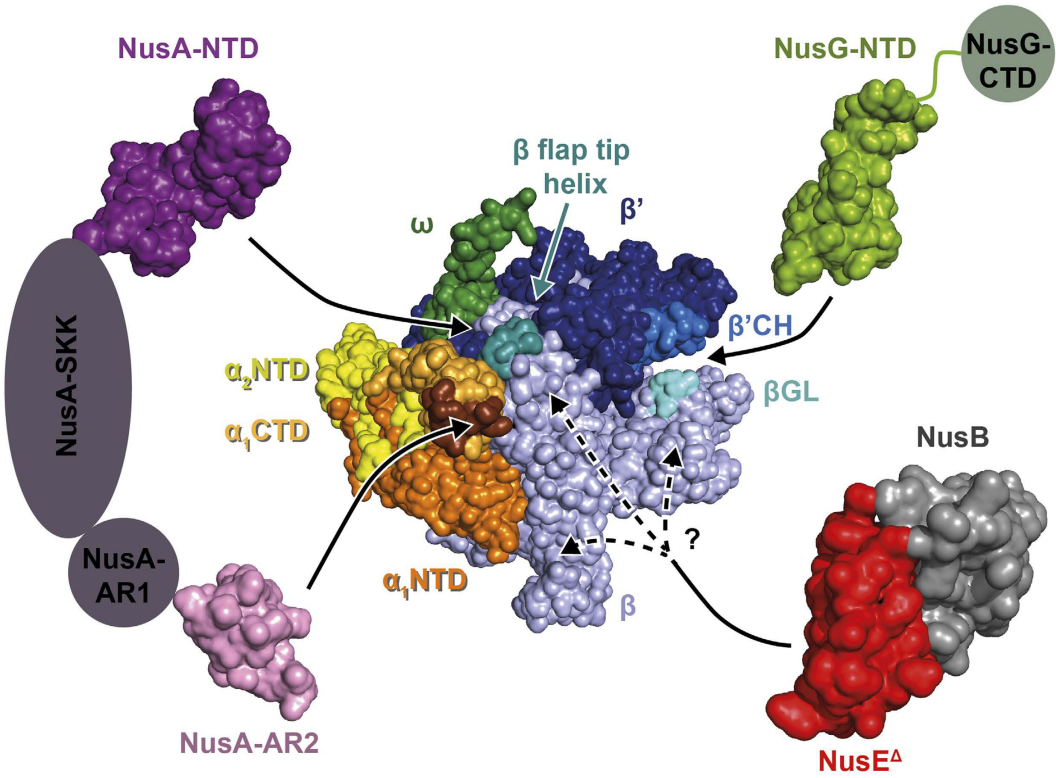


Figure 3. Nus factors binding sites on RNAP. RNAP is shown in surface representation with the NTD and CTD of α subunit 1 in bright and pale orange, respectively, the NTD of α subunit 2 in yellow, the β subunit in pale blue, the β' subunit in dark blue, the ω subunit in dark green. Nus factor binding sites are highlighted (NusA-AR2 binding site on α_1CTD, brown; βGL, cyan; β flap tip helix, turquoise; β'CH, bright blue). Nus factors are displayed in surface representation with linker regions or domains not studied in this work being drawn schematically (NusG, bright green; NusA, purple; $NusE^{\Delta}$, red, NusB, grey). Black arrows indicate the binding site of each Nus factor or domain. $NusE^{\Delta}$ interacts with the β subunit, but the exact binding site has no been identified yet. Protein Data Bank (PDB) codes: RNAP, 4KMU; NusA-NTD, 2KWP; NusA-AR2, 1WCN; NusB:$NusE^{\Delta}$, 3D3B; NusG-NTD, 2K06.

[^{1}H,^{15}N]-HSQC spectrum of ^{15}N-NusA-AR2 changed dramatically when isolated α was added (Fig. 5b), which verifies this interaction. ^{15}N-NusA-AR2 resonances corresponding to amino acid residues known to be located in the αCTD binding surface disappeared[5]. The signal intensity was only slightly diminished in the presence of β, and the spectrum of ^{15}N-NusA-AR2 was completely unaltered upon addition of β' or ω (Fig. 5c,d and Supplementary Fig. 5). Hence, we conclude that NusA-AR2 binds specifically to the α subunit (Fig. 3). Weak binding of NusA-AR1 to β was observed, just as for NusA-AR2. These interactions, however, may be unspecific due to the acidity of the AR domains[33].

Together with the CD spectra these interaction studies suggest that all subunits are functional and consequently correctly folded, although we cannot exclude that regions not interacting with NusA or NusG are not fully intact. Conventional NMR techniques thus allow qualitative studies of the interaction of RNAP with various transcription regulators, setting the stage for further biochemical and structural investigations.

Transcription factor NusE attaches to the β subunit. NusE is able to bind directly to RNAP, an interaction that is suggested to be involved in antitermination[21]. Thus, we asked which RNAP subunit was the target of NusE.

As NusE is only poorly soluble and tends to aggregate, we expressed and purified a NusE variant, $NusE^{\Delta}$, in which the ribosome binding loop is replaced by a single Ser, in complex with NusB[37]. $RNAP^{native}$ or the individual RNAP subunits were added to the NusB:^{15}N-$NusE^{\Delta}$ complex. While addition of α and ω had no effect on the 1D [^{1}H,^{15}N]-HSQC spectrum of ^{15}N-$NusE^{\Delta}$ (Supplementary Fig. 6a,b), $RNAP^{native}$ addition led to a loss of signals indicating binding of $NusE^{\Delta}$ to RNAP (Fig. 6a). A similar signal loss was obtained upon addition of β, demonstrating the formation of the $NusE^{\Delta}$:β complex (Fig. 6b). When β' was added to NusB:^{15}N-$NusE^{\Delta}$, the signal intensity was reduced by approximately 50%, an effect we attribute to weak or unspecific binding (Fig. 6c). To exclude the possibility that NusB alone binds to RNAP we performed a titration experiment with ^{15}N-NusB and $RNAP^{native}$ resulting in an unaltered

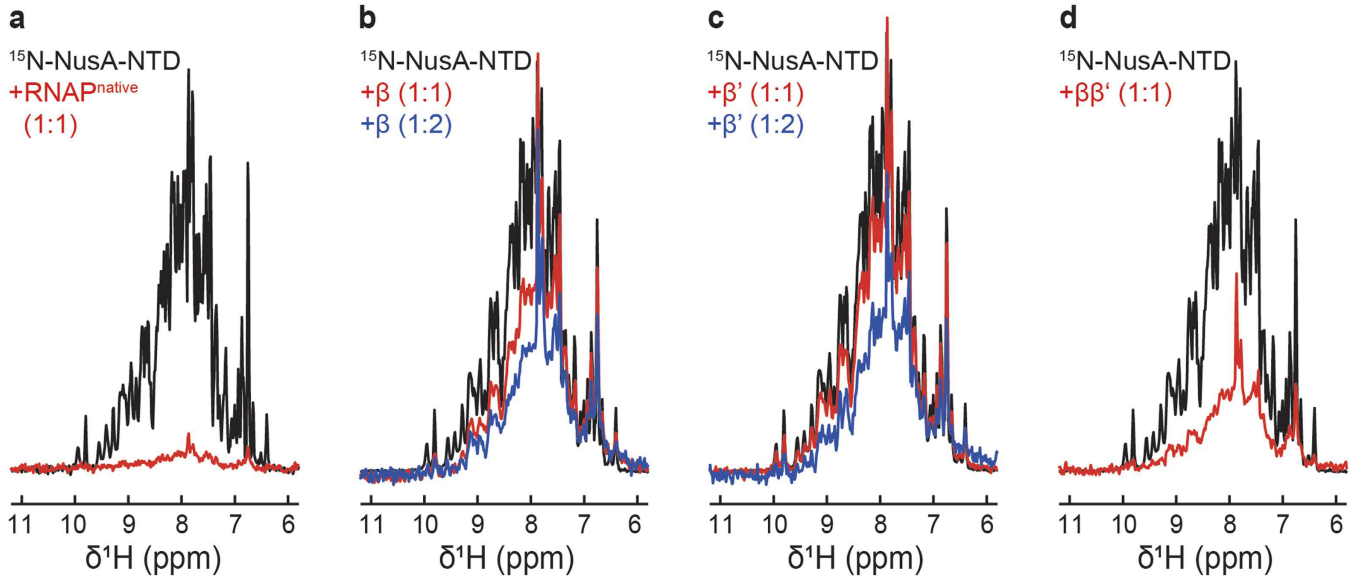


Figure 4. NusA-NTD interaction with RNAP, β, β' and ββ'. 1D [^{1}H,^{15}N]-HSQC spectra of the amide region of 30 μM ^{15}N-NusA-NTD in the absence, black, and in the presence of (**a**) RNAP^{native}, (**b**) β subunit, (**c**) β' subunit, or (**d**) ββ' complex; red, equimolar concentrations; blue, 1:2 molar ratio.

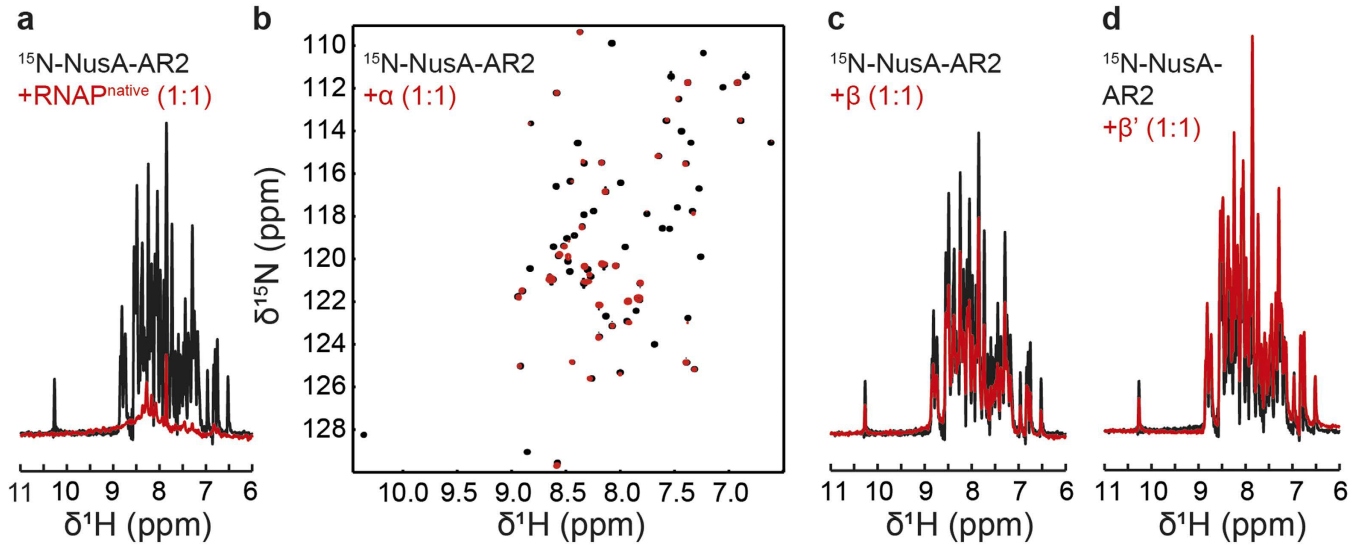


Figure 5. NusA-AR2 interaction with RNAP, α, β, and β'. [^{1}H,^{15}N]-HSQC spectra of the amide region of 30 μM ^{15}N-NusA-AR2 in the absence, black, and in the presence of equimolar concentrations, red, of (**a**) RNAP^{native} (1D spectra), (**b**) α subunit (2D spectra), (**c**) β subunit (1D spectra), or (**d**) β' subunit (1D spectra).

spectrum (Supplementary Fig. 6c). Crosslinking experiments using NusB:NusE$^{\Delta}$ and His$_6$-tagged RNAP^{native} in the presence of paraformaldehyde confirmed the formation of the NusB:NusE$^{\Delta}$:RNAP complex (Supplementary Fig. 6d,e). Addition of RNAP^{native} to ^{15}N-NusB:NusE$^{\Delta}$ led to a dramatic decrease of ^{15}N-NusB signal intensity, indicating that NusB is not released upon binding of NusE$^{\Delta}$ to RNAP (Fig. 6d). Thus, the NusB:NusE$^{\Delta}$ complex directly binds to RNAP *via* NusE$^{\Delta}$ and the β subunit is probably the key target of NusE$^{\Delta}$ (Fig. 3).

Although this might imply that the ribosome could directly interact with RNAP as NusE is part of the 30S subunit, we consider this scenario unlikely as the resulting supramolecular RNAP:ribosome complex would be very rigid and consequently gene expression would probably be impaired. Thus, we propose that the NusE:RNAP interaction might be involved in transcription antitermination as suggested earlier[21].

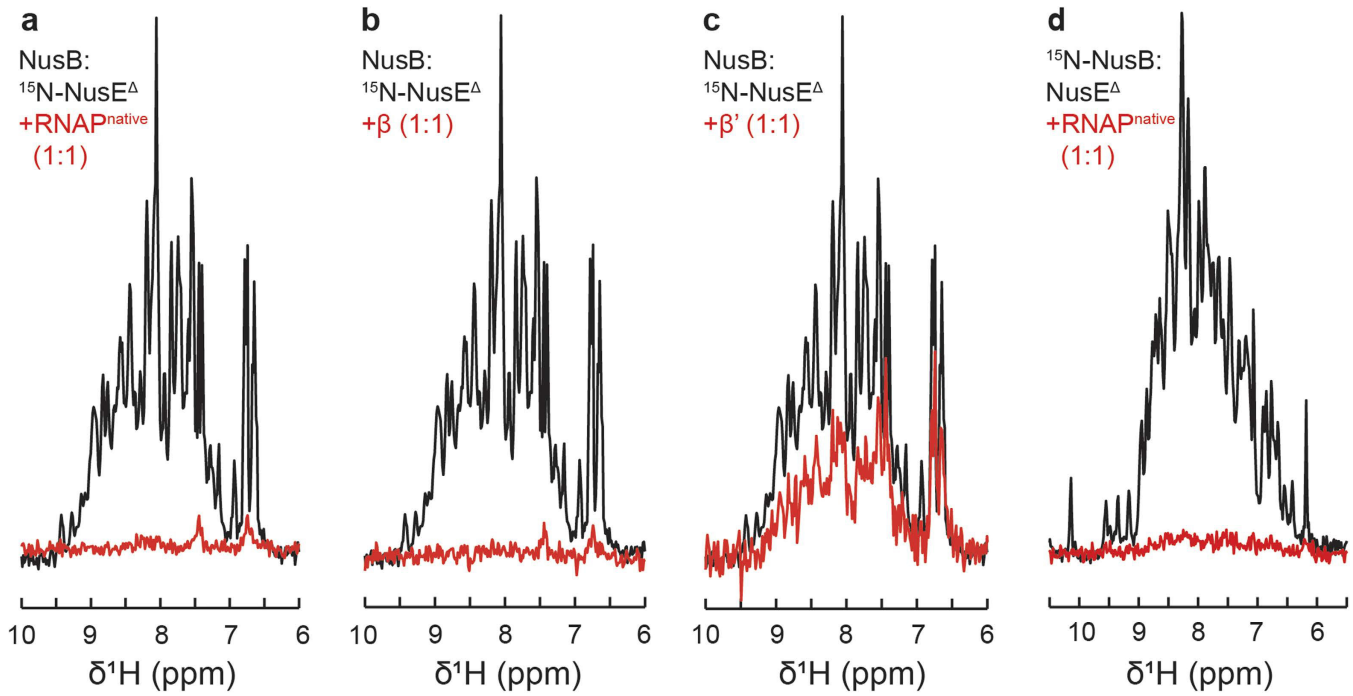


Figure 6. Interaction of NusE$^{\Delta}$ with RNAP, β, and β'. 1D [^{1}H,^{15}N]-HSQC spectra of the amide region of 30 μM NusB:^{15}N-NusE$^{\Delta}$ in the absence, black, and in the presence of equimolar concentrations, red, of (**a**) RNAP^{native}, (**b**) β subunit, or (**c**) β' subunit. (**d**) 1D [^{1}H,^{15}N]-HSQC spectra of the amide region of 30 μM ^{15}N-NusB:NusE$^{\Delta}$ in the absence, black, and in the presence of RNAP^{native} in equimolar concentration, red.

The isolated ω subunit does not interact with the isolated β′ subunit. The ω subunit of RNAP was proposed to have an essential function in folding of β' and in preventing it from aggregation as well as in promoting the assembly of $\alpha_2\beta$ with β'ω[7]. The signals of ^{15}N-ω are not diminished significantly upon addition of β', indicating that the two proteins do not interact (Supplementary Fig. 7). Yet, ω co-eluted with the other RNAP subunits in Ni^{2+} affinity chromatography after assembly (Fig. 1b), and ω was present in active RNAP. Thus, we conclude that ω binds only to unfolded or partially folded β'. Together with the analysis of its secondary structure (Fig. 1e and Supplementary Fig. 2) this, in turn, suggests that ω adopts its properly folded state either during RNAP assembly or during folding of β'.

NMR studies of RNAP. The [^{1}H,^{13}C]-TROSY heteronuclear multiple quantum coherence (HMQC) spectrum of deuterated RNAP^{native} with ^{1}H,^{13}C-labeled Ile, Leu, and Val methyl groups shows high signal dispersion, typical for a folded protein (Fig. 7a). However, owing to the size of RNAP (287 Val, 230 Ile, 349 Leu), many signals overlap.

Numerous αCTD signals could be assigned in RNAP^{native} by superposition of a [^{1}H,^{13}C]-HSQC spectrum of ^{13}C,^{15}N-αCTD and the spectrum of RNAP^{native} labeled as above (Fig. 7a), as the αCTD signals in RNAP^{native} are of higher intensity than signals of the rest of the RNAP due to the fact that this domain is flexibly connected to RNAP. A similar approach was used to assign signals in the RNAP^{native} spectrum that belong to the β' subunit (Fig. 7b). In this case, β' was deuterated and contained ^{1}H,^{13}C-labeled methyl groups of Ile, Leu, and Val residues. The signals of the isolated β' subunit are widely dispersed, and several of the RNAP^{native} signals can be assigned clearly to the β' subunit, since the chemical shifts are almost identical in the two spectra.

Addition of unlabeled NusG-NTD to methyl group labeled β' led to a significant decrease of some β' signals (Fig. 7c), indicating that the corresponding residues are affected by NusG-NTD binding. Two Ile and two Leu residues, which give rise to two and four signals in the Ile (^{13}C, 9–16 ppm) and Val/Leu (^{13}C, 17–29 ppm) region, respectively, are positioned directly in the NusG-NTD interaction site of the β'CH (Supplementary Fig. 8), matching the number of significantly affected β' signals. Other Ile, Leu, and Val residues are located in the vicinity of the interaction site and are probably affected by NusG-NTD binding as well (Supplementary Fig. 8). Hence, we conclude that the separately expressed and purified β' subunit is indeed functional in NusG-NTD binding.

In order to reduce the number of signals in the spectrum of methyl group labeled RNAP^{native}, we specifically labeled only the Ile, Val, and Leu methyl groups of the β' subunit with ^{1}H,^{13}C while all other residues of β' as well as the other subunits were deuterated (Fig. 7d). The signals in the resulting spectrum are as well dispersed as the signals of isolated β' (Fig. 7b), but new signals appear. Hence, by comparing the spectrum of methyl group labeled β' in RNAP with the one of methyl group labeled RNAP^{native}

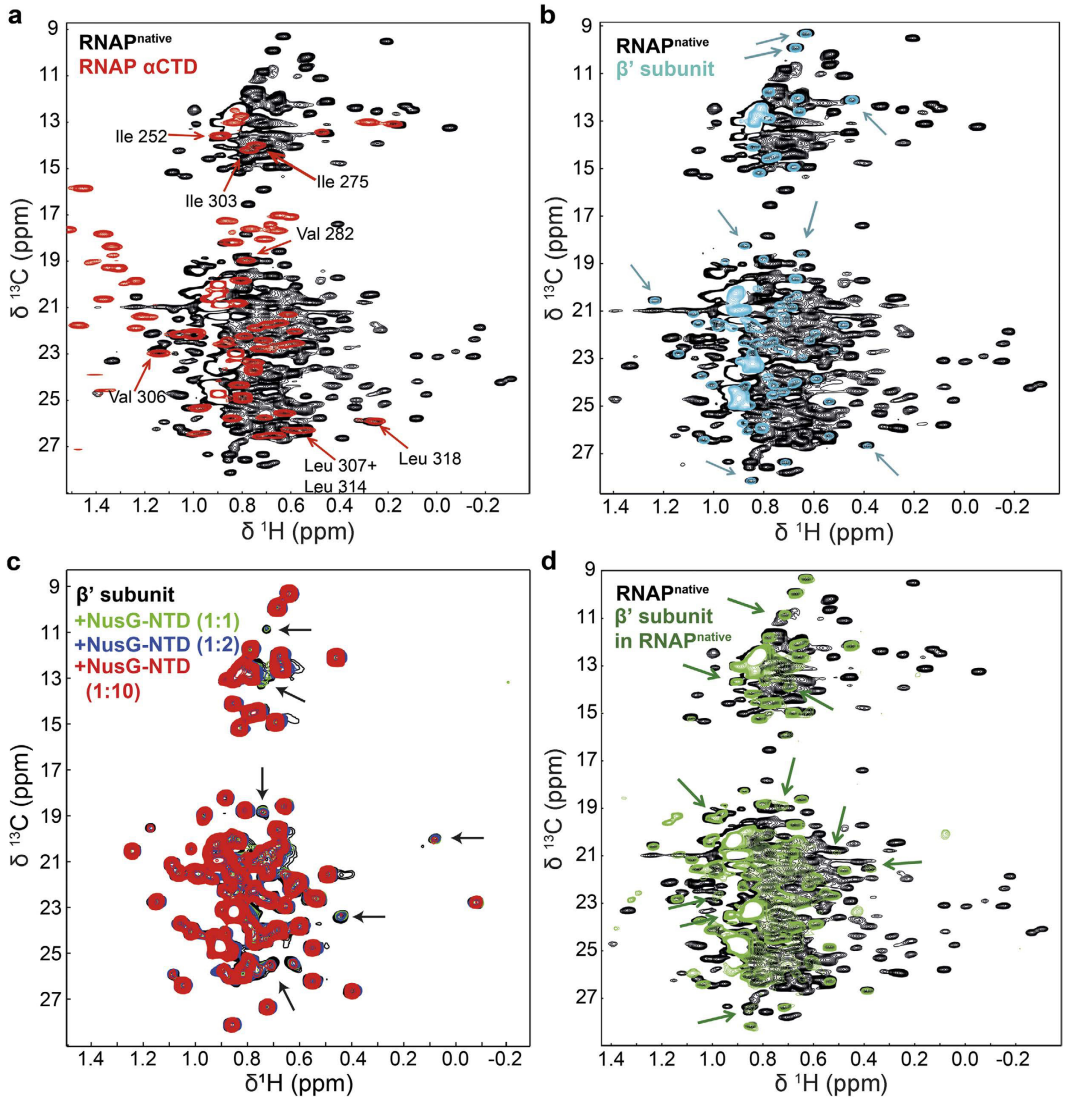


Figure 7. NMR studies of RNAP. C-H correlation spectra of $^{15}N,^{13}C$-RNAP αCTD; methyl group labeled RNAP^{native}; methyl group labeled β'; and methyl group labeled β' in reconstituted RNAP (other subunits deuterated). (**a**) Superposition of a [$^{1}H,^{13}C$]-HMQC spectrum of 30 μM RNAP^{native}, black, and a [$^{1}H,^{13}C$]-HSQC spectrum of 700 μM RNAP αCTD, red. Directly assigned peaks are labeled. (**b**) Superposition of [$^{1}H,^{13}C$]-HMQC spectra of 30 μM RNAP^{native}, black, and 2 μM β', cyan. Example peaks with identical chemical shift in RNAP^{native} and free β' are indicated by blue arrows. (**c**) Superposition of [$^{1}H,^{13}C$]-HMQC spectra of 2 μM β', before, black, and after addition of unlabeled NusG-NTD in a 1:1, 1:2, and 1:10 molar ratio (green, blue, and red, respectively). Arrows indicate signals that decrease significantly upon NusG-NTD addition. (**d**) Superposition of the [$^{1}H,^{13}C$]-HMQC spectra of RNAP^{native}, black, and β' in reconstituted RNAP, green. β' signals identical to signals of RNAP^{native} and those whose positions differ in free β' and β' in reconstituted RNAP are indicated by green arrows.

(Fig. 7d) more signals of RNAP^{native} could be assigned to the β' subunit than using the spectrum of isolated methyl group labeled β'. This is probably due to the fact that here β' was in its physiological context.

Thus, this work demonstrates that even heterooligomeric systems as complex as RNAP can be tackled by NMR spectroscopy, and, moreover, that intra- and interdomain dynamics and the transient interaction with regulatory factors can be studied. In fact, we expect that further refinement of the method we presented here by, e.g., specific labeling of parts of the RNAP subunits will lead to very major contributions to detailed studies of transcription factor:RNAP interactions by solution state NMR spectroscopy.

Methods

Assembly and purification of the RNAP and the ββ′ complex. All RNAP subunit genes were expressed separately (see Supplementary Methods), with the β' subunit being produced as a fusion protein carrying an N-terminal His_6 tag. Cell pellets from equal volumes of cell cultures of the individual

subunits were resuspended in denaturing lysis buffer (50 mM tris(hydroxymethyl)aminomethane (Tris)/HCl, pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 8 M urea, 1 mM dithiothreitol (DTT)) and combined. Cell lysis was performed with a microfluidizer, and the cell lysate was stirred for 1 h at room temperature. For the assembly of RNAP, the lysate was dialyzed against lysis buffer with decreasing urea concentrations (4 M, 1 M, 0.5 M, 0 M; 2 h each buffer at 4 °C). Finally, the extract was dialyzed overnight against buffer A (50 mM Tris/HCl, pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 10 mM imidazole). The dialysate was incubated for 1 h at 30 °C, centrifuged at 12,000 x *g* and 4 °C for 30 min, and the supernatant was applied to a HisTrap HP column (GE Healthcare, Munich, Germany). After washing with buffer A, elution was performed using a constant gradient with imidazole concentrations increasing up to 1 M in buffer A. RNAP containing fractions were combined and dialyzed against 50 mM Tris/HCl, pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 0.5 mM EDTA, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM DTT at 4 °C overnight. The protein solution was then concentrated by ultrafiltration (molecular weight cut-off (MWCO) = 10 kDa) and applied to a HiLoad 16/600 Superdex 200 pg column (GE Healthcare, Munich, Germany). The fractions of the main peaks from the SEC were concentrated separately by ultrafiltration (MWCO = 10 kDa), frozen in liquid nitrogen and stored at -80 °C.

The assembly and purification of ββ' was performed according to the protocol used for RNAP. However, the incubation step after removing urea was omitted and 37 mg protein were obtained from 1 l cultures.

Protein production and purification of RNAP^{native}. The genes for all subunits were expressed on the same plasmid from one promoter as an operon. Expression and purification are based on a slightly modified published protocol[38]. For the overexpression, the LB/M9 minimal medium[39,40] supplemented with ampicillin (100 μg/ml) was inoculated with a preculture to an OD_{600} of 0.03 and cells were grown at 37 °C. At OD_{600} ~ 0.2 the temperature was lowered to 16 °C. After 90 min, overexpression was induced by 0.5 mM IPTG and cells were grown overnight. The first purification step was performed using Ni-NTA Superflow cartridges (QIAGEN, Hilden) on an ÄKTA purifier system.

Isotopic labeling of proteins. ^{15}N- and ^{15}N-, ^{13}C-labeled proteins were obtained by growing *E. coli* in M9 minimal media[39,40] upon respective addition of $(^{15}NH_4)_2SO_4$ (Campro Scientific, Berlin, Germany) and ^{13}C-D-glucose (Spectra Stable Isotopes, Columbia, MD, USA) as the only nitrogen and carbon source. Expression and purification was the same as for proteins produced in LB medium (see Supplementary Methods).

The protocol for deuteration of proteins in which the methyl groups of Ile, Leu and Val residues are ^{1}H,^{13}C-labeled is based on a published method[26]. First, cells were slowly accustomed to D_2O (Campro Scientific, Berlin, Germany) in precultures (LB, M9 minimal medium in H_2O, M9 with 25% (v/v), 50% (v/v) and 100% (v/v) D_2O consecutively). In the 100% D_2O preculture and the main culture, deuterated glucose (Campro Scientific, Berlin, Germany) was added as the sole carbon source. The time for gene expression was doubled as compared to expression in H_2O. For methyl group labeling, 60 mg/L cell culture 2-keto-3-d_3-4-^{13}C-butyrate (isoleucine; Eurisotop, St. Aubin Cedex, France) and 100 mg/L cell culture 2-keto-3-methyl-d_3-3-d_1-4-^{13}C-butyrate (valine, leucine; Eurisotop, St. Aubin Cedex, France) were added 1 h prior to induction. To produce completely deuterated proteins without ^{13}C or ^{15}N label the final step was omitted.

RNAP activity assay. As RNAP^{native} and the assembled RNAPs do not contain the σ subunit for binding of a promoter region, a nucleic acid scaffold consisting of a template DNA without a promoter, a non-template DNA, and an RNA primer was used for the activity assay. The 24mer template (T24, 5'-GCCGCGCGCTTGCGGTCTGTCCC-3') and 14mer non-template (NT14, 5'-AACGCCAGACAGGG-3') DNA oligos overlap only partially to form a short downstream duplex DNA. The other end of T24 is complementary to the 16mer RNA primer (R16, 5'-GAGUCUGCGGCGCGCG-3') that is labeled with 6-carboxyfluorescein (6-FAM) at the 5'-end for visualization. These oligonucleotides are identical with the ones used to obtain the crystal structure of *Thermus thermophilus* elongation complex PDB code: 2O5I[41]).

The reactions were carried out in 20 mM Tris/HCl, pH 8.0, 40 mM KCl, 0.1 mM EDTA, 0.1 mM DTT. For a 50 μL reaction, 12 pmol of T24 and 10 pmol of R16 were mixed, heated to 75 °C for 5 min, and cooled to RT. 12 pmol of NT14 were added and incubated for 10 min at RT. 20 pmol RNAP were added and again incubated at RT for 10 min. To start the activity assay, 5 mM $MgCl_2$ and 2.5 μM of each NTP were added and incubated at 37 °C for 5 min. ATP and CTP were added for an RNA extension of 3 nt. When GTP was also added, the RNA was extended by 14 nt. The reaction samples were analyzed on a 20% (w/v) polyacrylamide/8.3 M urea gel and fluorescence was visualized by a Stella Imaging System (raytest, Straubenhardt, Germany). To compare the activities of RNAP^{active} and RNAP^{native}, the intensity of the strongest band from extended RNA was divided by the intensity of non-extended RNA primer.

CD measurements. Far-UV CD spectra were recorded on a Jasco J-810 spectropolarimeter (Jasco, Gross-Umstadt, Germany) with protein concentrations between 0.5 and 10 μM in 10 mM potassium phosphate buffer, pH 7.5. Spectra were accumulated ten times at 20 °C with an increment of 0.2 nm.

Measured ellipticity [Θ] was normalized against the protein concentration c in mM, the path length d in cm and the number of amino acids N according to equation (1).

$$[\Theta]_{MRW} = 100 \cdot [\Theta] \cdot /(c \cdot d \cdot N) \quad (1)$$

NMR spectroscopy. NMR measurements were conducted at 25 °C on Bruker *Avance* 600 MHz, 700 MHz, and 800 MHz spectrometers, the latter two equipped with cryogenically cooled probes. The interaction studies of transcription factors with RNAP^{native} and individual subunits were carried out in 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5, 50 mM NaCl, 5% (v/v) glycerol, 0.5 mM EDTA, 1 mM DTT. Methyl group and ^{15}N-labeled proteins were in 25 mM HEPES, pH 7.5, 50 mM NaCl, 5% (v/v) glycerol, 0.5 mM EDTA, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM DTT while [^{15}N,^{13}C]-αCTD was in 10 mM potassium phosphate, pH 6.4, 50 mM NaCl, 1 mM β-mercaptoethanol. 2D spectra were visualized and analyzed using NMRView[42], 1D spectra by Matlab (The MathWorks, Inc., Version 7.1.0.183). To compare different 1D [^{1}H,^{15}N]-HSQC spectra, the intensity was divided by the number of scans and the protein concentration.

Transverse relaxation rates of amide protons were determined with two-point measurements, using 1D [^{1}H,^{15}N]-HSQC experiments including a spin echo in the first insensitive nuclei enhancement by polarization transfer (INEPT) step[43]. Samples contained either 40 μM ^{15}N-NusA-NTD or 40 μM ^{15}N-NusA-NTD and an equimolar amount of β, β' or ββ' in 25 mM HEPES, pH 7.5, 50 mM NaCl, 5% (v/v) glycerol, 0.5 mM EDTA, 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM DTT. For the experiment with free NusA-NTD the difference between the two time points for the spin-echo experiments (Δt) was 10 ms, while it was 5 ms for all other measurements. The population-averaged observed R_2 was determined according to equation (2).

$$R_2{}^{obs} = x_{unbound} \cdot R_2{}^{NTD} + (1 - x_{unbound}) \cdot R_2{}^{NTD+partner} \quad (2)$$

$R_2{}^{NTD}$ is R_2 of free NusA-NTD and $R_2{}^{NTD+partner}$ is R_2 of the complex of NusA-NTD and β, β' or ββ'. Thus, the fraction of unbound NusA-NTD ($x_{unbound}$) was calculated using equation (3).

$$x_{unbound} = (R_2{}^{obs} - R_2{}^{NTD+partner})/(R_2{}^{NTD} - R_2{}^{NTD+partner}) \quad (3)$$

$R_2{}^{NTD}$ corresponds to $R_2{}^{obs}$of NusA-NTD and was experimentally determined to 50 s^{-1}. $R_2{}^{NTD+partner}$ was estimated based on the proportionality of R_2 and the molecular mass ($R_2{}^{NTD+\beta}$: 500 s^{-1}, $R_2{}^{NTD+\beta'}$: 500 s^{-1}, $R_2{}^{NTD+\beta\beta'}$: 1000 s^{-1}).

References

1. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9,** 85–98 (2011).
2. Burgess, R. R. Separation and characterization of the subunits of ribonucleic acid polymerase. *J. Biol. Chem.* **244,** 6168–6176 (1969).
3. Dove, S. L. & Hochschild, A. Conversion of the ω subunit of *Escherichia coli* RNA polymerase into a transcriptional activator or an activation target. *Genes Dev.* **12,** 745–754 (1998).
4. Jeon, Y. H., Yamazaki, T., Otomo, T., Ishihama, A. & Kyogoku, Y. Flexible linker in the RNA polymerase α subunit facilitates the independent motion of the C-terminal activator contact domain. *J. Mol. Biol.* **267,** 953–962 (1997).
5. Schweimer, K. *et al.* NusA interaction with the α-subunit of *E. coli* RNA polymerase is *via* the UP-element site and releases autoinhibition. *Structure* **19,** 945–954 (2011).
6. Ito, K., Iwakura, Y. & Ishihama, A. Biosynthesis of RNA polymerase in *Escherichia coli*. III. Identification of intermediates in the assembly of RNA polymerase. *J. Mol. Biol.* **96,** 257–271 (1975).
7. Ghosh, P., Ishihama, A. & Chatterji, D. *Escherichia coli* RNA polymerase subunit ω and its N-terminal domain bind full-length β' to facilitate incorporation into the $\alpha_2\beta$ subassembly. *Eur. J. Biochem.* **268,** 4621–4627 (2001).
8. Mustaev, A. *et al.* Modular organization of the catalytic center of RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 6641–6645 (1997).
9. Daube, S. S. & von Hippel, P. H. Interactions of *Escherichia coli* σ^{70} within the transcription elongation complex. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8390–8395 (1999).
10. Mukhopadhyay, J. *et al.* Translocation of σ^{70} with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* **106,** 453–463 (2001).
11. Mooney, R. A., Artsimovitch, I. & Landick, R. Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. *J. Bacteriol.* **180,** 3265–3275 (1998).
12. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7090–7095 (2000).
13. 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).
14. 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).
15. Borukhov, S., Lee, J. & Laptenko, O. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol. Microbiol.* **55,** 1315–1324 (2005).
16. Roberts, J. W., Shankar, S. & Filter, J. J. RNA polymerase elongation factors. *Annu. Rev. Microbiol.* **62,** 211–233 (2008).
17. Santangelo, T. J. & Artsimovitch, I. Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.* **9,** 319–329 (2011).
18. Burmann, B. M. *et al.* A NusE:NusG complex links transcription and translation. *Science* **328,** 501–504 (2010).

19. Mooney, R. A., Schweimer, K., Rösch, P., Gottesman, M. E. & 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).
20. Friedman, D. I., Schauer, A. T., Baumann, M. R., Baron, L. S. & Adhya, S. L. Evidence that ribosomal protein S10 participates in control of transcription termination. *Proc. Natl. Acad. Sci. U. S. A.* **78,** 1115–1118 (1981).
21. Mason, S. W. & Greenblatt, J. Assembly of transcription elongation complexes containing the N protein of phage λ and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. *Genes Dev.* **5,** 1504–1512 (1991).
22. Sekine, S., Tagami, S. & Yokoyama, S. Structural basis of transcription by bacterial and eukaryotic RNA polymerases. *Curr. Opin. Struct. Biol.* **22,** 110–118 (2012).
23. LeMaster, D. M. & Richards, F. M. NMR sequential assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteriation. *Biochemistry* **27,** 142–150 (1988).
24. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13585–13590 (1998).
25. Tugarinov, V. & Kay, L. E. Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723 residue enzyme malate synthase G. *J. Mol. Biol.* **327,** 1121–1133 (2003).
26. Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445,** 618–622 (2007).
27. Heil, A. & Zillig, W. Reconstitution of bacterial DNA-dependent RNA polymerase from isolated subunits as a tool for the elucidation of the role of the subunits in transcription. *FEBS Lett.* **11,** 165–168 (1970).
28. Borukhov, S. & Goldfarb, A. Recombinant *Escherichia coli* RNA polymerase: purification of individually overexpressed subunits and *in vitro* assembly. *Protein Expr. Purif.* **4,** 503–511 (1993).
29. Tang, H., Severinov, K., Goldfarb, A. & Ebright, R. H. Rapid RNA polymerase genetics: one-day, no-column preparation of reconstituted recombinant *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 4902–4906 (1995).
30. Gentry, D., Xiao, H., Burgess, R. & Cashel, M. The ω subunit of *Escherichia coli* K-12 RNA polymerase is not required for stringent RNA control *in vivo*. *J. Bacteriol.* **173,** 3901–3903 (1991).
31. Mukherjee, K., Nagai, H., Shimamoto, N. & Chatterji, D. GroEL is involved in activation of *Escherichia coli* RNA polymerase devoid of the ω subunit *in vivo*. *Eur. J. Biochem.* **266,** 228–235 (1999).
32. Katayama, A., Fujita, N. & Ishihama, A. Mapping of subunit-subunit contact surfaces on the β'subunit of *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **275,** 3583–3592 (2000).
33. Eisenmann, A., Schwarz, S., Prasch, S., Schweimer, K. & Rösch, P. The *E. coli* NusA carboxy-terminal domains are structurally similar and show specific RNAP- and λN interaction. *Protein Sci.* **14,** 2018–2029 (2005).
34. Worbs, M., Bourenkov, G. P., Bartunik, H. D., Huber, R. & Wahl, M. C. An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Mol. Cell* **7,** 1177–1189 (2001).
35. Toulokhonov, I., Artsimovitch, I. & Landick, R. Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* **292,** 730–733 (2001).
36. Yang, X. *et al.* The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep.* **10,** 997–1002 (2009).
37. Luo, X. *et al.* Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. *Mol. Cell* **32,** 791–802 (2008).
38. Artsimovitch, I., Svetlov, V., Murakami, K. S. & Landick, R. Co-overexpression of *Escherichia coli* RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J. Biol. Chem.* **278,** 12344–12355 (2003).
39. Meyer, O. & Schlegel, H. G. Biology of aerobic carbon monoxide-oxidizing bacteria. *Annu. Rev. Microbiol.* **37,** 277–310 (1983).
40. Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning - A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1994).
41. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448,** 157–162 (2007).
42. Johnson, B. A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278,** 313–352 (2004).
43. Iwahara, J., Tang, C. & Marius Clore, G. Practical aspects of ^{1}H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.* **184,** 185–195 (2007).

Acknowledgements

We thank Ramona Heißmann and Ulrike Persau for excellent technical support and Irina Artsimovitch for providing plasmids and for helpful discussions. The work was supported by grant Ro 617/17-1 (to P.R.) from the Deutsche Forschungsgemeinschaft.

Author Contributions

P.R. initiated and supervised the project. B.M.W., J.D., M.S. designed the cloning strategies. B.M.W., J.D., M.S., S.H.K. designed the expression and purification strategies, the CD experiments and the activity assays. J.D. and S.K. designed the crosslinking experiments. J.D., K.S., M.S., S.H.K. designed the NMR experiments. J.D. and M.S. conducted the experiments. J.D., M.S., B.M.W., P.R., S.H.K. wrote the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Drögemüller, J. *et al.* Exploring RNA polymerase regulation by NMR spectroscopy. *Sci. Rep.* **5**, 10825; doi: 10.1038/srep10825 (2015).

Supplementary Information

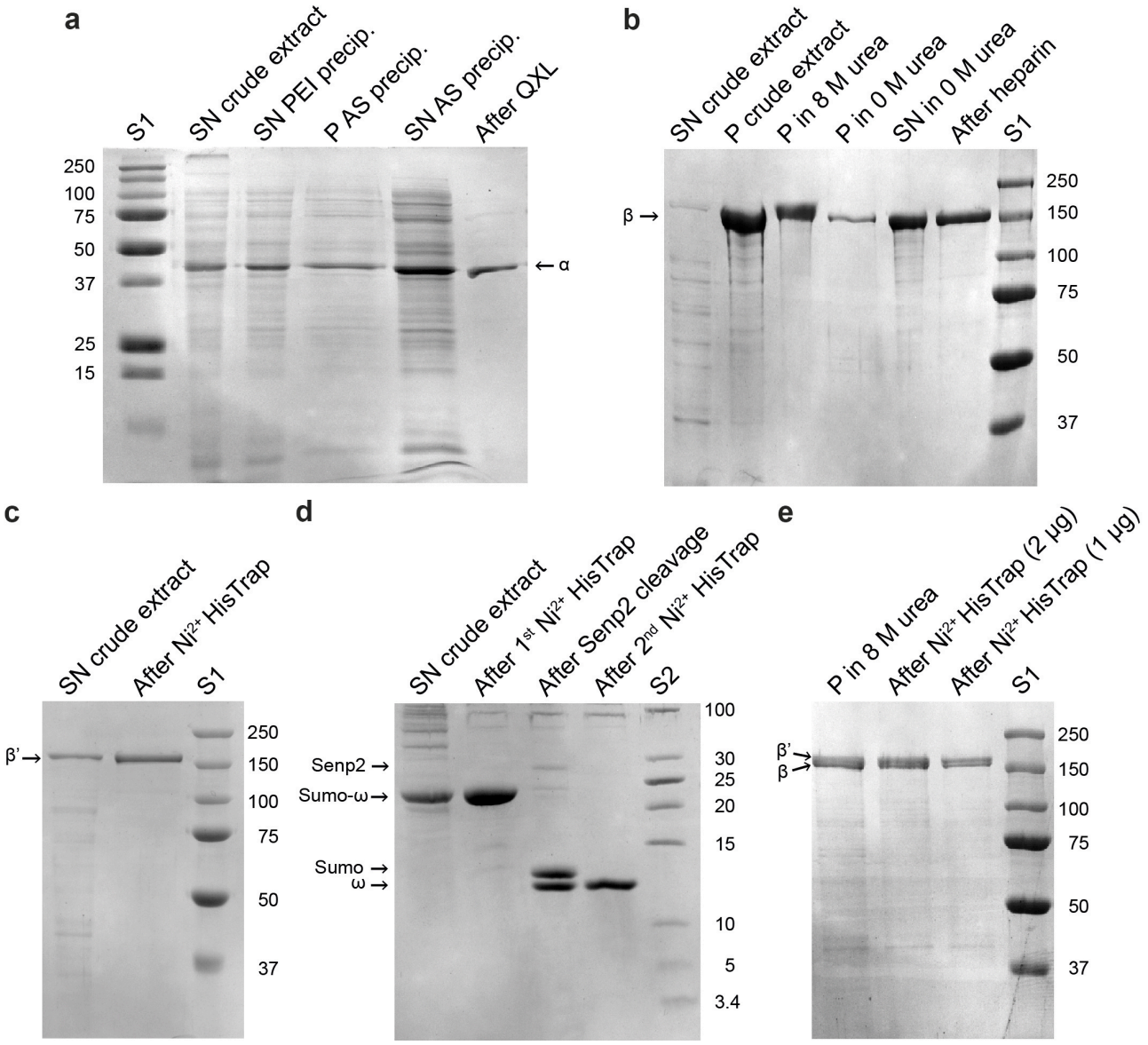
Exploring RNA polymerase regulation by NMR spectroscopy

Johanna Drögemüller[1,†], Martin Strauß[1,†], Kristian Schweimer[1], Birgitta M. Wöhrl[1], Stefan H. Knauer[1,*] & Paul Rösch[1]

[1]Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany

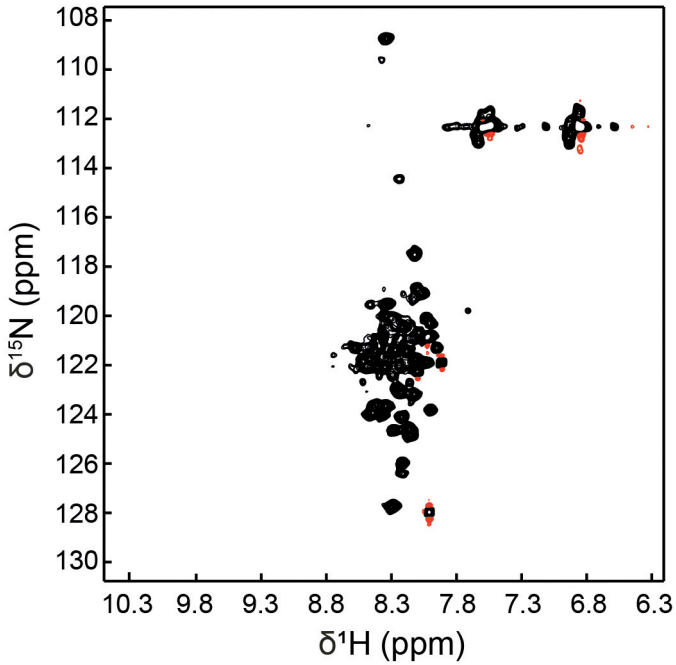
[†] These authors contributed equally to this work

Contents

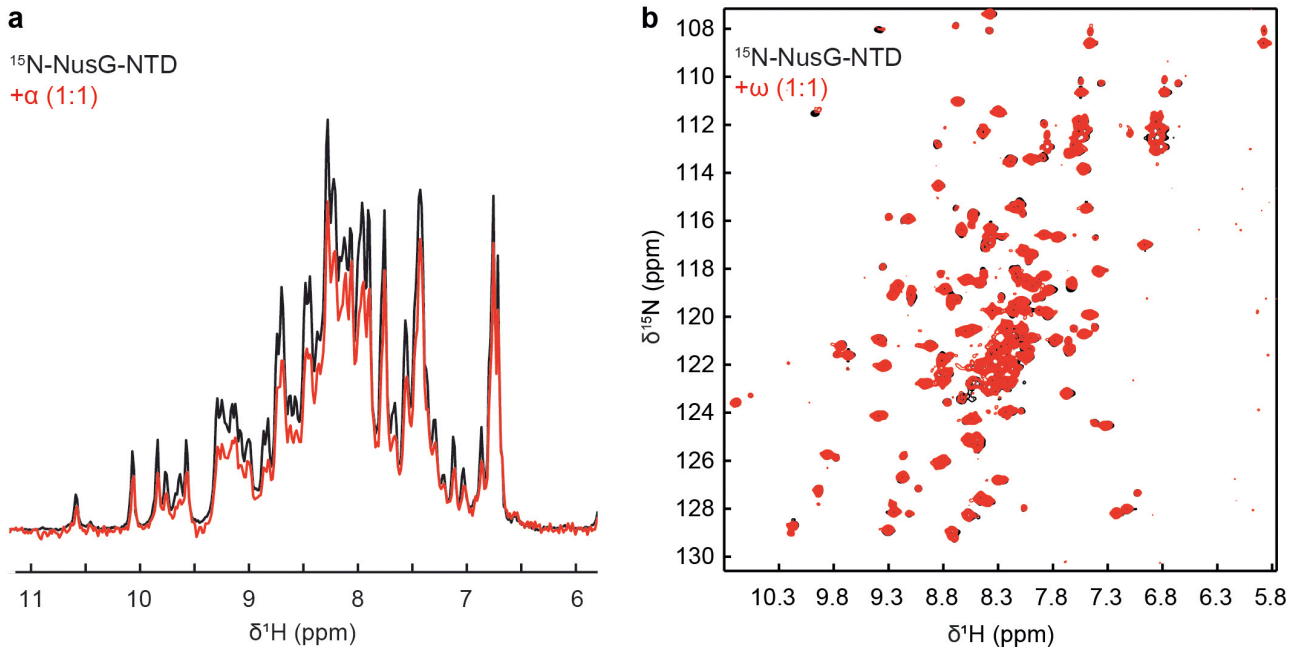


Supplementary Figure 1: Purification of individual RNAP subunits. 2 µg protein were applied to each lane. S1, Precision Plus Protein Standard (BioRad, Munich, Germany); S2, PageRuler Low Range Protein Ladder (Thermo Scientific, Schwerte, Germany); SN, supernatant; P, pellet; PEI, polyethylenimine, AS, ammonium sulfate (**a**) 19 % (w/v) SDS-polyacrylamide gel of aliquots taken from the fractions during α subunit purification after staining with Coomassie Blue. (**b**) 10 % (w/v) SDS-polyacrylamide gel of aliquots taken from the fractions during β subunit purification after staining with Coomassie Blue. (**c**) 10 % (w/v) SDS-polyacrylamide gel of aliquots taken from the fractions during β' subunit purification after staining with Coomassie Blue. (**d**) Schägger-Jagow gel[1] of aliquots taken from the fractions during ω subunit purification after staining with Coomassie

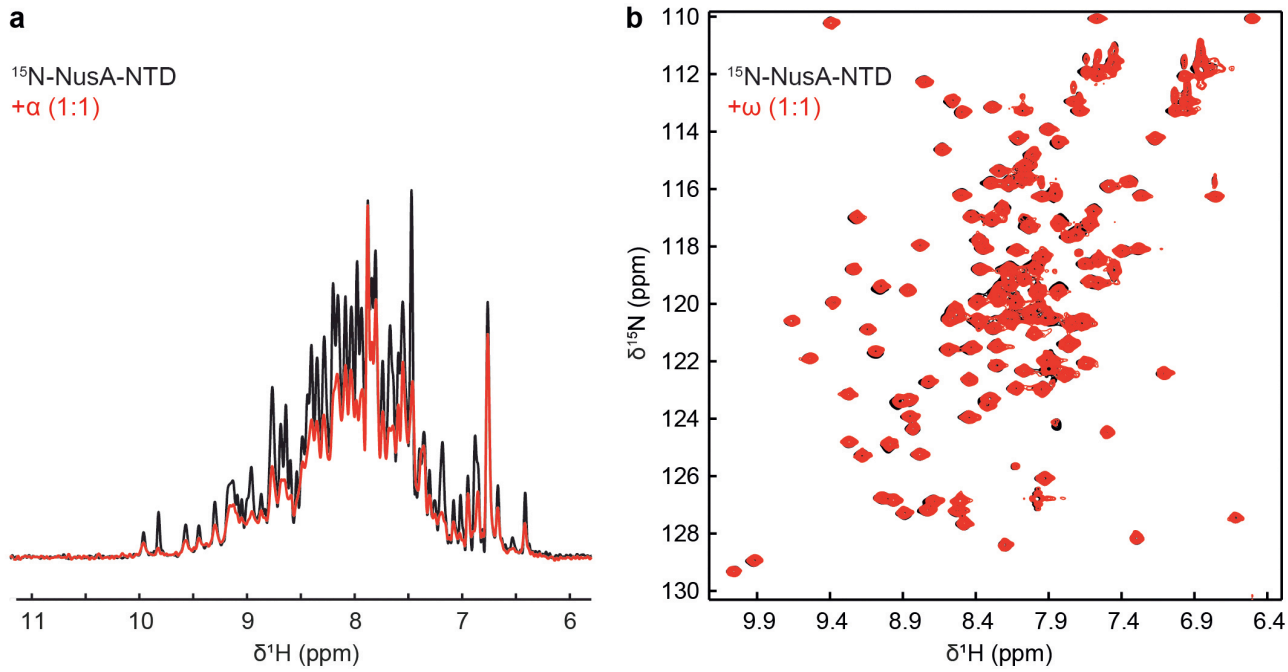
Blue. (**e**) 10 % (w/v) polyacrylamide gel of aliquots taken from the fractions during ββ' complex purification after staining with Coomassie Blue.



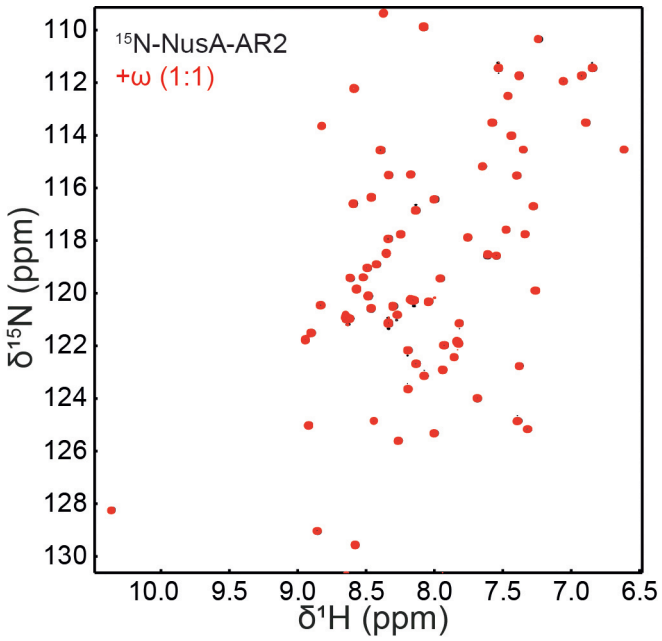
Supplementary Figure 2: [^{1}H,^{15}N]-HSQC spectrum of 300 μM ^{15}N-ω. Positive and negative signals are colored in black and red, respectively.



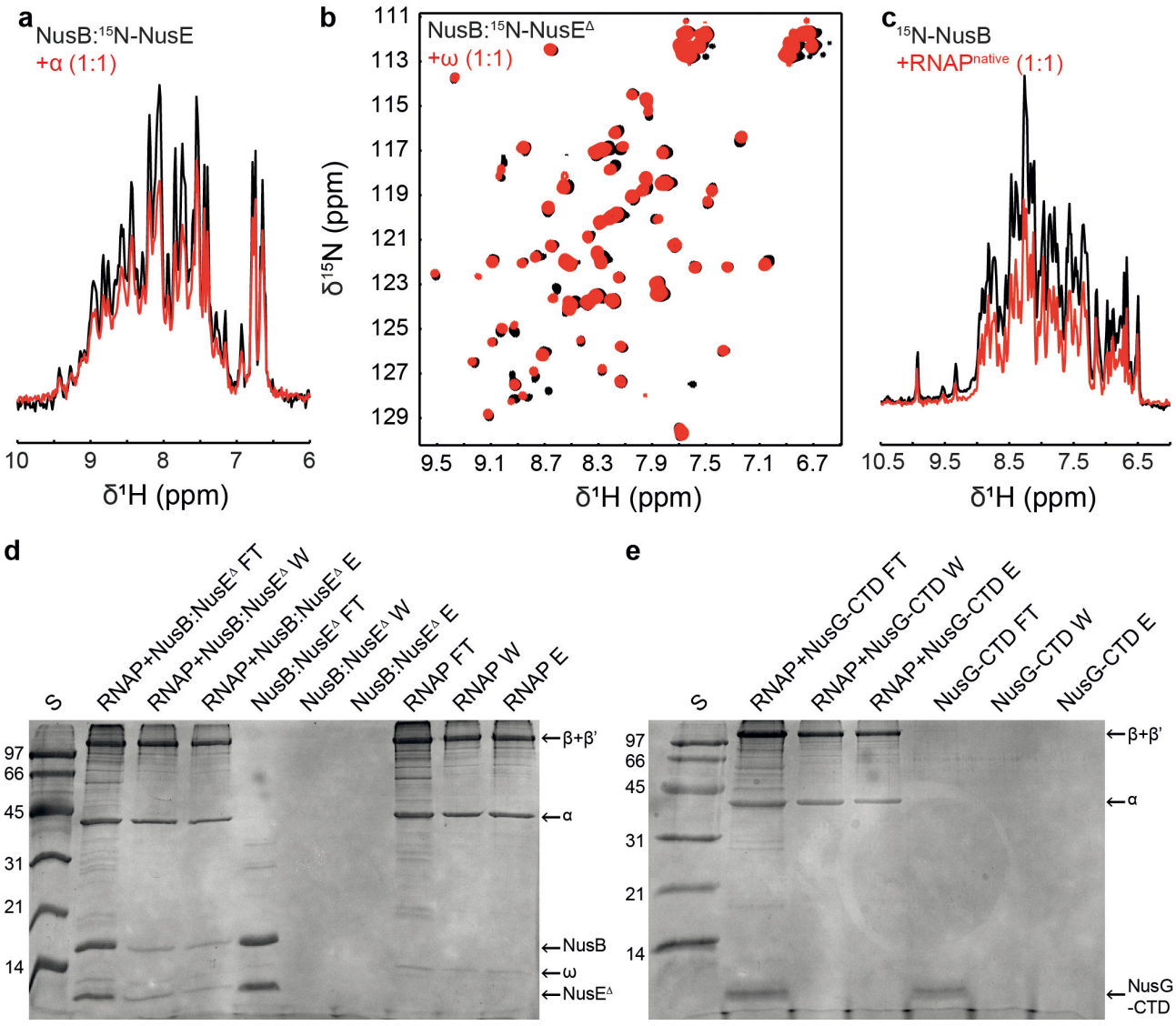
Supplementary Figure 3: NusG-NTD does not interact with isolated α or ω. (**a**) 1D [^{1}H,^{15}N]-HSQC spectra of 30 μM NusG-NTD in the absence, black, and in the presence, red, of an equimolar concentration of α. (**b**) 2D [^{1}H,^{15}N]-HSQC spectra of 100 μM NusG-NTD in the absence, black, and in the presence, red, of an equimolar concentration of ω.



Supplementary Figure 4: NusA-NTD does not interact with isolated α or ω. (**a**) 1D [^{1}H,^{15}N]-HSQC spectra of 30 μM NusA-NTD in the absence, black, and in the presence, red, of an equimolar concentration of α. (**b**) 2D [^{1}H,^{15}N]-HSQC spectra of 100 μM NusA-NTD in the absence, black, and in the presence, red, of an equimolar concentration of ω.

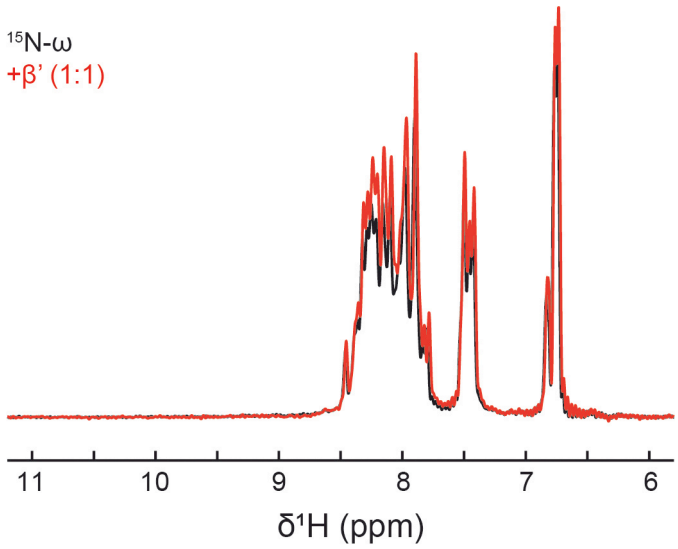


Supplementary Figure 5: NusA-AR2 does not interact with isolated ω. 2D [^{1}H,^{15}N]-HSQC spectra of 30 μM NusA-AR2 in the absence, black, and in the presence, red, of an equimolar concentration of ω.

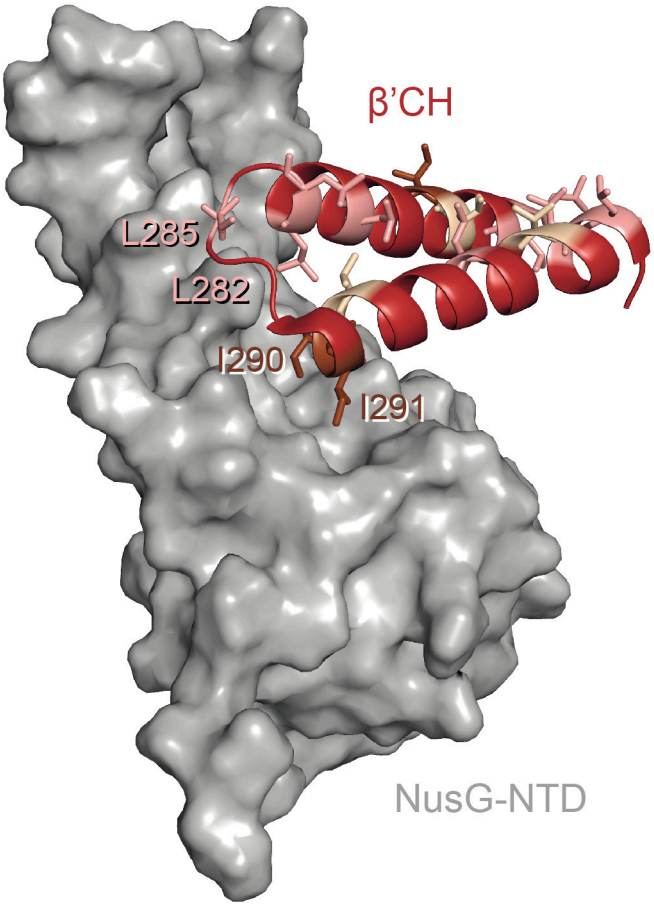


Supplementary Figure 6: Interaction studies of NusB and NusB:NusE$^{\Delta}$ with RNAP, isolated α and ω. (**a**) 1D [^{1}H,^{15}N]-HSQC spectra of 30 μM NusB:^{15}N-NusE$^{\Delta}$ in the absence, black, and in the presence, red, of an equimolar concentration of α. (**b**) 2D [^{1}H,^{15}N]-HSQC spectrum of 100 μM NusB:^{15}N-NusE$^{\Delta}$ in the absence, black, and in the presence, red, of an equimolar concentration of ω. (**c**) 1D [^{1}H,^{15}N]-HSQC spectra of 30 μM ^{15}N-NusB in the absence, black, and in the presence, red, of an equimolar concentration of RNAP^{native}. (**d,e**) Crosslinkling of RNAP and NusB:NusE$^{\Delta}$. 19 % (w/v) SDS-polyacrylamide gel after Ni^{2+} affinity chromatography and staining with Coomassie Blue. Crosslinking of RNAP and NusG-CTD was used as negative control. S: BioRad low range SDS-PAGE Standard (BioRad, Munich, Germany), FT: flow through, W: fraction of the

last washing step with 5 mM imidazole, E: eluate.



Supplementary Figure 7: The ω subunit does not interact with the β' subunit. 1D [^{1}H,^{15}N]-HSQC spectra of the amide region of 30 μM ^{15}N-ω subunit in the absence, black, and in the presence of an equimolar concentration of β', red.



Supplementary Figure 8: Model of NusG-NTD binding to the β'CH. The NusG-NTD: β'CH complex (PDB code: 2K06, NusG-NTD, surface representation; PDB code: 4KMU, β' clamp helices, ribbon representation) was modeled based on the crystal structure of *Pyrococcus furiosus* Spt4/5 binding to the RNAP clamp domain[2]. Ile (brown), Leu (pink) and Val (beige) residues in the β' clamp helices are represented as sticks.

Supplementary Methods

Cloning. Plasmids containing the genes *rpoA*, *rpoB*, *rpoC* and *rpoZ* were kindly provided by Irina Artsimovitch. *rpoB* was cloned from pIA942 into pET29b (Novagen, Madison, WI, USA) *via* *Bam*HI and *Nde*I. *rpoC* was cloned from pIA661 into pET29b *via Nde*I and *Hind*III restriction sites allowing the expression of *rpoC* with a hexahistidine tag at the C-terminus. For tagless production of *rpoZ* the gene was excised from pIA839 with its ribosome binding site *via Xba*I and *Hind*III and cloned into pET32a (Novagen, Madison, WI, USA). For expression of *rpoZ* with an N-terminal SUMO tag the *rpoZ* gene was cloned into pET28 derivative harboring the small ubiquitin-like modifier (SUMO) 1 gene *via Bam*HI and *Xho*I restriction sites.

The gene for NusA-NTD (1-125) was cloned using the Champion™ pET101 Directional TOPO© Expression Kit (Invitrogen, Carlsbad, CA, USA) with the following primers: Fwd-primer: 5'-CAC CAT GAA CAA AGA AAT TTT GGC-3'; Rev-primer: 5'-AGA ACC ACG CGG AAC CAG CAT CGC ACG TTC GGC TTC ACG-3'. The resulting *E. coli* expression vector pET101_NusA-NTD contains a C-terminal hexa-histidine tag and a thrombin cleavage site between NusA-NTD and the histidine tag.

Gene expression and protein purification. *rpoA* was expressed in *E. coli* BL21 (DE3) (Novagen, Madison, WI, USA) harboring the plasmid pIA287. Cells were grown in M9 minimal medium[3,4] containing 100 µg/ml ampicillin at 37 °C. At an optical density of 600 nm (OD_{600}) of ~ 0.7 expression was induced by 0.1 mM isopropyl-thiogalactoside (IPTG). Cells were harvested after 3 h (9,000 x *g*, 15 min, 4°C), resuspended in 50 mM Tris/HCl (pH 8.0) containing 500 mM NaCl and disrupted by a microfluidizer (Microfluidics, Newton, MA, USA). Nucleic acids were precipitated by addition of 0.6 % (v/v) polyethylenimine and removed by centrifugation (12,000 x *g,* 30 min, 4 °C). Subsequently, an ammonium sulfate precipitation (60 % (w/v)) was performed with the supernatant. After centrifugation (12,000 x *g*, 4 °C, 30 min) the supernatant was dialyzed against 20 mM Tris/HCl (pH 8.0) overnight at 4 °C and applied to a HiTrap QXL column (GE Healthcare,

Munich, Germany). After washing with 20 mM Tris/HCl (pH 8.0) elution was performed using a step gradient with increasing NaCl concentrations (0.25-1 M NaCl in 20 mM Tris/HCl (pH 8.0)). Fractions containing the target protein were combined, dialyzed against the required buffer, concentrated by ultrafiltration (VivaSpin units, molecular weight cut-off (MWCO) = 3.5 kDa, Sartorius Stedim Biotech GmbH, Göttingen, Germany) and stored at -80 °C after freezing with liquid nitrogen. 67 mg protein were obtained from a one liter culture.

rpoB was expressed in *E. coli* BL21 (DE3) harboring the pET29b/*rpoB* plasmid. Cells were grown in M9 minimal medium[3,4] containing 30 µg/ml kanamycin at 37 °C. At an OD_{600} of ~ 0.7 expression was induced by 1 mM IPTG. Cells were harvested 4 h after induction and lysed as described for *rpoA* using a buffer containing 50 mM Tris/HCl (pH 8.0), 500 mM NaCl, 5 % (v/v) glycerol, 1 mM DTT. After centrifugation (30 min, 4 °C, 12,000 x *g*) the pellet was resolved in 1 mM EDTA (pH 8.0), 1 mg/ml deoxycholic acid sodium salt, 20 mM DTT and lysozyme (0.2 mg/ml) and again centrifuged for 30 min at 12,000 x g and 4 °C. The pellet was then washed three times with the same buffer, subsequently three times with 50 mM Tris/HCl (pH 8.0), 50 mM NaCl, 10 mM EDTA, 5 mM DTT and once with H_2O. Finally, the pellet was resuspended in 50 mM Tris/HCl (pH 7.2), 8 M Urea, 500 mM NaCl and stirred for 1 h at room temperature. Urea was removed by dialysis against 50 mM Tris/HCl (pH 7.2), 5 % (v/v) glycerol, 500 mM NaCl, 0.5 mM EDTA, 1 mM DTT at 4 °C for 3 h followed by overnight dialysis using the same buffer without NaCl. The dialysate was centrifuged (30 min, 4 °C, 12,000 x *g*) and the supernatant was applied to a HiTrap Heparin HP column (GE Healthcare, Munich, Germany). After washing with 50 mM Tris/HCl (pH 7.2), 5 % (v/v) glycerol, 0.5 mM EDTA, 1 mM DTT elution was performed using a constant NaCl gradient up to 1 M in 50 mM Tris/HCl (pH 7.2), 5 % (v/v) glycerol, 0.5 mM EDTA, 1 mM DTT. Fractions containing pure β were combined and dialyzed against the required buffer before the protein solution was concentrated by ultrafiltration (MWCO = 10 kDa) and stored at -80 °C after freezing with liquid nitrogen. The yield was 53 mg protein per l culture.

rpoC was expressed in *E. coli* Rosetta (DE3) pLysS (Novagen, Madison, WI, USA). The recombinant protein harbored a seven amino acid linker followed by a hexahistidine tag (His_6) at the C-terminus. Cells were grown in M9 minimal medium[3,4] containing 30 µg/ml kanamycin and 34 µg/ml chloramphenicol at 37 °C. When an OD_{600} of ~ 0.5 was reached the temperature was lowered to 16 °C and gene expression was induced with 1 mM IPTG at an OD_{600} of 0.6-0.8. Cells were harvested 6 h after induction, resuspended and lysed as described above using buffer A (50 mM Tris/HCl (pH 7.5), 500 mM NaCl, 5 % (v/v) glycerol, 10 mM $MgCl_2$, 10 µM $ZnCl_2$, 10 mM imidazole). After centrifugation (30 min, 12,000 x *g*, 4 °C) the supernatant was applied to a HisTrap HP column (GE Healthcare, Munich, Germany). After washing with buffer A, elution was carried out using a step gradient with increasing imidazole concentrations (10-500 mM in buffer A). Fractions containing β' were combined. Following dialysis against the required buffer the protein solution was concentrated by ultrafiltration (MWCO = 10 kDa) and stored at -80 °C after shock freezing in liquid nitrogen. One liter culture yielded 15 mg protein.

The ω subunit with N-terminal His_6-SUMO tag was produced in *E. coli* Rosetta (DE3) pLysS harboring pET28M-SUMO1/*rpoZ*. Cells were grown in M9 minimal medium[3,4] in the presence of 30 µg/ml kanamycin and 34 µg/ml chloramphenicol at 37 °C until an OD_{600} of 0.4 was reached. The temperature was lowered to 25 °C and at an OD_{600} of 0.6-0.8 expression was induced with 1 mM IPTG. Cells were harvested after 4 h, resuspended and lysed as described above. In this case 25 mM Tris/HCl (pH 7.5), 300 mM NaCl, 10 mM imidazole was used for resuspension. After centrifugation (12,000 x *g*, 30 min, 4 °C), the supernatant was applied to a HisTrap HP column. After washing with 25 mM Tris/HCl (pH 7.5), 300 mM NaCl, 10 mM imidazole, elution was performed using a step gradient with increasing imidazole concentrations (10-500 mM in resuspension buffer). Fractions containing His_6-SUMO-ω were combined and cleaved during dialysis overnight against 25 mM Tris/HCl (pH 7.5), 300 mM NaCl by Senp2, a protease that cleaves directly after SUMO protein. The protein solution was reapplied to the HisTrap HP column.

Pure ω was found in the flow through, dialyzed against the required buffer, concentrated by ultrafiltration (MWCO = 3 kDa) and stored at -80 °C after freezing with liquid nitrogen with a yield of 3 mg protein per liter culture.

Tagless ω was used for *in vitro* assembly of RNAP and produced in *E. coli* Rosetta (DE3) pLysS containing pET32a/*rpoZ*. Cells were grown in M9 minimal medium[3,4] containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol at 37 °C. After induction with 0.1 mM IPTG at an OD_{600} of 0.6-0.8 cells were grown for another 3 h before harvesting (9,000 x *g*, 15 min).

NusA-NTD contained amino acids 1-125 and was produced in *E. coli* BL21 Star (DE3) (Invitrogen, Darmstadt, Germany) harboring pET101_NusA-NTD. Cells were grown at 37 °C in LB medium containing ampicillin (100 µg/ml) until an OD_{600} of 0.6 was reached. Then the temperature was lowered to 20 °C. After 30 min overexpression was induced by 1 mM IPTG. Cells were harvested after overnight growth, resuspended and lysed as described for *rpoA* using a buffer containing 20 mM Tris/HCl (pH 7.5), 150 mM NaCl, 10 % (v/v) glycerol, 20 mM imidazole. After centrifugation at 12,000 x *g* and 4 °C for 30 min, the supernatant was applied to a Ni^{2+}-NTA HiTrap column (GE Healthcare, Munich, Germany). After washing with 20 mM Tris/HCl (pH 7.5), 150 mM NaCl, 10 % (v/v) glycerol, 20 mM imidazole elution was performed *via* a step gradient with increasing imidazole concentrations (20 mM – 1 M imidazole in 20 mM Tris/HCl (pH 7.5), 150 mM NaCl, 10 % (v/v) glycerol, 20 mM imidazole). The fractions containing the NusA-NTD-His_6 fusion protein were combined and the protein was cleaved by thrombin (Novagen, Madison, WI, USA), during dialysis against 20 mM Tris/HCl (pH 7.5) at room temperature overnight. The protein solution was applied to a HiTrap QXL column which was subsequently washed with 20 mM Tris/HCl (pH 7.5) before elution was carried out *via* a step gradient with increasing NaCl concentrations (0 M-1 M NaCl in 20 mM Tris/HCl (pH 7.5)). The fractions containing NusA-NTD were combined and dialyzed against the required buffer. Finally, the protein solution was

concentrated by ultrafiltration (MWCO = 3 kDa), frozen in liquid nitrogen and stored at -80 °C.

The gene of the SUMO protease SENP2 was expressed in *E. coli* Rosetta (DE3) (Novagen, Madison, WI, USA) harboring the plasmid pET28b-senp2. Cells were grown in LB medium containing 30 µg/ml kanamycin and 34 µg/ml chloramphenicol at 37 °C. At OD_{600} ~ 0.7 expression was induced by 1 mM IPTG. Cells were harvested after 4 h (9,000 x *g*, 15 min, 4°C), resuspended in 40 mM Tris/HCl (pH 7.5) containing 500 mM NaCl, 10 mM imidazole and 5 mM DTT and disrupted by a microfluidizer. The supernatant was applied to a HisTrap HP column. Elution was performed using a step gradient with increasing imidazole concentrations (10-500 mM in resuspension buffer). The fractions containing SENP2 were combined, dialyzed against 5 mM Tris/HCl (pH 7.5), 250 mM NaCl, 10 mM DTT, 0.1 mM EDTA and concentrated by ultrafiltration (MWCO = 10 kDa). Finally the glycerol concentration was adjusted to 20 %, aliquots were frozen in liquid nitrogen and stored at -80 °C.

The production and purification of NusB:NusE$^{\Delta}$, NusB, RNAP α-CTD, NusG-NTD, NusG-CTD and NusA-AR2 were carried out as described previously (Refs. [5-7] for NusB:NusE$^{\Delta}$ and NusB, Ref. [8] for αCTD, Ref. [9] for NusG-NTD, Ref. [6] for NusG-CTD, Ref. [8] for NusA-AR2).

Formaldehyde crosslink. The crosslinking of RNAP and NusB:NusE$^{\Delta}$ was based on the SPINE method[10]. 7.7 nmol RNAP were mixed with 15.4 nmol NusB:NusE$^{\Delta}$ in 25 mM HEPES (pH 7.5), 100 mM NaCl and a 4 % (w/v) paraformaldehyde solution in the same buffer was added to a final concentration of 0.6 % (w/v). For the crosslink, the mixture was incubated at 37 °C for 20 min. 0.7 ml of Ni^{2+} chelating sepharose (50 % (w/v), GE Healthcare, Munich, Germany), equilibrated with 25 mM HEPES (pH 7.5), 100 mM NaCl, were added and incubated for 20 min at room temperature. Afterwards the mixture was transferred to a 2.5 ml gravity flow column and the flow trough was collected. The column was washed ten times with 1 ml of 25 mM HEPES (pH 7.5),

100 mM NaCl and seven times with 1 ml of the same buffer containing 5 mM imidazole. Bound protein was eluted with 25 mM HEPES (pH 7.5), 100 mM NaCl, 500 mM imidazole. The protein contained in 200 µl in the flow through, the last washing step and the eluate was precipitated with 50 µl 50 % (v/v) trichloroacetic acid (TCA) by incubation for 20 min on ice and subsequent centrifugation for 10 min at 15,000 x *g*. The pellet was dissolved in 50 µl 2x Roti (Roth, Karlsruhe, Germany). The crosslink was broken by boiling the solution for 20 min and the samples were analyzed by SDS-PAGE. The isolated RNAP, the isolated NusB:NusE$^{\wedge}$ complex as well as NusG-CTD in the absence and presence of RNAP as negative control were treated accordingly.

Programs. All structures were visualized using PyMOL[11].

Supplementary References

1. Schägger, H. & von Jagow, G. Tricine Sodium Dodecyl-Sulfate Polyacrylamide-Gel Electrophoresis for the Separation of Proteins in the Range from 1-Kda to 100-Kda. *Anal. Biochem.* **166,** 368-379 (1987).
2. 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).
3. Meyer, O. & Schlegel, H. G. Biology of aerobic carbon monoxide-oxidizing bacteria. *Annu. Rev. Microbiol.* **37,** 277-310 (1983).
4. Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning - A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1994).
5. Burmann, B. M., Luo, X., Wahl, M. C., Rösch, P. & 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).
6. Burmann, B. M. *et al*. A NusE:NusG complex links transcription and translation. *Science* **328,** 501-504 (2010).
7. Luo, X. *et al*. Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. *Mol. Cell* **32,** 791-802 (2008).
8. Schweimer, K. *et al*. NusA interaction with the α-subunit of *E. coli* RNA polymerase is via the UP-element site and releases autoinhibition. *Structure* **19,** 945-954 (2011).
9. Burmann, B. M., Schweimer, K., Scheckenhofer, U. & Rösch, P. Domain interactions of the transcription:translation coupling factor *E.coli* NusG are intermolecular and transient. *Biochem. J.* **435,** 783-789 (2011).
10. Herzberg, C. *et al*. SPINE: a method for the rapid detection and analysis of protein-protein interactions *in vivo*. *Proteomics* **7,** 4032-4035 (2007).

11. Schrödinger L. The PyMOL molecular graphics system, version 1.3. *Schrödinger, LLC, Mannheim, Germany* (2010).

9 Einzelarbeit D

Johanna Drögemüller*, Martin Strauß*, Kristian Schweimer, Marcel Jurk, Paul Rösch und Stefan H. Knauer (2015): Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Scientific Reports* **5**: 16428

* beide Autoren haben im gleichen Maße zur Arbeit beigetragen

SCIENTIFIC REPORTS

OPEN

Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy

Received: 15 July 2015
Accepted: 13 October 2015
Published: 12 November 2015

Johanna Drögemüller*, Martin Strauß*,‡, Kristian Schweimer, Marcel Jurk†, Paul Rösch & Stefan H. Knauer

In bacteria, RNA polymerase (RNAP), the central enzyme of transcription, is regulated by N-utilization substance (Nus) transcription factors. Several of these factors interact directly, and only transiently, with RNAP to modulate its function. As details of these interactions are largely unknown, we probed the RNAP binding surfaces of *Escherichia coli* (*E. coli*) Nus factors by nuclear magnetic resonance (NMR) spectroscopy. Perdeuterated factors with [^{1}H,^{13}C]-labeled methyl groups of Val, Leu, and Ile residues were titrated with protonated RNAP. After verification of this approach with the N-terminal domain (NTD) of NusG and RNAP we determined the RNAP binding site of NusE. It overlaps with the NusE interaction surface for the NusG C-terminal domain, indicating that RNAP and NusG compete for NusE and suggesting possible roles for the NusE:RNAP interaction, e.g. in antitermination and direct transcription:translation coupling. We solved the solution structure of NusA-NTD by NMR spectroscopy, identified its RNAP binding site with the same approach we used for NusG-NTD, and here present a detailed model of the NusA-NTD:RNAP:RNA complex.

Transcription of genomic information from DNA to RNA is the initial step in gene expression, with RNA polymerase (RNAP) being the key enzyme of this process in all domains of life[1]. Bacterial core RNAP consists of five subunits, 2x α, β, β′, and ω. While the α subunits promote the assembly of the enzyme and are target of many regulatory proteins[2–4], the β and β′ subunits form the active site and catalyze RNA synthesis[5,6]. The ω subunit is supposed to play a structural rather than a functional role. It binds to the N- and C-termini of the β′ subunit to prevent β′ aggregation until the ωβ′ complex is integrated into the RNAP[7]. During initiation of transcription the σ factor binds to core RNAP to form the holo enzyme, and σ is also essential for the recognition and melting of promoter regions (reviewed in[8]). The transcription cycle consists of three major phases: initiation, elongation, and termination. It is highly regulated by a multitude of transcription factors that bind to RNAP modifying its action. Prominent examples are the N utilization substance (Nus) factors that influence especially elongation and termination. Among all transcription factors NusG (Spt5 in archaea and eukaryotes) is unique as it is the only one that is universally conserved[9]. *Escherichia coli* (*E. coli*) NusG is a two-domain protein, with an N-terminal domain (NTD) and a C-terminal domain (CTD) connected *via* a flexible linker[10]. During elongation NusG-NTD binds to RNAP, enhancing the elongation rate and suppressing pauses[10,11]. To fulfill this function NusG-NTD contacts the β′ clamp helices (β′CH) and the β gate loop (βGL), closing the active site cleft so that the nucleic acids are locked and the transcription elongation complex (TEC) is stabilized (Fig. 1)[12,13]. Although NusG/Spt5-NTDs highly likely have the same function in all domains of life, NusG/Spt5-CTDs are targets of various interaction partners and thus serve as recruitment platform for further accessory factors. In *E. coli*, NusG-CTD binds to the termination factor Rho,

Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany. †Present address: Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany. ‡Present address: Columbia University Medical Center, New York, NY 10032, USA. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.H.K. (email: stefan.knauer@uni-bayreuth.de)

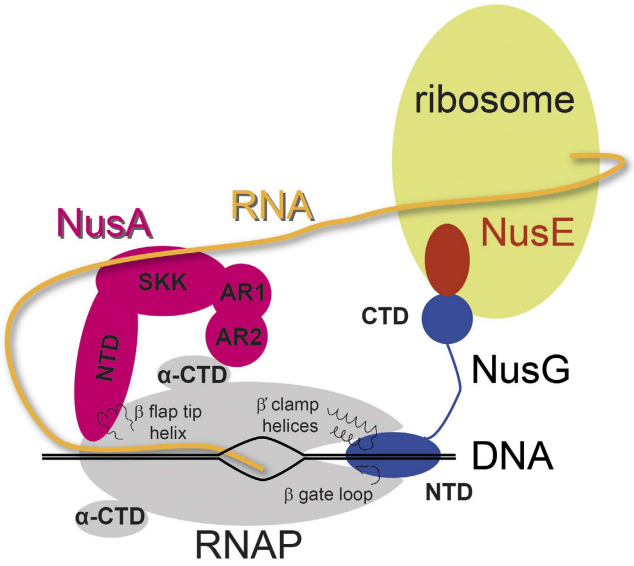


Figure 1. Schematic representation of transcription:translation coupling. NusA, pink, NusE, red; NusG, blue; RNAP, grey; ribosome, light green; DNA, black; RNA, yellow. In RNAP selected structural elements involved in Nus factor binding are indicated.

promoting Rho-dependent termination[14,15]. Additionally, *E. coli* NusG-CTD interacts with ribosomal protein S10 to couple transcription and translation (Fig. 1)[14]. S10 is identical to transcription factor NusE that forms a complex with NusB and as such is involved in antitermination[16]. In the multiprotein antitermination complex RNAP is modified to be able to read through termination signals, a process that is essential for efficient transcription of ribosomal RNA operons[17] or the DNA of lambdoid phages[18]. The NusE:NusB complex formed during antitermination binds to the single stranded, highly conserved *BoxA* RNA sequence[19] and is anchored to RNAP *via* NusE:NusG-CTD interaction[14]. However, NusE is also able to bind directly to RNAP where it remains during elongation[16,20]. This interaction may be involved in antitermination, and the binding site on RNAP is suggested to be located in the β subunit[20].

NusA is a multidomain protein consisting of an NTD, an S1, and two K-homology RNA binding domains, KH1 and KH2, the latter three forming the SKK domain. In *E. coli* and several other proteobacteria the NusA C-terminus comprises two acidic repeat domains, AR1 and AR2[21,22]. With its multitude of interaction partners, NusA is able to accomplish various functions. It modulates Rho-dependent and intrinsic termination, it either prolongs pauses or introduces new ones, and it is part of the antitermination complex (reviewed in[23,24]). NusA interacts directly with RNAP *via* NusA-NTD and NusA-AR2 (Fig. 1)[25]. While a high resolution solution nuclear magnetic resonance (NMR) structure is available for the complex of NusA-AR2 and the CTD of the RNAP α subunit (α-CTD)[4], the RNAP interaction surface of NusA-NTD is not experimentally defined in atomic detail. A low resolution electron microscopy structure of the *Bacillus subtilis* (*B. subtilis*) NusA-NTD:RNAP complex as well as initial binding models are available and all studies suggest that NusA-NTD binds to the flap region of the β subunit at the RNA exit channel[26–28]. However, the exact RNAP binding surface on NusA-NTD remains to be determined.

Knowledge of the RNAP interaction surfaces of transcription factors is crucial for the complete understanding of RNAP regulation. Owed to the molecular mass of RNAP (*E. coli* RNAP ~390 kDa), the main techniques to study RNAP:transcription factor complexes structurally in atomic detail are X-ray crystallography and electron microscopy. However, RNAP regulation heavily depends on transient interactions and dynamics, i.e. information not easily accessible by these techniques. Thus, we chose to study *E. coli* RNAP:Nus factor interaction by NMR spectroscopy to identify the RNAP binding surface of these transcription factors. Our approach is based on observations that even in systems >100 kDa methyl groups are excellent NMR probes as they are still mobile enough to produce highly resolved spectra with good signal intensities owed to their fast motions around the methyl axis[29].

Results and Discussion

RNAP interface of NusG-NTD. To identify the RNAP binding surface of transcription factors the methyl groups of Ile (δ1), Leu (δ1 or δ2), and Val (γ1 or γ2) residues of the respective, deuterated factor were labeled with [^{1}H,^{13}C] ([I,L,V]-labeled transcription factor; for clarity, all protein names without prefix refer to *E. coli* proteins). The titration of this [I,L,V]-labeled regulator with protonated RNAP was observed by two-dimensional (2D) [^{1}H,^{13}C]-methyl transverse relaxation optimized spectroscopy (TROSY). As a test case for the applicability of this method, we asked whether we were able to confirm

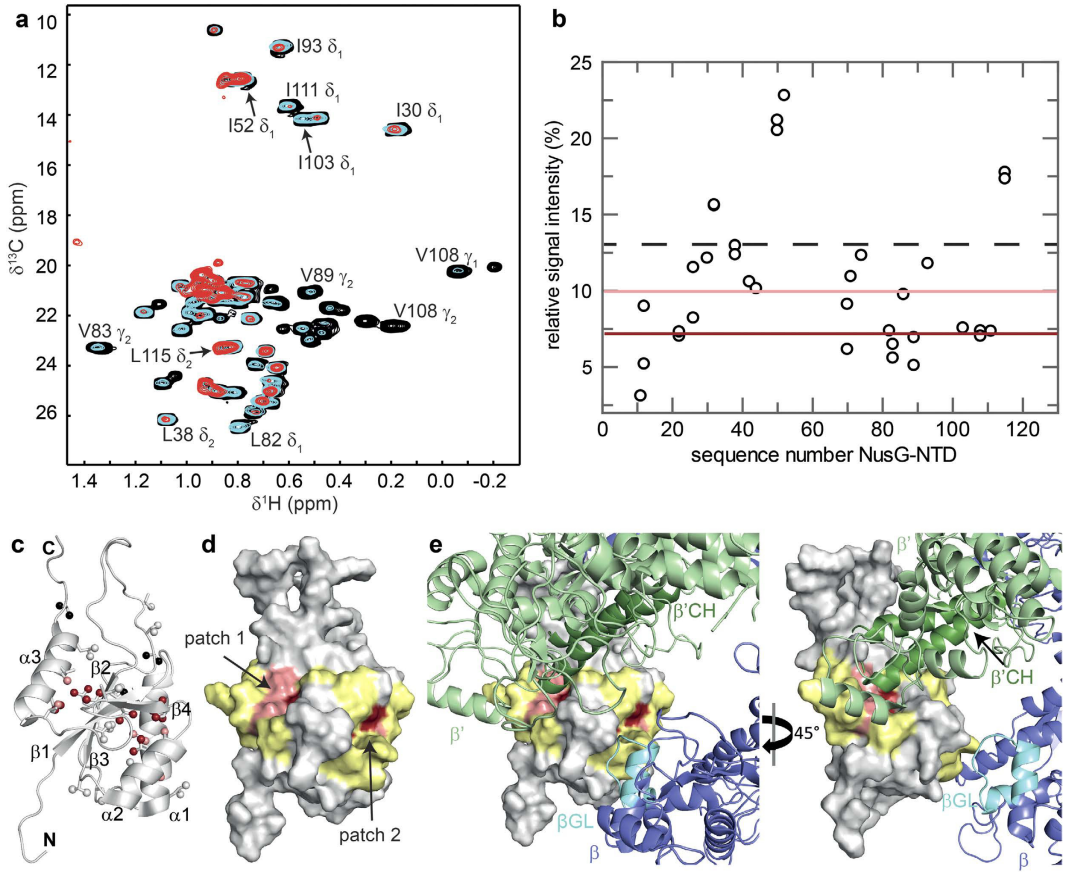


Figure 2. RNAP binding site of NusG-NTD. (**a**) Titration of [I,L,V]-NusG-NTD with protonated RNAP. Methyl-TROSY spectra of [I,L,V]-NusG-NTD in the absence, black, and in the presence of RNAP (1:1 molar ratio, cyan; 1:2 molar ratio, red). Selected signals are labeled. (**b**) Relative signal intensity of [I,L,V]-NusG-NTD after addition of RNAP in equimolar concentration *vs.* residue number of NusG-NTD. The dashed black line indicates the average relative signal intensity. Dark red and light red lines indicate the thresholds for strongly affected (55% of the average relative intensity) and slightly affected (75% of the average relative intensity) methyl groups, respectively. (**c**) Mapping of affected methyl groups onto the NusG-NTD structure (Protein Data Bank (PDB) ID: 2K06, cartoon representation, grey). Ile, Leu, and Val residues are in stick representation with the carbon atoms of their methyl groups as spheres. Strongly affected methyl groups, dark red; slightly affected methyl groups, light red; unaffected methyl groups, grey; unassigned methyl groups, black. Secondary structure elements and termini are labeled. (**d**) Mapping of affected residues onto the NusG-NTD structure (surface representation). For graphical illustration of the interaction site the complete amino acid was colored as affected in lieu of the methyl group. Colors are as in (**c**). Two amino acids on either side of affected Ile/Leu/Val residues are highlighted in yellow unless they were unaffected Ile/Leu/Val residues. (**e**) Model of NusG-NTD as in (**d**) bound to *E. coli* RNAP (PDB ID: 4KMU). The model is based on the structure of the *Pyrococcus furiosus* (*P. furiosus*) Spt4/5 complex bound to the RNAP clamp domain (PDB ID: 3QQC). NusG-NTD was superposed on Spt5 and RNAP β′ subunit on the clamp domain. As NusG-NTD and RNAP were treated as rigid bodies and no further optimization was carried out some minor clashes occur. β subunit, light blue; β′ subunit, light green; β′CH, dark green; βGL, cyan.

the RNAP binding surface of NusG-NTD. This surface is known from a crystallographic study of the archaeal Spt4/5 complex with the β′ clamp domain of RNAP and biochemical experiments on NusG and RfaH, the latter being a paralog of NusG[12,13].

Upon addition of RNAP, the methyl group signals of [I,L,V]-NusG-NTD decreased in intensity, but not uniformly over all signals (Fig. 2a), likely caused by a combination of several effects. First, a general loss of signal intensity is owed to [I,L,V]-NusG-NTD:RNAP complex formation as the molecular mass (MM) of the complex is roughly 30-fold that of [I,L,V]-NusG-NTD ($MM_{NusG\text{-}NTD}$ = 14 kDa, MM_{RNAP} = 389 kDa), resulting in severe line broadening. Second, by binding of [I,L,V]-NusG-NTD to RNAP, the specifically labeled methyl groups of [I,L,V]-NusG-NTD located in the binding interface get into close proximity of the RNAP protons, and the resulting intermolecular dipole-dipole interactions cause an additional contribution to relaxation, so that the signal intensity of methyl groups in the binding surface decreases more strongly than that of methyl groups located elsewhere in [I,L,V]-NusG-NTD.

Finally, signal intensities can be influenced by chemical exchange processes in the intermediate range of the NMR timescale. Quantitative analysis of signal intensities for the 1:1 complex revealed two patches in the protein structure where signal intensities changed noticeably (Fig. 2b,c). Patch 1 comprises residues in helix α3 and strands β1 and β3, while patch 2 is formed by residues located in helices α1 and α2, and these two patches are located at nearly opposite sides of NusG-NTD. No assigned, but unaffected methyl groups were found in either of these patches. This approach provides only information about Ile, Leu, and Val residues, but most likely additional amino acids, especially in the direct vicinity of the affected residues, are involved in the interaction. Thus we graphically extended the representation of patches 1 and 2 by including the two residues preceding and following each affected Ile, Leu, or Val residue, unless they were unaffected Ile, Leu, or Val residues, resulting in two continuous regions (Fig. 2d). In a model of NusG-NTD bound to RNAP based on the crystal structure of the archaeal Spt4/5: β′ clamp domain complex[12], residues of patch 1 are in direct proximity of the β′CH, indicating that we identified correctly the β′CH binding site (Fig. 2e). The NTD of RfaH, an *E. coli* paralog of NusG, not only interacts with the β′CH, but also binds to the βGL *via* His65, Thr66, and Thr67 which form an HTT motif located at the N-terminus of helix α2 (Supplementary Fig. 1)[13]. Although this interaction does not contribute significantly to the overall affinity of RfaH-NTD for RNAP it is essential for the antipausing activity of RfaH[13]. Similarly, structurally homologous residues in NusG-NTD (Ser79-His81) have been proposed to be involved in βGL binding, suggesting that this interaction is a general feature of NusG-like proteins[13]. NusG-NTD patch 2 corresponds to the RfaH region that is in immediate neighborhood of the βGL binding motif suggested for RfaH-NTD (Supplementary Fig. 1)[13]. Due to the absence of Ile, Leu, and Val residues in the NusG-NTD region that is structurally homologous to the HTT motif in RfaH, no direct information about this region is available in our experiments (Supplementary Fig. 1). Thus, we conclude that either the βGL binding surface in NusG-NTD differs slightly from the one in RfaH-NTD or that patch 2 constitutes only part of the βGL interaction surface or that residues of patch 2 are indirectly affected as they are located next to the actual binding site.

The clamp domain undergoes structural rearrangements during the transcription cycle, having closed and open conformations, and NusG-NTD/RfaH-NTD is proposed to lock the clamp in a closed state during elongation by making bridging contacts between the β′CH and the βGL so that the downstream DNA is completely encircled[13,30–33]. Hence, the elongation complex is stabilized and structural rearrangements that occur during pausing are prevented, which, in turn, leads to increased processivity. As we used core RNAP in our experiments the clamp is probably in an open state. Thus our findings indicate that in the absence of nucleic acids NusG-NTD contacts the β′CH and βGL either separately or simultaneously, suggesting that the RNAP claw is in a conformation that allows these contacts or that NusG-NTD induces a closed state.

Overall, the binding surfaces identified here are consistent with the previously published interaction sites of NusG-NTD, demonstrating that the present approach may be used to determine the RNAP binding surfaces of transcription factors in solution in a single experiment using intact RNAP and avoiding molecular alteration of the constituents. However, the limited number of NMR probes and their distribution over the structure restricts the structural resolution of the resulting binding site. Although we are not able to distinguish between methyl groups that are directly involved in the molecular interaction from those that are only indirectly affected, the careful interpretation of the surface representation allows us to identify the interaction surface.

RNAP interface of NusE. Transcription factor NusE/S10 not only interacts with RNAP *via* NusG, but it is also able to bind directly and specifically to the RNAP β subunit during transcription[14,16,20]. The function of this interaction is still unknown. In order to study the molecular details of this interaction we determined the RNAP binding surface of NusE with the same approach as for NusG-NTD. As NusE alone is very unstable and tends to aggregate we used a NusE variant that lacks the ribosome binding loop ($NusE^{\Delta}$) in complex with NusB for our experiments[34]. The presence of NusB does not influence the $NusE^{\Delta}$:RNAP interaction[20]. For the NMR titration, we labeled the methyl groups of Ile, Leu, and Val residues of $NusE^{\Delta}$ in the deuterated NusB:$NusE^{\Delta}$ complex with [^{1}H,^{13}C] ([I,L,V]-$NusE^{\Delta}$).

Upon addition of protonated RNAP, [I,L,V]-$NusE^{\Delta}$ methyl group signals decreased in varying proportion (Fig. 3a,b). All highly and slightly affected methyl groups are located in helices α1 and α2 as well as strands β1 and β4 (Fig. 3c). Inspection of the surface representation and the graphical extension as carried out for NusG-NTD result in a continuous patch (Fig. 3d). As the 7 Ile, 10 Leu, and 7 Val residues of $NusE^{\Delta}$ (86 residues overall) are distributed evenly over the sequence and the structure, our definition of the interaction surface is highly reliable. The RNAP binding site is opposite of the NusB:$NusE^{\Delta}$ interface and the ribosome integration site, i.e. the $NusE^{\Delta}$:RNAP interaction is not only possible within the context of the NusB:$NusE^{\Delta}$ complex, but also when NusE is integrated into the ribosome[35]. NusE could thus simultaneously accommodate the ribosome and the RNAP.

Interestingly, $NusE^{\Delta}$'s binding surface for RNAP strongly overlaps with that for NusG-CTD so that binding of $NusE^{\Delta}$ to RNAP and NusG-CTD should be mutually exclusive (Fig. 3e)[14]. Thus we asked whether NusG-CTD and RNAP compete for binding to NusE. We performed a [^{1}H,^{15}N]-heteronuclear single quantum coherence (HSQC) displacement experiment in which the complex NusB:[^{15}N]-$NusE^{\Delta}$:RNAP was titrated with NusG-CTD (Fig. 4a). In the one-dimensional (1D) [^{1}H,^{15}N]-HSQC spectra signals of [^{15}N]-$NusE^{\Delta}$ strongly decreased upon NusB:[^{15}N]-$NusE^{\Delta}$:RNAP complex formation as the increase

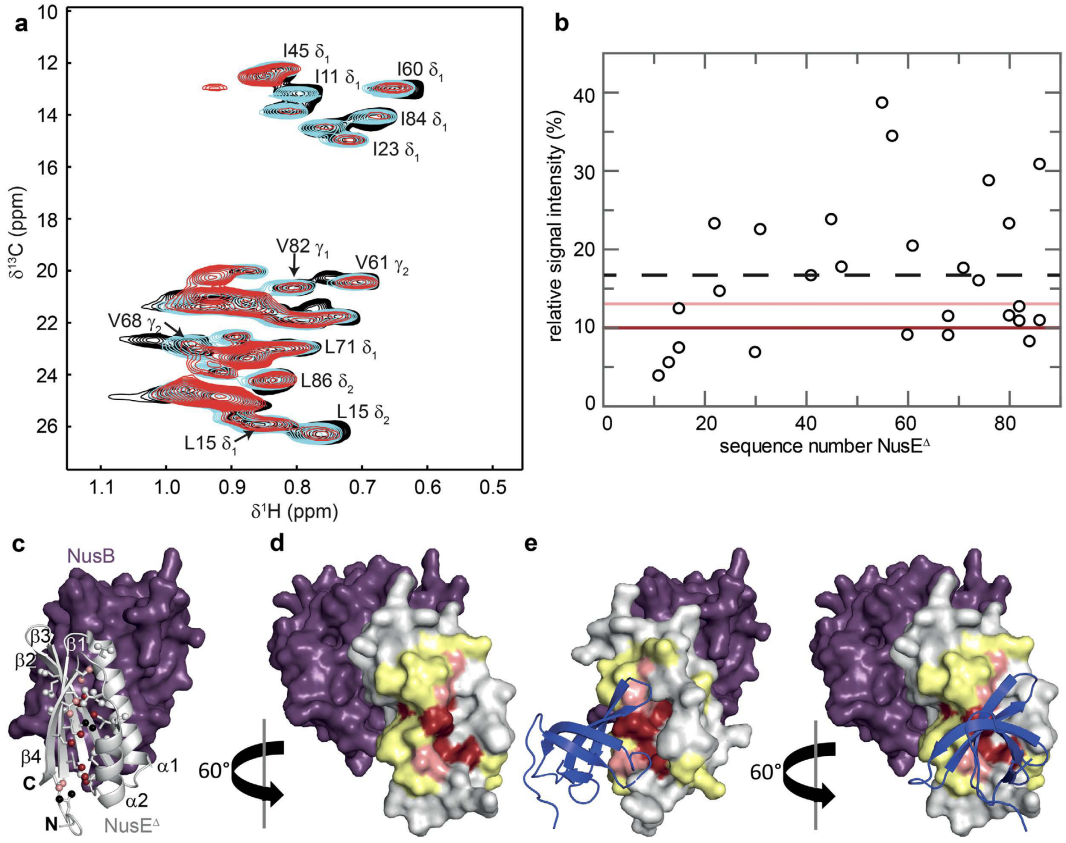


Figure 3. RNAP binding site of NusE$^{\Delta}$. (**a**) Titration of [I,L,V]-NusE$^{\Delta}$ with protonated RNAP (NusE$^{\Delta}$ being in complex with deuterated NusB). Methyl-TROSY spectra in the absence, black, and in the presence of RNAP (1:1 molar ratio, cyan; 1:2 molar ratio, red), with representative signal assignments. (**b**) Relative [I,L,V]-NusE$^{\Delta}$ signal intensity after addition of RNAP in a 1:2 molar ratio *vs.* amino acid sequence positions of NusE$^{\Delta}$. Dashed black line, average relative signal intensity; dark red and light red lines, thresholds for strongly affected (60% of the average relative intensity) and slightly affected (80% of the average relative intensity) methyl groups, respectively. (**c**) Mapping of affected methyl groups onto the NusB:NusE$^{\Delta}$ complex structure (PDB ID: 3D3B; NusB, purple; NusE$^{\Delta}$, light grey). NusB in surface, NusE$^{\Delta}$ in cartoon representation. Ile, Leu, and Val residues in NusE$^{\Delta}$ are represented as sticks with the carbon atoms of their methyl groups as spheres. Strongly affected methyl groups, dark red; slightly affected methyl groups, light red; unaffected methyl groups, grey; unassigned methyl groups, black. Secondary structure elements and termini are labeled. (**d**) Mapping of affected residues onto the NusB:NusE$^{\Delta}$ complex structure (surface representation). Colors are as in (**c**). For graphical illustration of the interaction site the complete amino acid was colored as affected in lieu of the methyl group. Two amino acids on either side of an affected Ile/Leu/Val residue are highlighted in yellow unless they were unaffected Ile/Leu/Val residues. (**e**) Structure of the NusB:NusE$^{\Delta}$:NusG-CTD complex. The NusE$^{\Delta}$:NusG-CTD complex (PDB ID: 2KVQ, NusG-CTD in blue cartoon representation) was superposed on the NusB:NusE$^{\Delta}$ complex from (**d**).

of the molecular mass leads to significant line broadening. Titration with NusG-CTD reversed this effect, demonstrating the displacement of RNAP from NusB:[^{15}N]-NusE$^{\Delta}$. The corresponding 2D [^{1}H,^{15}N]-HSQC spectra show that released NusB:[^{15}N]-NusE$^{\Delta}$ binds to NusG-CTD (Supplementary Fig. 2). Thus, NusG-CTD can abstract NusE$^{\Delta}$ from RNAP. Next, we asked whether in reverse RNAP can displace NusG-CTD from the NusB:NusE$^{\Delta}$:NusG-CTD complex. We titrated NusB:NusE$^{\Delta}$:[^{15}N]-NusG-CTD with RNAP and followed the titration by recording 2D [^{1}H,^{15}N]-HSQC spectra (Fig. 4b,c). Addition of NusB:NusE$^{\Delta}$ to [^{15}N]-NusG-CTD led to changes in the chemical shifts of [^{15}N]-NusG-CTD signals typical for NusB:NusE$^{\Delta}$:[^{15}N]-NusG-CTD complex formation. Those changes were reversed by about 50% when RNAP was added in 3-fold molar excess, as expected on disruption of the NusB:NusE$^{\Delta}$:NusG-CTD complex by NusE:RNAP interaction. Thus, RNAP and NusG-CTD compete for NusE$^{\Delta}$ with similar low micromolar K_D values (NusB:NusE$^{\Delta}$:NusG-CTD: 50 μM)[14].

These competition experiments support the notion of overlapping binding sites of NusE for NusG-CTD and RNAP, and they show that NusG-CTD can interact with NusE in the presence of RNAP. The complexes NusE:RNAP and NusE:NusG:RNAP *via* NusG are thus in a delicate equilibrium that can easily be influenced by other regulators such as transcription factors or certain RNA sequences. Overall, formation

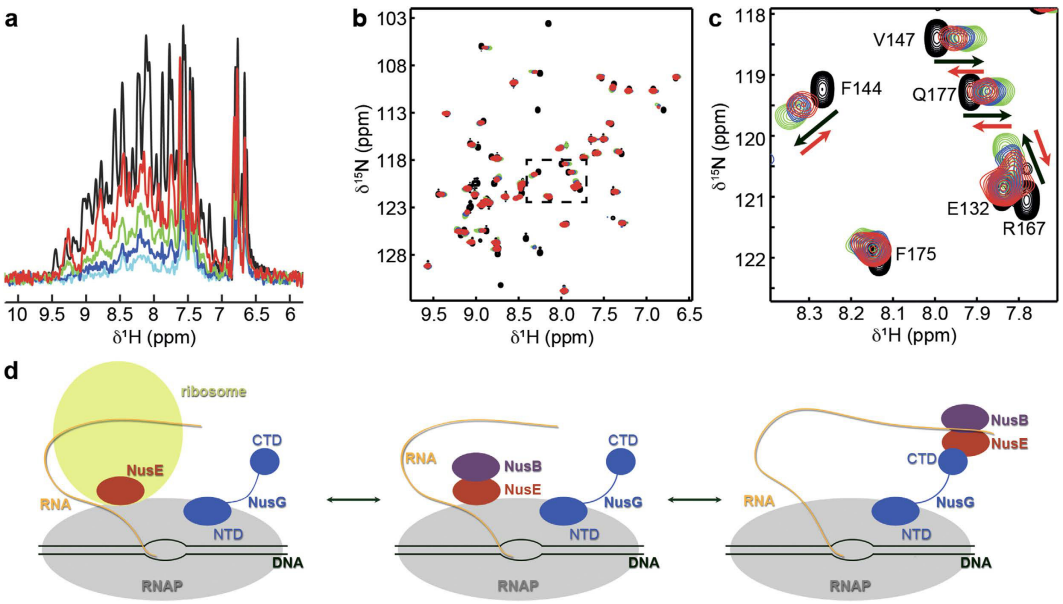


Figure 4. Competition of RNAP and NusG-CTD for NusE binding. (**a**) Displacement of RNAP from NusB:NusE$^{\Delta}$ by NusG-CTD. 1D [^{1}H,^{15}N]-HSQC spectra of free NusB:[^{15}N]-NusE$^{\Delta}$, black, NusB:[^{15}N]-NusE$^{\Delta}$ in the presence of RNAP in equimolar concentration, light blue, and NusB:[^{15}N]-NusE$^{\Delta}$ in the presence of RNAP and NusG-CTD (molar ratio 1:1:1, dark blue; 1:1:3, green; 1:1:10, red). (**b**) Displacement of NusB:NusE$^{\Delta}$ from NusG-CTD by RNAP. 2D [^{1}H,^{15}N]-HSQC spectra of [^{15}N]-NusG-CTD, black, [^{15}N]-NusG-CTD in the presence of NusB:NusE$^{\Delta}$ in equimolar concentration, green, and [^{15}N]-NusG-CTD in the presence of NusB:NusE$^{\Delta}$ and RNAP (molar ratio 1:1:1, blue; 1:1:3, red). (**c**) Detail of the rectangular region in (**b**). Black arrows indicate the chemical shift changes that occur upon addition of NusB:NusE$^{\Delta}$ to [^{15}N]-NusG-CTD, red arrows show the changes upon subsequent addition of RNAP. (**d**) Schematic representation of the potential functions of a direct NusE:RNAP interaction. Color code as in Fig. 1.

of the NusE:RNAP complex might play various roles during transcription (Fig. 4d). It might be involved either in transcription:translation coupling as the ribosome could directly contact RNAP *via* S10, e.g. when the RNA tether is relatively short, or in transcription antitermination where NusB:NusE is part of the antitermination complex[14,16,19]. The amount of free NusE that is not bound to the ribosome is estimated to be very low, but it is essential for transcription antitermination[36]. Thus tethering of NusE or the NusB:NusE complex to RNAP might be an early event in transcription antitermination to increase the local NusE concentration. NusE would remain bound to the TEC until transferred to NusG-CTD during assembly of the antitermination complex. As ribosomal operons comprise a very high density of transcribing RNAPs with high elongation rates[37], tethering NusE directly to RNAP would ensure fast and efficient transcription antitermination in these operons.

Solution structure of NusA-NTD from *E. coli*. The six domains comprising transcription factor NusA associates with RNAP *via* NusA-NTD, which is necessary and sufficient for the enhancement of pausing during transcription[27]. To determine the solution structure of NusA-NTD by NMR spectroscopy we initially tried a construct containing amino acids Met1-Ile137 carrying an N-terminal His_9-tag, NusA(1–137). The high degree of heterogeneity in the peak intensities as well as the spectral overlap in the [^{1}H,^{15}N]-HSQC spectrum of the [^{15}N]-labeled protein, however, prevented further analysis (Supplementary Fig. 3). A shorter construct, NusA-NTD$^{\Delta}$, consisting of amino acids Met1-Met125 and a cleavable C-terminal His_6-tag, led to homogeneous signal intensities with non-overlapping signals in the [^{1}H,^{15}N]-HSQC spectra (Supplementary Fig. 3) and allowed nearly complete backbone and side chain resonance assignment. No resonances were found for residues Asp103, Arg104, Thr106, Thr107, and Gln108. These are located in a flexible loop so that severe line broadening may occur caused by either fast solvent exchange or conformational exchange on the intermediate chemical shift time scale. Structure determination was performed on the basis of 1565 distance and 193 dihedral restraints derived from multiple NMR experiments (Table 1).

NusA-NTD$^{\Delta}$ comprises four α-helices (α1: Asn2–Ala17, α2: Pro19–Glu40, α3: Leu77–Glu85, α4: Thr106–Ala124) and four β-strands (β1: Val45–Asp50, β2: Asp55–Val65, β3: Glu74–Thr76, β4: Gly90–Gln96) and its structure resembles that of NusA-NTDs from other bacteria[22,28,38,39]. It is L-shaped, with a globular head and a mainly α-helical body (Fig. 5a and b). In the latter α1, α2, α4, β1, and β2 surround an elongated hydrophobic core, and the long β2 strand protrudes into the globular head. The C-terminal

Distance restraints	total	1507
	intraresidual	329
	sequential	386
	medium range	321
	long range	471
Hydrogen bond restraints		58
Dihedral restraints		193
Restraint violations	rms distance violation (Å)	0.006 (±0.0011)
	max. distance violation (Å)	0.11
	rms dihedral violation (°)	0.05 (±0.02)
	max. dihedral violation (°)	0.8
	rmsd bond length (Å)	0.00070 (±0.00009)
	rmsd bond angle (°)	0.13 (±0.012)
Atomic coordinate precision	backbone atoms (Å)	0.80[a]
	all heavy atoms (Å)	1.13[a]
Ramachandran plot statistics[b]	most favored regions (%)	90.5
	additional allowed regions (%)	8.8
	generously allowed regions (%)	0.2
	disallowed regions (%)	0.5

Table 1. Experimental constraints for structure calculation of NusA-NTD$^{\Delta}$. [a]residues Met1-Arg123. [b]determined by PROCHECK-NMR.

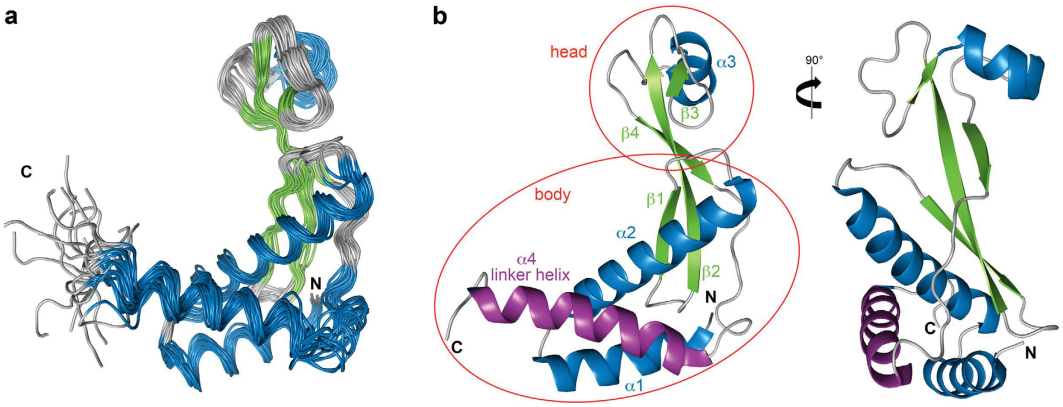


Figure 5. Solution structure of NusA-NTD$^{\Delta}$. (**a**) Structural ensemble of the 20 accepted lowest energy structures in ribbon representation colored according to secondary structure (α-helices, blue; β-strands, green; loops, grey). (**b**) Cartoon representation of the calculated structure with the lowest energy. Secondary structure elements are colored as in (**a**) and labeled. Helix α4 is highlighted in purple, the head and body parts are indicated.

helix α4 connects NusA-NTD and the NusA-SKK domain (linker helix). The globular head comprises α3, β3, β4, and the N-terminal part of β2. While the head is mainly acidic, the body exhibits large basic patches (Supplementary Fig. 4).

To date structures of NusA proteins from different bacteria are available, and although all NusA-NTDs are similar in their overall architecture, they differ in the position of the linker helix (Supplementary Fig. 5a–f). For NusA-NTD from *B. subtilis* (*Bs*NusA-NTD), NMR data suggest that this helix occurs in two alternative conformations in solution[28]. However, we have no indication for the presence of multiple conformations of helix α4 in NusA-NTD$^{\Delta}$. Moreover, unambiguous [^{15}N]-nuclear Overhauser enhancement spectroscopy (NOESY) cross peaks between hydrophobic amino acids could be observed in NMR experiments, demonstrating a direct interaction between helix α4 and helices α1 and α2 in NusA-NTD$^{\Delta}$ (Supplementary Fig. 5g). As crystal structures of full length NusA from *Thermotoga maritima* (*Tm*NusA, protein data bank (PDB) IDs: 1HH2, 2L2F), *Mycobacterium tuberculosis* (*Mt*NusA, PDB ID: 1K0R) and *Planctomyces limnophilus* (*Pl*NusA, PDB ID: 4MTN) show that the NusA-SKK domain is connected

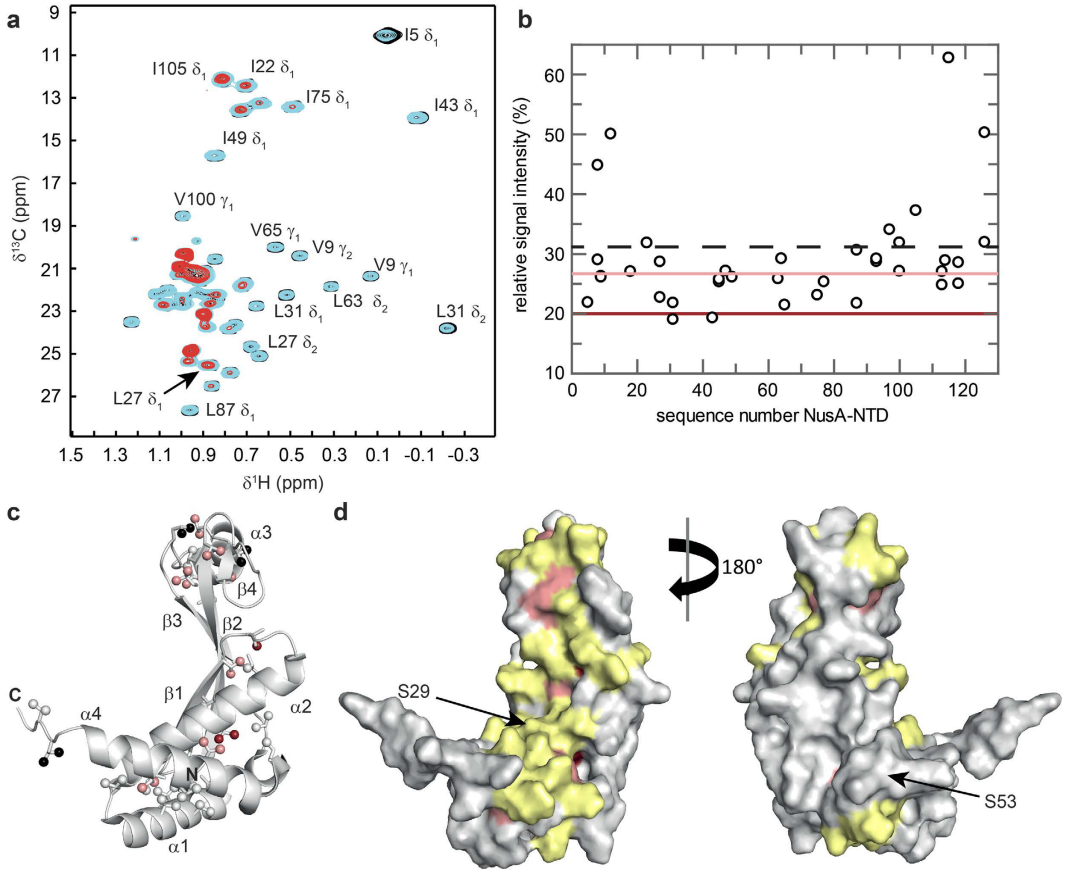


Figure 6. RNAP binding site of NusA-NTD$^{\Delta}$. (**a**) Titration of [I,L,V]-NusA-NTD$^{\Delta}$ with RNAP. Methyl-TROSY spectra of [I,L,V]-NusA-NTD$^{\Delta}$ in the absence, black, and in the presence of RNAP (1:1 molar ratio, cyan; 1:2 molar ratio, red), with assignment of representative signals. (**b**) Relative [I,L,V]-NusA-NTD$^{\Delta}$ signal intensity after addition of RNAP in equimolar concentration *vs*. amino acid sequence positions of NusA-NTD$^{\Delta}$. Dashed black line, average relative signal intensity; dark red and light red lines, thresholds for strongly affected (65% of the average relative intensity) and slightly affected (85% of the average relative intensity) residues, respectively. (**c**) Mapping of affected methyl groups onto the NusA-NTD$^{\Delta}$ structure. NusA-NTD$^{\Delta}$ (grey) in cartoon representation. Ile, Leu, and Val residues are in stick representation with the carbon atoms of their methyl groups as spheres. Strongly affected methyl groups, dark red; slightly affected methyl groups, light red; unaffected methyl groups, grey; unassigned methyl groups, black. (**d**) Mapping of affected residues onto the NusA-NTD$^{\Delta}$ structure (surface representation). For graphical illustration of the interaction site the complete amino acid was colored as affected in lieu of the methyl group. Colors are as in (**c**). Two amino acids on either side of an affected Ile/Leu/Val residue are highlighted in yellow unless they were unaffected Ile/Leu/Val residues. The positions of Ser29 and Ser53 are marked by black arrows.

to the linker helix by only a short loop, this helix might be responsible for the correct positioning of NusA-SKK for RNA binding.

Comparing NusA-NTD structures it is striking that *Mt*NusA-NTD and *Pl*NusA-NTD lack the globular head (Supplementary Fig. 5a–e), which is proposed to interact with the β′ subunit of RNAP[40]. This might indicate a different mode of action/binding of *Mt*NusA and *Pl*NusA compared to other NusAs.

RNAP interface of NusA-NTD. NusA-NTD is supposed to bind to RNAP by interacting with the β flap tip helix of the β flap region, which forms the outer wall of the RNA exit channel. To date, available complex models are based on a low-resolution electron microscopy structure, cleavage experiments, targeted amino acid exchanges and NMR experiments using a short β flap construct[26–28]. Here we used complete RNAP to determine the RNAP binding site of NusA-NTD$^{\Delta}$ by applying the same approach as for NusG-NTD and NusE$^{\Delta}$. Methyl group labeled NusA-NTD$^{\Delta}$ ([I,L,V]-NusA-NTD$^{\Delta}$) was titrated with protonated RNAP leading to a non-uniform decrease of [I,L,V]-NusA-NTD$^{\Delta}$ methyl group signals (Fig. 6a). Again, the normalized signal intensity decrease in the 1:1 complex was analyzed to identify highly and slightly affected methyl groups (Fig. 6b). These are located mainly on the concave side of the body and in the acidic head (Fig. 6c). Inspection of the surface representation suggests that the β-sheet on the concave side of NusA-NTD$^{\Delta}$ is the center of the interaction surface, although it contains only a

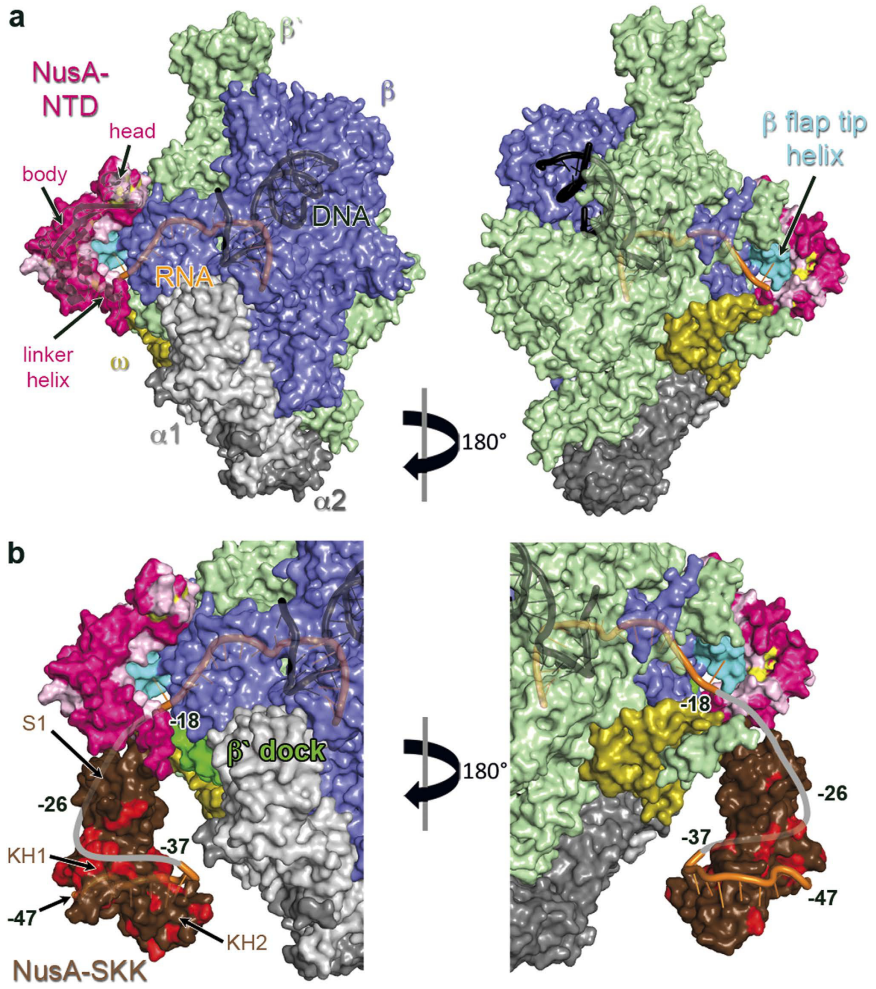


Figure 7. Model for the binding of NusA-NTD$^{\Delta}$ to elongating RNAP. (**a**) NusA-NTD$^{\Delta}$ (cartoon and surface representation, pink) is docked to elongating *Tt*RNAP (PDB ID: 2O5I, surface representation). Residues in NusA-NTD$^{\Delta}$ that are affected by RNAP binding are highlighted in yellow and two amino acids on either side of an affected Ile/Leu/Val residue are colored in light pink unless they were unaffected Ile/Leu/Val residues. α_1, light grey; α_2, dark grey; β, blue; β′, pale green; ω, olive; β flap tip helix, teal; RNA, orange; DNA, black. (**b**) Binding of exiting RNA by NusA. The orientation of NusA-NTD$^{\Delta}$ is the same as in (**a**), the position of *Tm*NusA-SKK was modeled by superposing *Tm*NusA-NTD (PDB ID: 1L2F) on NusA-NTD$^{\Delta}$. RNA was taken from the *Mt*NusA-SKK:RNA complex (PDB ID: 2ASB). Representation of NusA-NTD$^{\Delta}$, *Tt*RNAP and nucleic acids as in (**a**). The β′ dock domain is highlighted in green. *Tm*NusA-SKK (brown) is in surface representation with residues affected by RNA binding highlighted in red according to Schweimer *et al.*[4]. The grey line shows a possible path of exiting RNA, the estimated base numbers are indicated.

limited number of Ile, Leu, or Val residues resulting in a low structural resolution (Fig. 6d). Our binding site is in accordance with cleavage experiments using NusA variants NusA(S29C) and NusA(S53C), that indicated that S29 is located in the NusA:RNAP interface, while S53 is at the opposite side of NusA-NTD (Fig. 6d)[27]. Moreover, our results generally agree with mutational analyses showing that the concave side of the β-sheet is involved in NusA-NTD:β flap interaction[28].

Model of the NusA:RNAP complex. NusA has various effects on transcription elongation and termination with the NusA-NTD:RNAP interaction being probably one key step within the regulatory mechanism[27]. NusA-NTD contacts the RNA exit channel by binding to the β flap tip helix of the β flap region, but the resolution of the electron microscopy structure of a NusA-NTD:RNAP complex was too low to unambiguously determine the orientation of NusA-NTD bound to RNAP[26]. Cleavage and crosslinking experiments on the one hand and mutational analyses as well as NMR studies on *Bs*NusA-NTD and a short β flap construct on the other hand lead to two binding models[27,28].

We used our NMR data to dock NusA-NTD$^{\Delta}$ to the β flap tip helix of elongating *Thermus thermophilus* RNAP (*Tt*RNAP, PDB ID: 2O5I) using HADDOCK[41] (Fig. 7a). In the model most reliable according to HADDOCK, the body of NusA-NTD$^{\Delta}$ binds the β flap tip helix *via* its concave side, which

is in accordance with other models[27,28]. The body is oriented towards the RNA exit channel so that the globular head interacts with the β′ subunit, the latter being in agreement with previous findings that the β′ subunit might also be involved in NusA-NTD binding[20,40]. This orientation allows a tight interaction with the *Tt*RNAP and is similar to the orientations suggested in earlier models[27,28], although the absolute position of NusA-NTD$^{\Delta}$ strongly depends on the residues chosen as restraints and the position of the β flap tip helix.

Next, we integrated the NusA-SKK domain into the model (Fig. 7b). As the structure of *E. coli* NusA-SKK is not available and as the position of the linker helix is similar in *Pl*NusA and NusA-NTD$^{\Delta}$, we first used the crystal structure of *Pl*NusA as template. This, however, led to heavy steric clashes of the *Pl*NusA-SKK domain and *Tt*RNAP which could be prevented by rotating the *Pl*NusA-SKK domain away from the *Tt*RNAP, using the 3-4 residues following the linker helix as anchor. Alternatively, the linker helix itself might rotate slightly. Thus, we modeled the position of *Tm*NusA-SKK by superposing *Tm*NusA-NTD (PDB ID: 1L2F) on NusA-NTD$^{\Delta}$, and we added a short piece of RNA from the *Mt*NusA-SKK:RNA complex structure (PDB ID: 2ASB, Fig. 7b). Either way, the NusA-SKK domain can be positioned correctly for RNA binding. As NusA-NTD is necessary and sufficient for enhancing transcriptional pausing and recognizes duplex RNA[27], exiting RNA might first contact a basic patch on the helical bundle of the NusA-NTD body (Supplementary Fig. 4), which is in direct vicinity of the RNAP exit channel. The RNA then wraps around the NusA-SKK domain, which, in turn, recognizes specific RNA signals (Fig. 7b)[4,42,43]. Crosslinking experiments showed that the RNA region −16 to −23 lies near the NusA-NTD in full-length NusA and that the −34 to −40 region of exiting RNA contacts the NusA-KH2 domain[27], which is consistent with our model. Moreover, the NusA-S1 domain is placed in the vicinity of the β′ dock domain, being in accordance with a genetically shown NusA-S1:β′ dock interaction[44] and cleavage experiments using Fe(III)-(*S*)-2-[4-(2-bromoacetamido)benzyl]ethylenediaminetetraacetic acid (FeBABE)[27]. The position of the C-terminus of NusA-SKK roughly orientates the two NusA-AR domains towards the α-subunits of RNAP and thus localizes NusA-AR2 close to the α-CTD, sterically simplifying a NusA-AR2:α-CTD interaction[4].

Finally, it has been speculated that reorientation of helix α4 stabilizes RNA hairpins[28]. However, not only does NusA exhibit large conformational plasticity, but, in addition, the β flap tip helix is also a highly mobile element[28]. During the transcription cycle the flexibility of the β flap tip helix is important for the regulation of the size of the RNA exit channel, of which the β flap forms the outer wall. Thus, we suggest that the orientation of NusA-NTD bound to RNAP as well as the position of helix α4 may vary, depending on the position of the β flap tip helix. Moreover, this structural flexibility is complemented by the other NusA domains, which are all elastically connected.

Outlook. In this conceptually simple single-experiment approach to identify the RNAP interaction surface of transcription factors with NMR spectroscopy (i) complete RNAP is used, (ii) probes in the transcription factor are directly monitored and, most importantly, (iii) none of the interaction partners needs to be modified. In the future, the method will be refined and used to study these interactions in more detail. Moreover, this approach is very general and can thus be transferred to other systems, with a small binding partner interacting with a supramolecular complex.

Materials and Methods

Cloning. The gene coding for *Ec*NusA-NTD(1–137) was cloned into pET19b *via Blp*I and *Bam*HI. The resulting *E. coli* expression vector pET19b_NusA-NTD_1-137 codes for a His_9 tag fused to the N-terminus of NusA-NTD, cleavable by PreScission protease.

Gene expression and protein purification. NusG-NTD was produced and purified as described[45], as was NusA-NTD$^{\Delta 20}$, the NusB:NusE$^{\Delta}$ complex[34,46] and RNAP[20].

Expression of *nusA-NTD(1–137)* was carried out in *E. coli* BL21 (λ DE3) (Novagen, Madison, WI, USA) harboring pET19b_NusA-NTD_1-137. Lysogeny broth (LB) medium supplemented with 100 μg/ml ampicillin was inoculated with a preculture to an optical density at 600 nm (OD_{600}) of 0.2 and cells were grown at 37 °C until they reached an OD_{600} of 0.7. The temperature was lowered to 20 °C and 30 min later overexpression was induced with 2 mM IPTG. After overnight growth, cells were harvested by centrifugation (9,000 x *g*, 15 min, 4 °C) and dissolved in 20 mM tris(hydroxymethyl)aminomethane (Tris)/HCl (pH 7.9), 100 mM NaCl, 10% (v/v) glycerol, 5 mM β-mercaptoethanol, 10 mM imidazole (buffer A). Cell disruption was carried out with a microfluidizer (Microfluidics, Newton, MA, USA). Having centrifuged the lysate (12,000 × *g*, 30 min, 4 °C), the supernatant was applied to a Ni-NTA column (Qiagen, Hilden, Germany), and subsequently the column was washed with buffer A. A step gradient with increasing imidazole concentrations (10–500 mM in buffer A) was used for elution. Fractions containing His_9-NusA-NTD(1–137) were combined and cleaved during overnight dialysis against 50 mM Tris/HCl (pH 8.0), 150 mM NaCl (molecular weight cut-off (MWCO) 3,500 Da) by PreScission protease (GE Healthcare, Munich, Germany). The protein solution was then dialyzed against 50 mM Tris (pH 7.4), 1 mM dithiothreitol (DTT, buffer B) and reapplied to the Ni-NTA column connected to a QXL FF column (GE Healthcare, Munich, Germany). After washing with buffer B, the Ni-NTA column was removed and the QXL FF column was eluted using a step gradient with increasing NaCl concentrations (0–1 M NaCl in buffer B). Fractions containing pure NusA-NTD(1-137) were dialyzed against the

required buffer, concentrated by ultrafiltration (MWCO 3,000 Da) and stored at −80 °C after freezing with liquid nitrogen.

Proteins were uniformly labeled with ^{15}N or ^{15}N,^{13}C by growing *E. coli* in M9 minimal medium[41,42] with addition of $(^{15}NH_4)_2SO_4$ (Campro Scientific, Berlin, Germany) or $(^{15}NH_4)_2SO_4$ and ^{13}C-D-glucose (Spectra Stable Isotopes, Columbia, MD, USA) as only nitrogen and carbon source. Expression and purification was the same as for proteins produced in LB medium. Methyl group labeling of Ile, Leu and Val residues with [^1H,^{13}C] in deuterated proteins was performed as described previously[20].

NMR spectroscopy. NMR spectroscopic experiments were conducted on Bruker *Avance* 600 MHz, 700 MHz and 800 MHz spectrometers, the latter two equipped with cryogenically cooled probes. For resonance assignment of NusA-NTD$^{\Delta}$, standard double and triple resonance through-bond experiments were recorded[47,48]. The protein was in 10 mM potassium phosphate buffer (pH 6.4) containing 50 mM NaCl at 298 K. NMR data were processed using in-house routines (Apodization, Fourier transformation, phase correction and baseline correction) and visualized with NMRView[49]. Distance restraints for structure calculation were derived from [^{15}N]-edited and [^{13}C]-edited NOESY spectra with mixing times of 100–120 ms. NOESY cross peaks were classified according to their relative intensities and converted to distance restraints with the following upper limits: 3.0 Å, strong; 4.0 Å, medium; 5.0 Å, weak; 6.0 Å, very weak. Experimental NOESY spectra were validated semi-quantitatively against back-calculated spectra to confirm the assignment and to avoid bias of upper distance restraints by spin-diffusion. Hydrogen bonds were included for backbone amide protons in regular secondary structure if the amide proton did not show a water exchange cross peak in the [^{15}N]-edited NOESY spectrum. Backbone dihedral restraints were obtained from chemical shift data by using TALOS[50]. Existence of a hydrogen bond was assumed if the acceptor of a slowly exchanging amide proton, based on the absence of a water exchange peak in the [^{15}N]-edited NOESY spectrum, could be identified unambiguously from the results of initial structure calculations. For each hydrogen bond the distance between the amide proton and the acceptor was restrained to less than 2.3 Å and the distance between the amide nitrogen and the acceptor to less than 3.1 Å.

The structure calculation was performed with the program XPLOR-NIH 2.1.2[51] using a three-step simulated annealing protocol with floating assignment of prochiral groups including a conformational database potential[52]. For the final iteration 80 structures were calculated, the 20 structures of lowest energy were accepted and further analyzed with the programs XPLOR-NIH 2.1.2 and PROCHECK-NMR[53].

TROSY spectra[29] were recorded using [I,L,V]-labeled protein samples (20 μM) in 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, pH 7.5), 50 mM NaCl, 5% (v/v) glycerol, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 10 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM DTT in 99.9% D_2O at 298 K. Unlabeled, protonated RNAP in the same buffer was added in two steps (ratios 1:1, 1:2). Non-stereo-specific assignments of methyl groups of NusG-NTD and NusE$^{\Delta}$ were taken from previous studies[10,46]. Signal intensities were normalized by protein concentration and number of scans. As pulse lengths changed less than 1% upon RNAP addition, the influence of these changes on the intensity were neglected. For each titration step the ratio of remaining signal intensities and signal intensities in the spectrum of the free transcription factor were calculated, yielding relative signal intensities. Next, the mean value of all relative 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 slightly affected. Additionally, Leu and Val residues were considered as affected, when at least one of the two signals showed a significant intensity decrease. Only unambiguously assigned signals were used in the analysis.

Proteins for the displacement experiments of [^{15}N]-NusE$^{\Delta}$:NusB from RNAP by NusG-CTD and of NusE$^{\Delta}$:NusB from [^{15}N]-NusG-CTD by RNAP were in 25 mM HEPES, pH 7.5, 100 mM NaCl at 298 K. Separate samples for [^{15}N]-NusE$^{\Delta}$:NusB (50 μM) and [^{15}N]-NusE$^{\Delta}$:NusB:RNAP (25 μM each) were prepared. For the displacement experiments NusG-CTD was added (stock concentration: 1050 μM). Similarly, separate samples for [^{15}N]-NusG-CTD (50 μM) and [^{15}N]-NusG-CTD: NusE$^{\Delta}$:NusB (25 μM each) were prepared. For the displacement experiments RNAP was added from a 117 μM stock. The titrations were followed by recording 1D or 2D [^1H,^{15}N]-HSQC spectra after each titration step. 1D spectra were normalized by protein concentration and number of scans. As pulse lengths changed less than 1% upon RNAP addition, the influence of these changes on the intensity were neglected.

Docking and Molecular Modeling. The NusG-NTD:RNAP complex was generated based on the crystal structure of Spt4/5 bound to the clamp domain from *P. furiosus* (PDB ID: 3QQC). *E. coli* NusG-NTD (PDB ID: 2K06, model 1) was superposed on Spt5 (chain D, root mean square deviation (r.m.s.d.) 1.2 Å). *Ec*RNAP (PDB ID: 4KMU) was positioned by superposing the β′ subunit (chain D) on the clamp domain (chain A, r.m.s.d. 2.4 Å).

Docking of NusA-NTD$^{\Delta}$ (model 1) to elongating *Tt*RNAP (PDB ID: 2O5I) was carried out using the HADDOCK webserver[41]. Residues in NusA-NTD$^{\Delta}$ that were experimentally determined to be affected by RNAP binding (Leu27, Leu31, Ile43, Val45) were defined as active residues. Solvent exposed residues in the β flap tip helix were chosen as active residues (chain C, residues Arg772, Leu773, Ser776, Ile777). Passive residues were automatically determined by HADDOCK. The coordinates of the β flap tip helix in the docked complex relative to the deposited coordinates of NusA-NTD$^{\Delta}$ are shown in Supplementary

Table 1. After docking NusA-NTD$^{\Delta}$ to *Tt*RNAP, the position of the NusA-SKK domain was modeled with two alternative procedures. First, *Pl*NusA (PDB ID: 4MTN) was superposed on NusA-NTD$^{\Delta}$ (residues G3-D73 of *Pl*NusA; residues Met1-Thr101 of NusA-NTD$^{\Delta}$). To avoid clashes with *Tt*RNAP the *Pl*NusA-SKK was rotated manually around residues in the linker between *Pl*NusA-NTD and *Pl*NusA-SKK (residues Arg107-Gln109) using PyMOL[54]. In the second approach *Tm*NusA (PDB ID: 1L2F) was superposed on NusA-NTD$^{\Delta}$ using residues 1–101. Finally, the *Mt*NusA-SKK:RNA complex (PDB ID: 2ASB, residues Ser108-Gly333 of *Mt*NusA-SKK) was superposed on *Tm*NusA-SKK (residues Glu132-Leu344) to position the RNA. RNA base numbers were estimated.

Programs. All structures were visualized with PyMOL[54]. The Adaptive Poisson-Boltzmann Solver (APBS)-Plugin and the PDB2PQR server were used for the determination of the charge surface potential[55,56]. Superpositions of different NusA-NTDs were done with LSQMAN[57], omitting the linker helix (residues Met1-Thr101 of NusA-NTD$^{\Delta}$, residues Met1-Asn101 of *Tm*NusA (PDB ID: 1L2F, 1HH2), residues Met1-Asp101 of *Bs*NusA (PDB ID: 2MT4), residues Met1-Phe79 of *Mt*NusA (PDB ID: 2K0R), residues Gly3-Asp73 of *Pl*NusA (PDB ID: 4MTN)). All other superpositions were carried out by PyMOL[54].

References

1. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9,** 85–98 (2011).
2. Ito, K., Iwakura, Y. & Ishihama, A. Biosynthesis of RNA polymerase in *Escherichia coli*. III. Identification of intermediates in the assembly of RNA polymerase. *J. Mol. Biol.* **96,** 257–271 (1975).
3. Jeon, Y. H., Yamazaki, T., Otomo, T., Ishihama, A. & Kyogoku, Y. Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain. *J. Mol. Biol.* **267,** 953–962 (1997).
4. Schweimer, K. *et al.* NusA interaction with the α-subunit of *E. coli* RNA Polymerase is *via* the UP-element site and releases autoinhibition. *Structure* **19,** 945–954 (2011).
5. Mustaev, A. *et al.* Modular organization of the catalytic center of RNA polymerase. *Proc. Natl. Acad. Sci. USA.* **94,** 6641–6645 (1997).
6. Zaychikov, E. *et al.* Mapping of catalytic residues in the RNA polymerase active center. *Science* **273,** 107–109 (1996).
7. Ghosh, P., Ishihama, A. & Chatterji, D. *Escherichia coli* RNA polymerase subunit omega and its N-terminal domain bind full-length beta′ to facilitate incorporation into the alpha2beta subassembly. *Eur. J. Biochem.* **268,** 4621–4627 (2001).
8. Haugen, S. P., Ross, W. & Gourse, R. L. Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat. Rev. Microbiol.* **6,** 507–519 (2008).
9. 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).
10. Mooney, R. A., Schweimer, K., Rösch, P., Gottesman, M. E. & 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).
11. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. USA.* **97,** 7090–7095 (2000).
12. 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).
13. 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).
14. Burmann, B. M. *et al.* A NusE:NusG Complex Links Transcription and Translation. *Science* **328,** 501–504 (2010).
15. Li, J., Mason, S. W. & Greenblatt, J. Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. *Genes Dev.* **7,** 161–172 (1993).
16. Mason, S. W. & Greenblatt, J. Assembly of transcription elongation complexes containing the N protein of phage lambda and the *Escherichia coli* elongation factors NusA, NusB, NusG, and S10. *Genes Dev.* **5,** 1504–1512 (1991).
17. Aksoy, S., Squires, C. L. & Squires, C. Evidence for antitermination in *Escherichia coli* RRNA transcription. *J. Bacteriol.* **159,** 260–4 (1984).
18. Gottesman, M. E. & Weisberg, R. A. Little lambda, who made thee? *Microbiol. Mol. Biol. Rev.* **68,** 796–813 (2004).
19. Nodwell, J. R. & Greenblatt, J. Recognition of *boxA* antiterminator RNA by the *E. coli* antitermination factors NusB and ribosomal protein S10. *Cell* **72,** 261–268 (1993).
20. Drögemüller, J. *et al.* Exploring RNA polymerase regulation by NMR spectroscopy. *Sci. Rep.* **5,** 10825–10835 (2015).
21. Eisenmann, A., Schwarz, S., Prasch, S., Schweimer, K. & Rösch, P. The *E. coli* NusA carboxy-terminal domains are structurally similar and show specific RNAP- and lambdaN interaction. *Protein Sci.* **14,** 2018–29 (2005).
22. Worbs, M., Bourenkov, G. P., Bartunik, H. D., Huber, R. & Wahl, M. C. An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Mol. Cell* **7,** 1177–1189 (2001).
23. Borukhov, S., Lee, J. & Laptenko, O. Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol. Microbiol.* **55,** 1315–1324 (2005).
24. Roberts, J. W., Shankar, S. & Filter, J. J. RNA polymerase elongation factors. *Annu. Rev. Microbiol.* **62,** 211–233 (2008).
25. Mah, T. F., Li, J., Davidson, A. R. & Greenblatt, J. Functional importance of regions in *Escherichia coli* elongation factor NusA that interact with RNA polymerase, the bacteriophage lambda N protein and RNA. *Mol. Microbiol.* **34,** 523–537 (1999).
26. Yang, X. *et al.* The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep.* **10,** 997–1002 (2009).
27. Ha, K. S., Toulokhonov, I., Vassylyev, D. G. & Landick, R. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing *via* interaction with the RNA exit channel of RNA polymerase. *J. Mol. Biol.* **401,** 708–725 (2010).
28. Ma, C. *et al.* RNA polymerase-induced remodelling of NusA produces a pause enhancement complex. *Nucl. Acids Res.* **43,** 2829–2840 (2015).
29. Tugarinov, V. & Kay, L. E. An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR* **28,** 165–172 (2004).
30. Epshtein, V., Dutta, D., Wade, J. & Nudler, E. An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463,** 245–249 (2010).
31. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292,** 1876–1882 (2001).

32. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. Structural basis of transcriptional pausing in bacteria. *Cell* **152,** 431–441 (2013).
33. Davis, C. A., Bingman, C. A., Landick, R., Record, M. T. Jr & Saecker, R. M. 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,** 7833–7838 (2007).
34. Luo, X. *et al.* Structural and functional analysis of the *E. coli* NusB-S10 transcription antitermination complex. *Mol. Cell* **32,** 791–802 (2008).
35. Schuwirth, B. S. *et al.* Structures of the Bacterial Ribosome at 3.5 A Resolution. *Science* **310,** 827–834 (2005).
36. Greive, S. J., Lins, A. F. & von Hippel, P. H. Assembly of an RNA-protein complex. Binding of NusB and NusE (S10) proteins to *boxA* RNA nucleates the formation of the antitermination complex involved in controlling rRNA transcription in *Escherichia coli*. *J. Biol. Chem.* **280,** 36397–36408 (2005).
37. Vogel, U. & Jensen, K. F. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J. Bacteriol.* **176,** 2807–2813 (1994).
38. Shin, D. H. *et al.* Crystal structure of NusA from *Thermotoga maritima* and functional implication of the N-terminal domain. *Biochemistry* **42,** 13429–13437 (2003).
39. Gopal, B. *et al.* Crystal structure of the transcription elongation/anti-termination factor NusA from *Mycobacterium tuberculosis* at 1.7 A resolution. *J. Mol. Biol.* **314,** 1087–1095 (2001).
40. Traviglia, S. L., Datwyler, S. A., Yan, D., Ishihama, A. & Meares, C. F. Targeted protein footprinting: where different transcription factors bind to RNA polymerase. *Biochemistry* **38,** 15774–15778 (1999).
41. de Vries, S. J., van Dijk, M. & Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5,** 883–897 (2010).
42. Beuth, B., Pennell, S., Arnvig, K. B., Martin, S. R. & Taylor, I. A. Structure of a *Mycobacterium tuberculosis* NusA-RNA complex. *EMBO J.* **24,** 3576–3587 (2005).
43. Prasch, S. *et al.* RNA-binding specificity of *E. coli* NusA. *Nucl. Acids Res.* **37,** 4736–4742 (2009).
44. Ito, K. & Nakamura, Y. Localization of *nusA*-suppressing amino acid substitutions in the conserved regions of the beta' subunit of *Escherichia coli* RNA polymerase. *Mol. Gen. Genet.* **251,** 699–706 (1996).
45. 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).
46. Burmann, B. M., Luo, X., Wahl, M. C., Rösch, P. & Gottesman, M. E. Fine tuning of the *E. coli* NusB:NusE complex affinity to *BoxA* RNA is required for processive antitermination. *Nucl. Acids Res.* **38,** 314–326 (2010).
47. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectrosc.* **34,** 93–158 (1999).
48. Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Rance, M. & Skelton, N. J. in Protein NMR spectroscopy: principles and practice (Academic Press, Boston, 2007).
49. Johnson, B. A. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278,** 313–352 (2004).
50. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13,** 289–302 (1999).
51. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160,** 66–74 (2003).
52. Schweimer, K. *et al.* Structural investigation of the binding of a herpesviral protein to the SH3 domain of tyrosine kinase Lck. *Biochemistry* **41,** 5120–5130 (2002).
53. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8,** 477–486 (1996).
54. Schrödinger L. The PyMOL molecular graphics system, version 1.3. *Schrödinger, LLC, Mannheim, Germany* (2010).
55. 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).
56. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucl. Acids Res.* **32,** W665–7 (2004).
57. Kleywegt, G. J. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr. D Biol. Crystallogr.* **52,** 842–857 (1996).

Acknowledgements

We thank Ramona Heissmann for excellent technical assistance and Björn M. Burmann as well as Stefan Prasch for initial work on NusA-NTD and NusA-NTD$^{\Delta}$. The work was supported by grants Ro 617/17-1 and Ro 617/21-1 (to P.R.) from the Deutsche Forschungsgemeinschaft, and the Ludwig-Schaefer-Scholarschip 2015 from Columbia University Medical Center (to P.R.). We also thank Björn M. Burmann for carefully reading the manuscript and helpful discussions.

Author Contributions

P.R. and S.H.K. conceived and designed the project. M.S., J.D. and S.H.K. designed the NMR experiments. J.D. and M.S. conducted the experiments. M.J. and K.S. solved the NusA-NTD structure. J.D., M.S., K.S., S.H.K. and P.R. wrote the manuscript.

Additional Information

Accession codes: Chemical shifts of NusA-NTD$^{\Delta}$ have been deposited in the Biological Magnetic Resonance Bank Databank, accession number 16868. The atomic coordinates of the NusA-NTD$^{\Delta}$ structure have been deposited in the Protein Data Bank, accession number 2KWP.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Drögemüller, J. *et al.* Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci. Rep.* **5**, 16428; doi: 10.1038/srep16428 (2015).

Supplementary Information

Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy

Johanna Drögemüller[1,§], Martin Strauß[1,§], Kristian Schweimer[1], Marcel Jurk[2], Paul Rösch[1], Stefan H Knauer[1,*]

[1] Lehrstuhl Biopolymere und Forschungszentrum für Bio-Makromoleküle, Universität Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany

[2] present address: Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany

[§] These authors contributed equally to this work

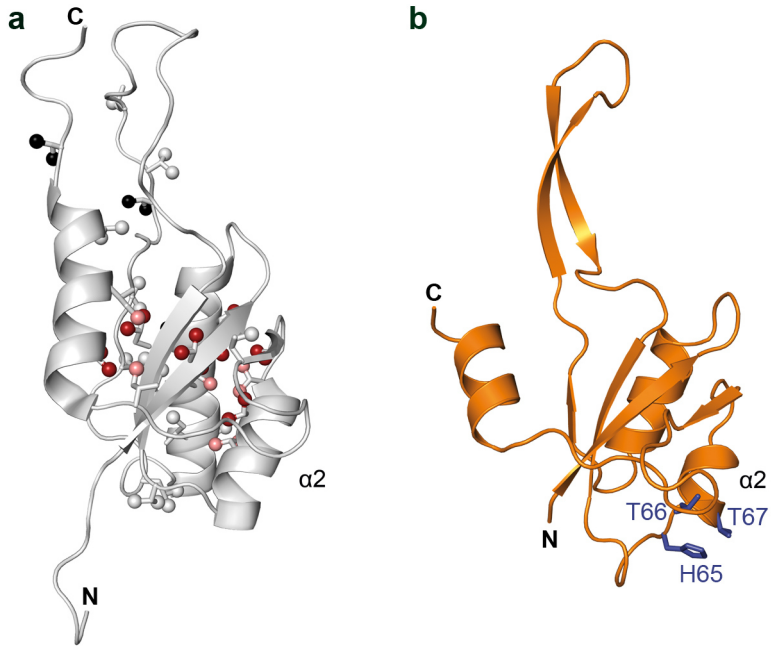
* corresponding author

Contents

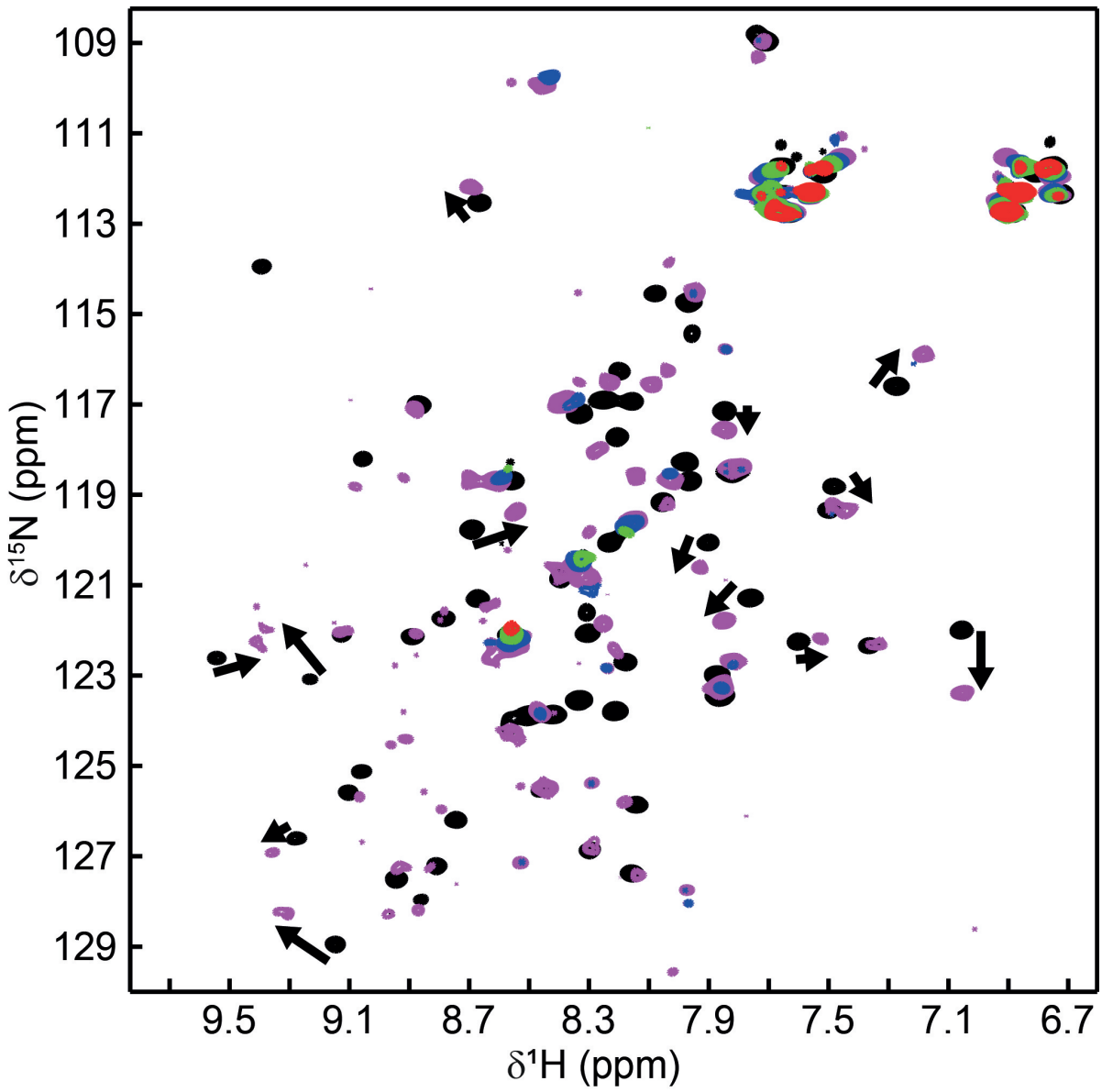
Supplementary Table 1: Coordinates of the β flap tip helix in the modeled NusA-NTD$^{\Delta}$:*Tt*RNAP complex. The table is an extract of the PDB file of elongating *Tt*RNAP (residues 767-781 of the β subunit; PDB ID: 2O5I) docked to NusA-NTD$^{\Delta}$ as described in the Material and Methods section, giving the position of *Tt*RNAP relative to the deposited coordinates of NusA-NTD$^{\Delta}$ (PDB ID: 2KWP).

```
ATOM  10777  N   PRO C 767       3.453  -1.849 -42.571  1.00 90.52           N
ATOM  10778  CA  PRO C 767       2.398  -2.073 -41.584  1.00 91.06           C
ATOM  10779  C   PRO C 767       2.633  -1.165 -40.362  1.00 92.21           C
ATOM  10780  O   PRO C 767       3.412  -0.213 -40.472  1.00 92.68           O
ATOM  10781  CB  PRO C 767       2.524  -3.559 -41.274  1.00 90.38           C
ATOM  10782  CG  PRO C 767       3.028  -4.129 -42.533  1.00 90.03           C
ATOM  10783  CD  PRO C 767       4.098  -3.134 -42.897  1.00 90.44           C
ATOM  10784  N   THR C 768       2.013  -1.418 -39.212  1.00 92.37           N
ATOM  10785  CA  THR C 768       2.261  -0.517 -38.082  1.00 93.42           C
ATOM  10786  C   THR C 768       2.739  -1.281 -36.849  1.00 94.34           C
ATOM  10787  O   THR C 768       2.104  -2.254 -36.442  1.00 94.97           O
ATOM  10788  CB  THR C 768       0.994   0.286 -37.704  1.00 92.74           C
ATOM  10789  OG1 THR C 768       0.329   0.726 -38.894  1.00 92.51           O
ATOM  10790  CG2 THR C 768       1.369   1.533 -36.866  1.00 91.14           C
ATOM  10791  N   PRO C 769       3.867  -0.844 -36.239  1.00 95.40           N
ATOM  10792  CA  PRO C 769       4.475  -1.458 -35.040  1.00 95.44           C
ATOM  10793  C   PRO C 769       3.577  -1.368 -33.769  1.00 95.51           C
ATOM  10794  O   PRO C 769       3.583  -2.294 -32.943  1.00 95.59           O
ATOM  10795  CB  PRO C 769       5.803  -0.706 -34.892  1.00 96.45           C
ATOM  10796  CG  PRO C 769       6.147  -0.342 -36.332  1.00 95.78           C
ATOM  10797  CD  PRO C 769       4.806   0.125 -36.849  1.00 95.78           C
ATOM  10798  N   GLU C 770       2.827  -0.273 -33.609  1.00 93.98           N
ATOM  10799  CA  GLU C 770       1.962  -0.077 -32.434  1.00 93.34           C
ATOM  10800  C   GLU C 770       0.563  -0.681 -32.666  1.00 93.16           C
ATOM  10801  O   GLU C 770      -0.118  -1.081 -31.718  1.00 92.30           O
ATOM  10802  CB  GLU C 770       1.849   1.427 -32.127  1.00 92.37           C
ATOM  10803  CG  GLU C 770       3.152   2.249 -32.385  1.00 90.19           C
ATOM  10804  CD  GLU C 770       3.957   2.594 -31.128  1.00 88.39           C
ATOM  10805  OE1 GLU C 770       4.437   1.679 -30.437  1.00 87.21           O
ATOM  10806  OE2 GLU C 770       4.113   3.798 -30.832  1.00 87.50           O
ATOM  10807  N   GLU C 771       0.159  -0.730 -33.938  1.00 93.66           N
ATOM  10808  CA  GLU C 771      -1.116  -1.314 -34.376  1.00 95.22           C
ATOM  10809  C   GLU C 771      -0.991  -2.838 -34.494  1.00 96.29           C
ATOM  10810  O   GLU C 771      -1.993  -3.557 -34.560  1.00 96.81           O
ATOM  10811  CB  GLU C 771      -1.531  -0.719 -35.743  1.00 95.59           C
ATOM  10812  CG  GLU C 771      -2.798  -1.302 -36.423  1.00 95.80           C
ATOM  10813  CD  GLU C 771      -2.986  -0.867 -37.893  1.00 96.05           C
ATOM  10814  OE1 GLU C 771      -4.129  -0.944 -38.403  1.00 95.71           O
ATOM  10815  OE2 GLU C 771      -1.999  -0.457 -38.544  1.00 95.66           O
ATOM  10816  N   ARG C 772       0.255  -3.318 -34.526  1.00 95.94           N
ATOM  10817  CA  ARG C 772       0.562  -4.755 -34.598  1.00 96.21           C
ATOM  10818  C   ARG C 772       0.310  -5.431 -33.247  1.00 97.11           C
ATOM  10819  O   ARG C 772      -0.155  -6.577 -33.193  1.00 97.59           O
ATOM  10820  CB  ARG C 772       2.031  -4.975 -34.992  1.00 94.95           C
ATOM  10821  CG  ARG C 772       2.497  -6.441 -35.071  1.00 92.48           C
ATOM  10822  CD  ARG C 772       1.926  -7.149 -36.290  1.00 91.56           C
ATOM  10823  NE  ARG C 772       1.943  -6.309 -37.490  1.00 90.45           N
ATOM  10824  CZ  ARG C 772       1.929  -6.773 -38.739  1.00 90.61           C
ATOM  10825  NH1 ARG C 772       1.906  -8.079 -38.978  1.00 91.27           N
ATOM  10826  NH2 ARG C 772       1.936  -5.935 -39.759  1.00 89.96           N
ATOM  10827  N   LEU C 773       0.629  -4.710 -32.164  1.00 97.02           N
ATOM  10828  CA  LEU C 773       0.423  -5.199 -30.800  1.00 97.16           C
ATOM  10829  C   LEU C 773      -1.069  -5.199 -30.435  1.00 96.83           C
ATOM  10830  O   LEU C 773      -1.557  -6.174 -29.874  1.00 97.78           O
ATOM  10831  CB  LEU C 773       1.212  -4.334 -29.791  1.00 96.99           C
ATOM  10832  CG  LEU C 773       2.759  -4.388 -29.822  1.00 96.94           C
```

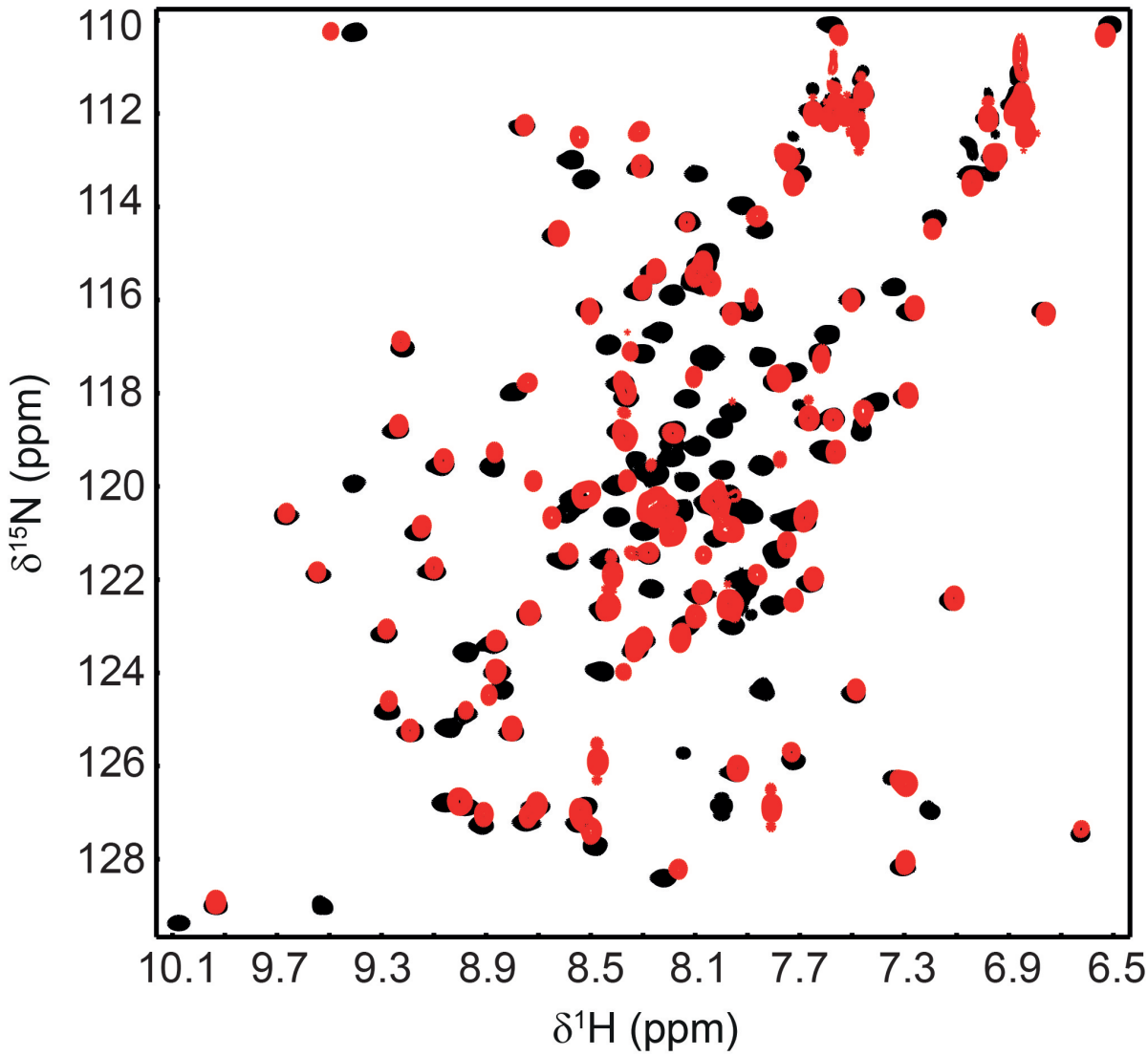
```
ATOM  10833  CD1 LEU C 773       3.329  -3.401 -28.800  1.00 96.92           C
ATOM  10834  CD2 LEU C 773       3.246  -5.800 -29.502  1.00 97.02           C
ATOM  10835  N   LEU C 774      -1.789  -4.120 -30.755  1.00 95.96           N
ATOM  10836  CA  LEU C 774      -3.229  -4.033 -30.473  1.00 95.46           C
ATOM  10837  C   LEU C 774      -3.998  -5.145 -31.174  1.00 95.47           C
ATOM  10838  O   LEU C 774      -4.830  -5.817 -30.562  1.00 96.22           O
ATOM  10839  CB  LEU C 774      -3.806  -2.692 -30.942  1.00 93.90           C
ATOM  10840  CG  LEU C 774      -5.346  -2.636 -30.934  1.00 92.29           C
ATOM  10841  CD1 LEU C 774      -5.830  -2.770 -29.504  1.00 91.81           C
ATOM  10842  CD2 LEU C 774      -5.833  -1.345 -31.545  1.00 91.22           C
ATOM  10843  N   ARG C 775      -3.735  -5.325 -32.462  1.00 95.29           N
ATOM  10844  CA  ARG C 775      -4.402  -6.381 -33.208  1.00 94.68           C
ATOM  10845  C   ARG C 775      -4.209  -7.740 -32.520  1.00 94.67           C
ATOM  10846  O   ARG C 775      -5.184  -8.404 -32.214  1.00 94.44           O
ATOM  10847  CB  ARG C 775      -3.860  -6.445 -34.654  1.00 94.56           C
ATOM  10848  CG  ARG C 775      -4.360  -5.341 -35.613  1.00 92.12           C
ATOM  10849  CD  ARG C 775      -5.880  -5.356 -35.749  1.00 90.64           C
ATOM  10850  NE  ARG C 775      -6.342  -4.568 -36.888  1.00 89.53           N
ATOM  10851  CZ  ARG C 775      -7.616  -4.420 -37.234  1.00 88.70           C
ATOM  10852  NH1 ARG C 775      -8.567  -5.009 -36.526  1.00 88.08           N
ATOM  10853  NH2 ARG C 775      -7.937  -3.684 -38.292  1.00 88.26           N
ATOM  10854  N   SER C 776      -2.950  -8.120 -32.264  1.00 94.65           N
ATOM  10855  CA  SER C 776      -2.568  -9.411 -31.644  1.00 94.82           C
ATOM  10856  C   SER C 776      -3.212  -9.627 -30.236  1.00 94.23           C
ATOM  10857  O   SER C 776      -3.590 -10.761 -29.910  1.00 94.33           O
ATOM  10858  CB  SER C 776      -1.021  -9.512 -31.521  1.00 94.77           C
ATOM  10859  OG  SER C 776      -0.360  -9.645 -32.780  1.00 94.97           O
ATOM  10860  N   ILE C 777      -3.334  -8.580 -29.415  1.00 93.70           N
ATOM  10861  CA  ILE C 777      -3.901  -8.698 -28.057  1.00 92.75           C
ATOM  10862  C   ILE C 777      -5.350  -9.252 -28.072  1.00 93.16           C
ATOM  10863  O   ILE C 777      -5.631 -10.255 -27.404  1.00 93.55           O
ATOM  10864  CB  ILE C 777      -3.896  -7.308 -27.305  1.00 92.10           C
ATOM  10865  CG1 ILE C 777      -2.455  -6.821 -27.095  1.00 91.21           C
ATOM  10866  CG2 ILE C 777      -4.534  -7.449 -25.920  1.00 91.63           C
ATOM  10867  CD1 ILE C 777      -2.357  -5.456 -26.433  1.00 90.12           C
ATOM  10868  N   PHE C 778      -6.256  -8.615 -28.820  1.00 93.48           N
ATOM  10869  CA  PHE C 778      -7.673  -9.029 -28.887  1.00 93.33           C
ATOM  10870  C   PHE C 778      -7.928 -10.029 -30.029  1.00 94.16           C
ATOM  10871  O   PHE C 778      -8.818 -10.872 -29.934  1.00 93.84           O
ATOM  10872  CB  PHE C 778      -8.583  -7.804 -29.105  1.00 92.40           C
ATOM  10873  CG  PHE C 778      -8.687  -6.861 -27.918  1.00 91.76           C
ATOM  10874  CD1 PHE C 778      -7.540  -6.454 -27.198  1.00 91.29           C
ATOM  10875  CD2 PHE C 778      -9.943  -6.350 -27.533  1.00 91.20           C
ATOM  10876  CE1 PHE C 778      -7.638  -5.546 -26.104  1.00 90.92           C
ATOM  10877  CE2 PHE C 778     -10.067  -5.440 -26.441  1.00 91.24           C
ATOM  10878  CZ  PHE C 778      -8.906  -5.037 -25.723  1.00 90.85           C
ATOM  10879  N   GLY C 779      -7.153  -9.925 -31.108  1.00 95.41           N
ATOM  10880  CA  GLY C 779      -7.313 -10.826 -32.242  1.00 97.01           C
ATOM  10881  C   GLY C 779      -6.406 -10.473 -33.421  1.00 97.94           C
ATOM  10882  O   GLY C 779      -6.809  -9.697 -34.309  1.00 98.21           O
ATOM  10883  N   GLU C 780      -5.190 -11.038 -33.414  1.00 98.62           N
ATOM  10884  CA  GLU C 780      -4.154 -10.817 -34.448  1.00 98.20           C
ATOM  10885  C   GLU C 780      -4.735 -10.916 -35.859  1.00 97.77           C
ATOM  10886  O   GLU C 780      -4.483 -10.050 -36.707  1.00 97.44           O
ATOM  10887  CB  GLU C 780      -3.002 -11.842 -34.299  1.00 98.74           C
ATOM  10888  CG  GLU C 780      -3.479 -13.315 -34.185  1.00 98.09           C
ATOM  10889  CD  GLU C 780      -2.821 -14.257 -35.186  1.00 97.77           C
ATOM  10890  OE1 GLU C 780      -3.289 -15.415 -35.292  1.00 97.28           O
ATOM  10891  OE2 GLU C 780      -1.848 -13.846 -35.859  1.00 97.54           O
ATOM  10892  N   LYS C 781      -5.507 -11.978 -36.099  1.00 97.33           N
ATOM  10893  CA  LYS C 781      -6.147 -12.202 -37.392  1.00 96.31           C
ATOM  10894  C   LYS C 781      -7.292 -11.208 -37.620  1.00 95.87           C
ATOM  10895  O   LYS C 781      -8.485 -11.533 -37.511  1.00 96.54           O
ATOM  10896  CB  LYS C 781      -6.675 -13.642 -37.491  1.00 95.28           C
ATOM  10897  CG  LYS C 781      -7.331 -14.173 -36.226  1.00 93.86           C
ATOM  10898  CD  LYS C 781      -8.176 -15.409 -36.492  1.00 93.01           C
ATOM  10899  CE  LYS C 781      -7.389 -16.500 -37.197  1.00 92.99           C
ATOM  10900  NZ  LYS C 781      -6.056 -16.712 -36.568  1.00 93.33           N
```



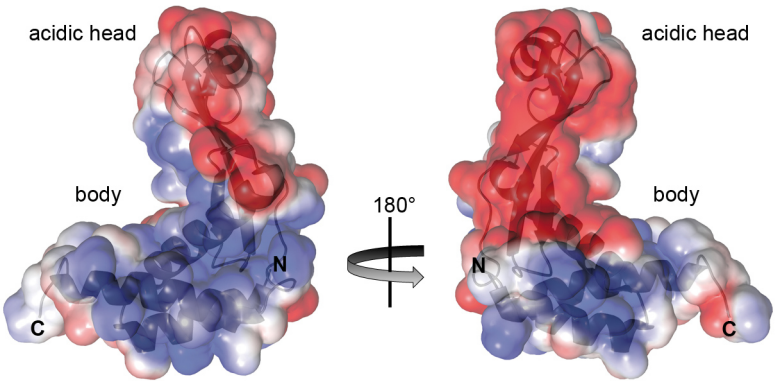
Supplementary Figure 1: Structures of NusG-NTD and RfaH-NTD. (**a**) RNAP binding site of NusG-NTD. Structure of NusG-NTD (PDB ID: 2K06) in cartoon representation, grey. Ile, Leu, and Val residues are shown as sticks with the carbon atoms of their methyl groups represented as spheres. Strongly affected methyl groups, dark red; slightly affected methyl groups, light red; unaffected methyl groups, grey; unassigned methyl groups, black. (**b**) βGL binding motif of RfaH-NTD. Structure of RfaH-NTD (PDB ID: 2OUG) in cartoon representation, orange. Residues involved in βGL binding are shown as blue sticks and labeled.



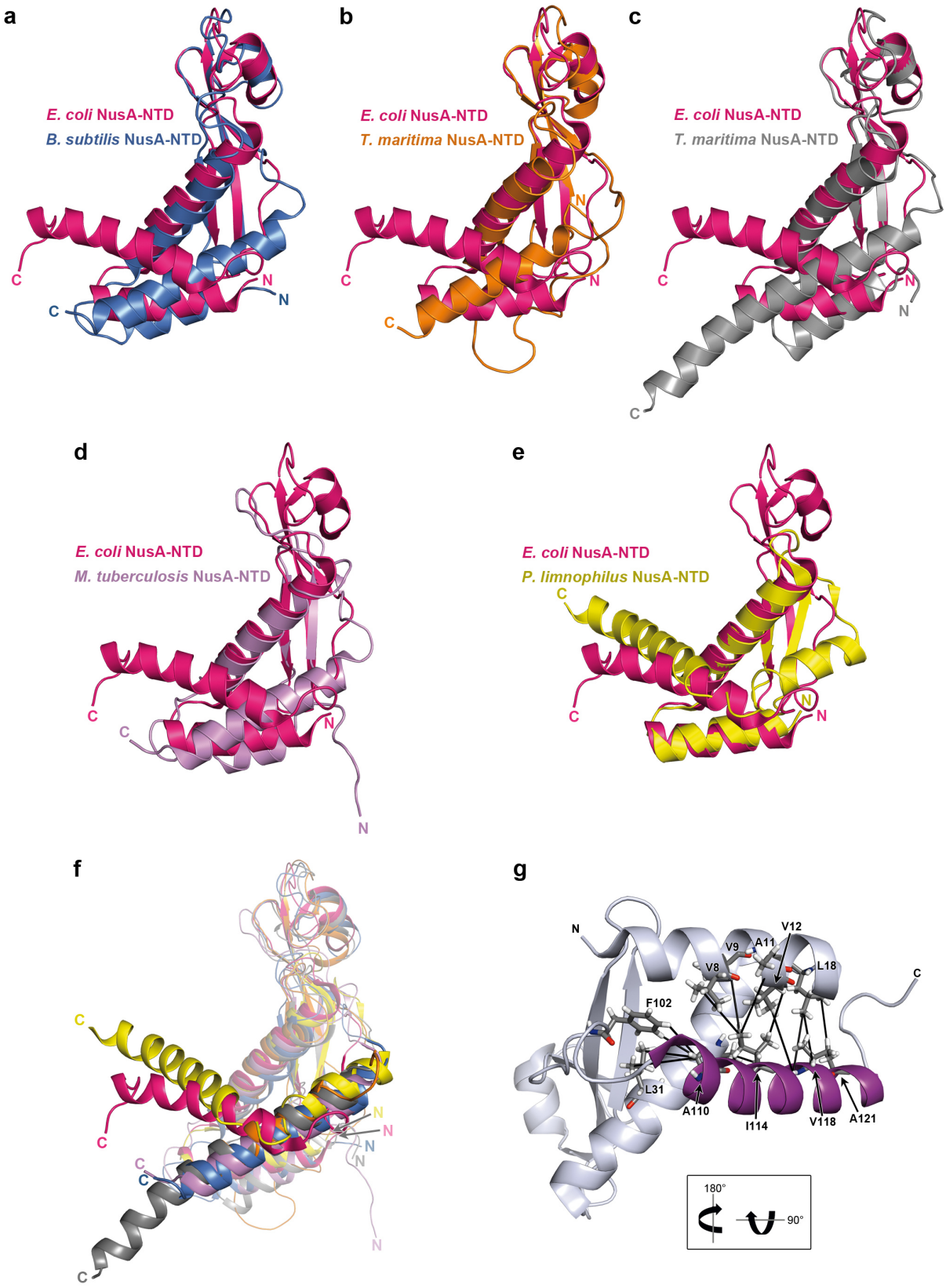
Supplementary Figure 2: Displacement of RNAP from NusE$^{\Delta}$ by NusG-CTD. 2D [^{1}H,^{15}N]-HSQC spectra of free NusB:[^{15}N]-NusE$^{\Delta}$, black, NusB:[^{15}N]-NusE$^{\Delta}$ in the presence of RNAP in equimolar concentration, red, and NusB:[^{15}N]-NusE$^{\Delta}$ in the presence of RNAP and NusG-CTD (molar ratio 1:1:1, green; 1:1:3, blue; 1:1:10, purple). Black arrows indicate the chemical shift changes that occur upon complex formation of NusG-CTD and NusB:[^{15}N]-NusE$^{\Delta 1}$.



Supplementary Figure 3: Superposition of the [^{1}H,^{15}N]-HSQC spectra of [^{15}N]-NusA-NTD(1-137), red, and [^{15}N]-NusA-NTD$^{\Delta}$, black. The protein concentration was 400 µM in each sample.



Supplementary Figure 4: Electrostatic potential molecular surface of NusA-NTD$^{\Delta}$. NusA-NTD$^{\Delta}$ in cartoon and surface representation. The electrostatic surface potential is colored from -2 kT/e, red, to +2 kT/e, blue.



Supplementary Figure 5: Comparison of NusA-NTD structures. (**a-e**) Superposition of NusA-NTD$^{\Delta}$ (pink) with (**a**) *Bs*NusA-NTD (blue, PDB ID: 2MT4, root mean square deviation

(r.m.s.d.) 1.8 Å), (**b**) *Tm*NusA-NTD (orange, PDB ID: 1HH2, r.m.s.d. 1.9 Å) (**c**) *Tm*NusA-NTD (grey, PDB ID: 1L2F, r.m.s.d. 1.7 Å), (**d**) *Mt*NusA-NTD (violet, PDB ID: 1K0R, r.m.s.d. 1.8 Å), and (**e**) *Pl*NusA-NTD (yellow, PDB ID: 4MTN, r.m.s.d. 1.4 Å). The linker helix was not used for the superpositions. **(f)** Superposition of NusA-NTD structures shown in (**a-e**). The linker helix is shown in bright colors. **(g)** NOE network fixing the position of the linker helix in NusA-NTD$^{\Delta}$ (cartoon representation, grey; the linker helix is highlighted in purple). The inset indicates how the molecule is rotated in respect to (**a**). Residues participating in the NOE network are labeled and shown as sticks (carbon atoms, dark grey; nitrogen atoms, blue; oxygen atoms, red; hydrogen atoms, white). Unambiguously identified NOEs are shown as black lines. For clarity only one NOE is displayed per methyl group (using the corresponding methyl carbon atom as center).

Supplementary References

1. Burmann, B. M. *et al*. A NusE:NusG Complex Links Transcription and Translation. *Science* **328**, 501-504 (2010).

9 Danksagung

An dieser Stelle möchte ich mich bei allen Menschen bedanken, die zum Gelingen dieser Arbeit beigetragen haben. Zuallererst gilt mein Dank Prof. Dr. Paul Rösch für das interessante Projekt mit viel Gestaltungsfreiraum unter hervorragenden Arbeitsbedingungen und für sein Interesse an meiner Arbeit.

Dr. Stefan Knauer danke ich für viel Unterstützung, Interesse und eine hohe Diskussionsbereitschaft. Bei Prof. Dr. Birgitta Wöhrl möchte ich mich für wertvolle Ratschläge zur Planung von Laborarbeiten und bei Kristian Schweimer für seine umfangreiche Unterstützung bei NMR-Fragen bedanken.

Anja Groh danke ich für anregende Diskussionen über alternative Methoden, Geld zu verdienen und außerdem, wie auch Gudrun Wagner und Violaine Zigan, für ihre Unterstützung bei Verwaltungsangelegenheiten. Rainer Hofmann danke ich für schnelle Hilfe bei allen Computerfragen und -problemen. Beim TA-Team möchte ich mich für viel Unterstützung bei Laborangelegenheiten bedanken. Ramona Heißmann danke ich für ihren pragmatischen Umgang mit allen Fragen und Problemen im Labor und für viele in großen Mengen gereinigte Proteine, manchmal auch noch RNase-frei. Uli Persau danke ich für ihre Energie und Fröhlichkeit, die mich immer wieder aufgebaut haben, und ihre molekularbiologische Unterstützung. Andy Hager danke ich für ihren herzlichen Einsatz bei Laborganisatorischen Dingen und die schnelle Bearbeitung jedes Materialengpasses.

Bei Anna Schneider, Christian Seutter von Loetzen, Dr. Olivia Hartl-Spiegelhauer, Dr. Maximilian Hartl, Dr. Berit Leo, Dr. Philipp Weiglmeier und Britta Zimmermann bedanke ich für die angenehme Arbeitsatmosphäre, anregende Diskussionen während und nach der Arbeit und für Hilfe und Unterstützung wenn ich sie gebraucht habe. Ganz besonders danke ich Martin Strauß für sehr gute Zusammenarbeit während des RNAP-Projekts, viel Verständnis und eine sehr gute Freundschaft.

Daniel danke ich vor Allem für seine Geduld.

10 Erklärung

(§ 5 Nr. 4 PromO)
Hiermit erkläre ich, dass keine Tatsachen vorliegen, die mich nach den gesetzlichen Bestimmungen über die Führung akademischer Grade zur Führung eines Doktorgrades unwürdig erscheinen lassen.

(§ 8 S. 2 Nr. 5 PromO)
Hiermit erkläre ich mich damit einverstanden, dass die elektronische Fassung meiner Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung hinsichtlich der eigenständigen Anfertigung der Dissertation unterzogen werden kann.

(§ 8 S. 2 Nr. 7 PromO)
Hiermit erkläre ich eidesstattlich, dass ich die Dissertation selbständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe. Ich habe die Dissertation nicht bereits zur Erlangung eines akademischen Grades anderweitig eingereicht und habe auch nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden.

(§ 8 S. 2 Nr. 9 PromO)
Hiermit erkläre ich, dass ich keine Hilfe von gewerbliche Promotionsberatern bzw. -vermittlern in Anspruch genommen habe und auch künftig nicht nehmen werde.

…….
Ort, Datum, Unterschrift