

ABSTRACT

Title of Dissertation: INVESTIGATION OF THE NOVEL
ANTITERMINATION MECHANISM OF
LOAP, A NUSG SPECIALIZED PARALOG

Madison Danielle Jermain, Doctor of
Philosophy, 2025

Dissertation directed by: Professor Wade C. Winkler, Department of
Cell Biology and Molecular Genetics

The transcription factor NusG plays an important role in the transcription elongation complex across all domains of life. During elongation, the NusG-NTD forms contacts to RNAP while the NusG-CTD dynamically interacts with other proteins, affecting processivity, transcription termination, and in some bacteria, transcription-translation coupling. NusG is integral to antitermination complexes, where it associates with additional factors to modify the TEC in order to bypass termination events at distinct gene clusters. In Gammaproteobacteria, these complexes enhance the expression of ribosomal RNA operons, bacteriophage operons, and CRISPR arrays. Specialized NusG paralogs, such as LoaP, RfaH, or UpxY, are widespread in bacteria and likely form their own unique antitermination complexes. RfaH, the only well-studied paralog, binds to a paused TEC via a specific non-template DNA sequence, then prevents

Rho termination while enhancing transcription-translation coupling of targeted operons. It is unclear whether other NusG paralogs use similar mechanisms.

LoaP is primarily encoded by Bacillota, Actinomycetota and Spirochaetota, and while regulons influenced by LoaP have only been identified in a few organisms, the recurring location of *loaP* adjacent to long operons responsible for the biosynthesis of secondary metabolites suggests a regulatory relationship. Previously, *Bacillus velezensis* LoaP was shown to enhance transcription of the difficidin and macrolactin antibiotic synthesis operons, but its antitermination determinants were unknown. Presented herein is the reconstitution of difficidin operon (*dfn*) antitermination activity in vitro following the purification of RNA polymerase and additional transcription factors encoded by *Bacilli*. The *dfn* 5' leader region contains a small RNA hairpin ligand essential for LoaP antitermination activity both in vivo and in vitro, and an intrinsic terminator that is highly dependent on NusA. The addition of LoaP specifically antagonizes NusA, thereby promoting readthrough of the termination region without the need for additional factors. These findings demonstrate the basic requirements of the LoaP class of NusG paralogs and indicate that the LoaP antitermination mechanism differs significantly from that of the RfaH – and more broadly, the Gram-positive – paradigm.

INVESTIGATION OF THE NOVEL ANTITERMINATION MECHANISM OF LOAP, A
NUSG SPECIALIZED PARALOG

by

Madison Danielle Jermain

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2025

Advisory Committee:

Dr. Wade Winkler, Chair

Dr. Jason Kahn, Dean's Representative

Dr. Brantley Hall

Dr. Najib El-Sayed

Dr. Jiqiang Ling

© Copyright by
Madison Danielle Jermain
2025

Dedication

To my parents, Tom and Cyndy Jermain, who instilled in me a vast love for learning. Your unwavering encouragement and faith in me have made me both the scientist and woman I am today. Thank you for always believing in the little girl who called her bedroom her “lab” and who never stopped asking “why?”.

Acknowledgements

Thank you to my advisor, Dr. Wade Winkler, for being ever supportive and for encouraging my scientific exploration throughout every stage of this dissertation. Thank you for allowing me the freedom to develop my skillset not only as an independent researcher, but also a scientific communicator and mentor, and for a plethora of conversations about dogs, books, running and baking.

To the members of the Winkler lab, thank you for the close friendships we have built in lieu of failed experiments and major life events. Thank you for the time you have dedicated to discussing projects and supplying comments on papers and presentations. Your support and collaboration mean the world to me.

I am also sincerely thankful to the members of my advisory committee, for your generous commitment of time to insightful conversations and constructive criticism. Your expertise and feedback have strengthened both this dissertation work and my abilities as a scientific scholar.

A special thanks to my undergraduate mentor, Dr. Michael Carter, whose microbiology lectures and undergraduate research lab immediately derailed my plans to be a forensic scientist. I never expected to love bacterial metabolism so much, but I blame you for it.

Finally, to Tanner, who has spent countless hours rambling about bacterial RNA metabolism with me over the years, first as a fellow grad student and then as my partner. You quickly became my favorite person to talk about science with and so much of this dissertation would not have been possible without your love and support.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
List of Abbreviations	x
Chapter 1: Introduction	1
Post-Initiation Transcriptional Regulation in Bacteria	1
Roles of Universal Transcription Factors	1
Modes of Bacterial Transcription Termination	6
Transcription Attenuation & Processive Antitermination	12
Processive Antitermination Mechanisms	15
Phage λ N Antitermination	15
Ribosomal RNA Antitermination	18
Antitermination via <i>cis</i> -acting RNA elements	20
Specialized NusG Paralogs	23
Antitermination by a NusG Specialized Paralog, RfaH	23
Others (UpxY, TaA, ActX)	26
LoaP	27
Chapter 2: Structure of LoaP bound to RNA hairpin target	31
Copyright Notice	31
Introduction	31
Results	33
LoaP- <i>dfn</i> Crystal Structure	33
Binding measurements of RNA ligands containing substitutions of key functional groups	37
Discussion	43
Materials and Methods	45
Crytallography	45
Differential Radial Capillary Action of Ligand Assay (DRaCALA)	45
Chapter 3: Reconstitution of a <i>Bacillus subtilis</i> elongation complex in vitro	47
Introduction	47
Results	50
Overexpression of His-tagged Transcription Factors	50
Precipitation techniques enhance recovery of transcription factors	53
Nickel Affinity Chromatography	56
Ion Exchange Chromatography	57
Twin-Strep-tagged <i>rpoC</i> increases purity of active RNAP _{BS}	58
Discussion	62
Materials and Methods	65
Strain Construction	65
NusA _{Bv} purification protocol	65
NusB _{BS/Bv} purification protocol	67
NusG _{Ec/BS/Bv} purification protocol	68

S10 (NusE) <i>Ec/Bs/Bv</i> purification protocol	69
Rho <i>Bv</i> purification protocol	71
Twin-Strep-tagged RNAP <i>Bs</i> purification protocol	72
Synchronized In Vitro Transcription	73
Chapter 4: <i>Bacillus velezensis</i> LoaP promotes antitermination of antibiotic synthesis operons by antagonizing NusA.....	74
Introduction.....	74
Results.....	79
LoaP is selectively responsible for production of polyketide secondary metabolites in <i>Bacillus velezensis</i>	79
In vitro transcription of the <i>dfn</i> leader region	84
NusA enhances termination efficiency of the <i>dfnA</i> intrinsic terminator.....	85
LoaP antagonizes NusA in the <i>dfn</i> leader region.....	89
LoaP antitermination activity is dependent on an RNA hairpin located within the <i>dfn</i> leader region	90
A transcriptional pause within the <i>dfn</i> leader region moderately affects LoaP antitermination activity.....	93
Recruitment of LoaP to the TEC is novel	97
LoaP enacts readthrough at additional <i>dfn</i> terminators.....	99
Discussion	102
Materials and Methods.....	108
Strain Construction	108
Protein Expression and Purification.....	109
Synchronized In Vitro Transcription	109
Differential Radial Capillary Action of Ligand Assays (DRaCALAs)	110
Fluorescent Reporters	111
Mass Spectrometry Analysis.....	112
Chapter 5: Future Directions.....	116
Introduction.....	116
Conserved vs. enriched residues in LoaP proteins.....	119
Heterologous express of LoaP with potential for drug discovery.....	122
rRNA Antitermination in <i>B. subtilis</i>	125
Final Thoughts	128
Appendices.....	130
References.....	138

List of Tables

Table 2-1: Summarized interaction sites in the LoaP-*dfn* complex

Table 2-2: Dissociation constants of LoaP mutant-*dfn* interactions

List of Figures

Figure 1-1: Roles of the universal transcription factor NusG differ across organisms

Figure 1-2: NusG bridges interaction between RNAP and the leading ribosome or Rho

Figure 1-3: Modes of Bacterial Transcription Termination

Figure 1-4: Transcription Attenuation and Processive Antitermination exhibit regulatory control over termination events

Figure 1-5: Antitermination by Lambda (λ) Phage

Figure 1-6: Antitermination of *E. coli* rRNA operons

Figure 1-7: Sequence and secondary structure of RNA elements involved in antitermination

Figure 1-8: The NusG specialized paralog RfaH undergoes a conformational change following recruitment to the ntDNA

Figure 1-9: Arrangement of the diffcadin operon in *B. velezensis*

Figure 2-1: Structure of the LoaP-*dfn* hairpin complex

Figure 2-2: LoaP interactions with a bulged pyrimidine and the UNCG tetraloop are critical for RNA-binding

Figure 2-3: Predicted model of LoaP bound to the TEC

Figure 2-4: Structural comparison of NusG specialized paralogs

Figure 3-1: The yield of overexpressed NusA_{Bv} benefits from a shorter induction at a higher temperature

Figure 3-2: Precipitation techniques enhance the recovery of NusA_{Bv}

Figure 3-3: SDS-PAGE gels resolving the purification steps for 10xHis-SUMO-NusA_{Bv}

Figure 3-4: Purification of a Twin-Strep-tagged *B. subtilis* RNAP

Figure 3-5: Twin-Strep-tagged *B. subtilis* RNAP possesses transcription activity in vitro

Figure 4-1: Polyketide antibiotics produced by *B. velezensis*

Figure 4-2: LoaP is selectively responsible for the production of polyketide secondary metabolites in *Bacillus velezensis*

Figure 4-3: Abundance of peptides unique to LoaP or proteins encoded by the *dfn*, *mln* and *bae* operons

Figure 4-4: Synchronized transcription of the *dfn* 5' UTR reveals a RNAP_{Bs} pause site

Figure 4-5: NusA enhances termination efficiency of the *dfnA* intrinsic terminator

Figure 4-6: LoaP antagonizes NusA-mediated intrinsic termination in vitro

Figure 4-7: LoaP antitermination activity is dependent on an RNA hairpin located within the *dfnA* leader region

Figure 4-8: A pause in the *dfn* leader region is specific to RNAP_{Bs}

Figure 4-9: A transcriptional pause within the *dfnA* leader region moderately affects antitermination activity

Figure 4-10: Deletion of the *dfn* leader pause sequence in vitro

Figure 4-11: Investigation of LoaP recruitment to the TEC

Figure 4-12: LoaP enacts readthrough at additional *dfn* terminators

Figure 4-13: LoaP antitermination at the *dfnI* terminator requires the *dfn* hairpin

Figure 4-14: The molecular mechanism for LoaP antitermination is novel

Figure 5-1: Residues enriched in the NGN domain of LoaP_{Bv} have implications for readthrough activity

Figure 5-2: Heterologous overexpression of LoaP proteins does not impact cellular fitness

Figure 5-3: LoaP proteins from other species are structurally similar but may differ mechanistically

Figure 5-4: *B. subtilis* does not show evidence of antitermination for the expression of *rrn* operons in a Rho-dependent reporter system

List of Abbreviations

μM: micromolar

10xHis: 10 sequential histidine residues

2xYT: 2x yeast extract tryptone media

6xHis: 6 sequential histidine residues

AS: ammonium sulfate

AT: antitermination

ATP: adenosine triphosphate

Cryo-EM: cryogenic electron microscopy

CTD: C-terminal domain

DNA: deoxyribonucleic acid

dsDNA: double-stranded deoxyribonucleic acid

DTT: 1,4-dithiothrietol

EAR: *eps*-associated RNA

eps: exopolysaccharides

HGT: horizontal gene transfer

IMAC: immobilized metal affinity chromatography

IPTG: isopropyl β-D-1-thiogalactopyranoside

K_d : dissociation constant

KOW: Krypides-Ouzounis-Woese motif

LB: lysogeny broth

LC-MS/MS: liquid chromatography-mass spectrometry

LoaP: long operon-associated protein

mM: millimolar

mRNA: messenger ribonucleic acid

NBO: non-bridging oxygen

NGN: NusG N-terminal domain

nM: nanomolar

nm: nanometer

NTD: N-terminal domain

ntDNA: non-template deoxyribonucleic acid

NTP: nucleotide triphosphate

Nus: N-utilization substance

NusG^{SP}: NusG specialized paralog

nut: N-utilization site

OD_{600nm}: optical density at 600 nm

ops: operon polarity suppressor

PA: processive antitermination

PAGE: polyacrylamide gel electrophoresis

PCR: polymerase chain reaction

PEI: polyethyleneimine

pI: isoelectric point

PKS: polyketide synthesis

pM: picomolar

PMSF: phenylmethane sulfonyl fluoride or phenylmethylsulfonyl fluoride

RNA: ribonucleic acid

RNAP: RNA polymerase

put: polymerase utilization RNA

rNTP: ribonucleotide triphosphate

rRNA: ribosomal ribonucleic acid

rut: Rho-utilization site

SDS: sodium dodecyl sulfate

SUMO: small ubiquitin-related modifier

TEC: transcription elongation complex

tRNA: transfer ribonucleic acid

Chapter 1: Introduction

Post-Initiation Transcriptional Regulation in Bacteria

Roles of Universal Transcription Factors

The ability to rapidly respond and adapt to the surrounding environment is a hallmark characteristic of microbial life. Bacterial species have evolved a wealth of mechanisms to promptly regulate the expression of their genes in response to fluctuating metabolic and environmental factors, in an effort to maintain or improve cellular fitness. Gene expression, or the process by which genetic information is converted to functional cellular factors, involves the transcription of genes to produce RNA, and often the translation of messenger RNA (mRNA) molecules to produce functional proteins. While gene expression can be regulated at any stage, the transcription complex is a typical regulatory checkpoint. Throughout the process of transcription, RNA polymerase (RNAP) – a relatively stable complex comprised of α_2 , β , β' and ω subunits¹ – is responsible for synthesizing RNA via phases of initiation^{2,3}, elongation^{4,5} and termination⁶. Although this core RNAP complex, or apoenzyme, is catalytically active, it is incapable of enacting transcription alone².

Instead, an additional class of proteins, known as sigma (σ) factors, are required in order to target the otherwise non-specific DNA-binding activity of RNAP to specific promoter sites^{7,8}. Sigma factors are highly dynamic proteins essential to the formation of transcription initiation complexes. These proteins operate by binding a conserved location on the surface of RNAP while additional domains interact with characteristic sequences of promoter DNA elements⁸. This works to effectively position RNAP at precise transcription start sites¹. Additionally, sigma factors promote transcription initiation by helping to unwind the double-stranded DNA (dsDNA)

template at the transcription start, ensuring the proper positioning of template DNA within the active site of RNAP, only when bound to the correct promoter sequence⁸⁻¹⁰.

While all bacteria encode for a housekeeping sigma factor responsible for the expression of a majority of genes – many of which are essential – most bacteria encode for additional, alternative sigma factors. When expressed, these alternative sigma factors redirect the pool of free RNAP to alternative promoter sites². This allows the modulation of bacterial gene expression by temporally influencing which promoter sites RNAP can bind, leading to direct changes in transcript production.

Aside from the association of sigma factors, additional factors make transcription initiation a key regulatory step. These DNA-binding proteins control the rate of transcription initiation at target genes by binding at or near promoter regions to subsequently enhance or suppress RNAP binding^{11,12}. In turn, these actions serve to either activate or repress gene expression at specific locales, giving these transcription modulators the apt classifications of transcription “activators” or “repressors”.

Following initiation, transcription elongation relies on additional regulatory mechanisms wherein each nucleotide incorporation cycle exists as a potential checkpoint, with RNAP depending on pausing events, halt sites and occasionally, additional cellular factors to coordinate the successful production of a functional RNA molecule^{4,13}. During the transition between transcription initiation and elongation, sigma factors dissociate from the surface of RNAP and are replaced by the universal transcription factor NusG¹⁴ (**Figure 1-1A**), which associates with RNAP at the same location and likely facilitates promoter escape. NusG (Spt5 in eukaryotes) is notably the only universally conserved family of transcription factors¹⁵. While NusG is present in all species which utilize a bacteria-type RNAP, its ancient, and nearly always essential, function

is further evidenced by its presence in most transcription elongation complexes (TECs) across the genome¹⁶.

Although the regulatory effects of the housekeeping, core NusG can differ across diverse bacteria, all NusG proteins possess a similar structural core. The highly conserved NusG N-terminal domain (NGN; **Figure 1-1A**) is entirely unique to the NusG family of proteins, with all NGN domains forming similar contacts with RNAP^{15,17}. More specifically, the NGN domain is responsible for interacting with the two largest subunits of RNAP (β and β') at the clamp helices and gate loop, flexible modules which close over the DNA binding channel during elongation¹⁸. In doing so, NusG stabilizes the closed conformation of the RNAP clamp during elongation, enhancing the processivity of the enzyme while discouraging RNAP pausing or premature release of the nascent RNA, an outcome that is irreversible^{19,20}.

Transcription pausing is an essential function across all domains of life^{21,22}. In addition to tuning the rate of nascent RNA synthesis, pausing events can kinetically favor the recruitment of regulatory factors, allow for co-transcriptional RNA folding, which promotes proper RNA secondary structure, contribute to the coupling of transcription and translation machinery, and are an obligatory step in transcription termination^{23,24}. Elemental pauses, or pausing events which occur when RNAP encounters consensus pause sequences, can be brief and are quickly overcome. These pausing events often loosen or reduce contacts between the RNAP clamp and the DNA template, reducing or preventing nucleotide addition²¹. The lifetimes of elemental pauses are heavily influenced by RNA secondary structure (hairpin-stabilized pausing) in addition to transcription elongation factors such as the exit channel-binding protein, NusA^{25,26}.

However, pausing events can also precede TEC arrest via backtracking (backtracked pausing). Backtracking often occurs as a proofreading mechanism in response to the

misincorporation of nucleotides, or in response to obstacles such as DNA-binding proteins or a weak DNA/RNA duplex that causes the TEC to revert back to a region of greater stability²⁷. Such events often trigger TEC pausing and arrest, resulting in an inactive complex as the nascent RNA is extruded through the RNAP active site, blocking nucleotide addition²⁸. While backtracking can be resolved by forward translocation, intrinsic transcript cleavage by RNAP or factor-assisted transcript cleavage^{29,30}, NusG combats backtracking altogether via NGN domain interactions with the upstream non-template DNA (ntDNA), occasionally in a sequence-specific manner^{31,32}. These contacts help to stabilize the upstream DNA duplex at the edge of the 12-14 nucleotide (nt) transcription bubble, inhibiting backtracking.

The NusG-CTD (**Figure 1-1A**) contains a Kyrpides-Ouzounis-Woese (KOW) motif common among RNA-binding proteins^{17,33}. Within the NusG-CTD, this motif serves as a protein binding platform, allowing NusG to complex with additional factors, namely Rho³⁴ (**Figure 1-2B**) or ribosomal protein S10 (NusE)³⁵ (**Figure 1-2A**), which allow NusG to govern the fate of nascent transcripts either by favoring transcript termination or transcription-translation coupling (discussed further below). In this way, accessory factors, operating within multicomponent nucleoprotein complexes, modulate elongation by regulating TEC rate and stability³⁶⁻³⁹. By controlling pause entry and lifetime, as well as the formation of RNA secondary structure, elongation factors either drive RNA synthesis away from or toward termination.

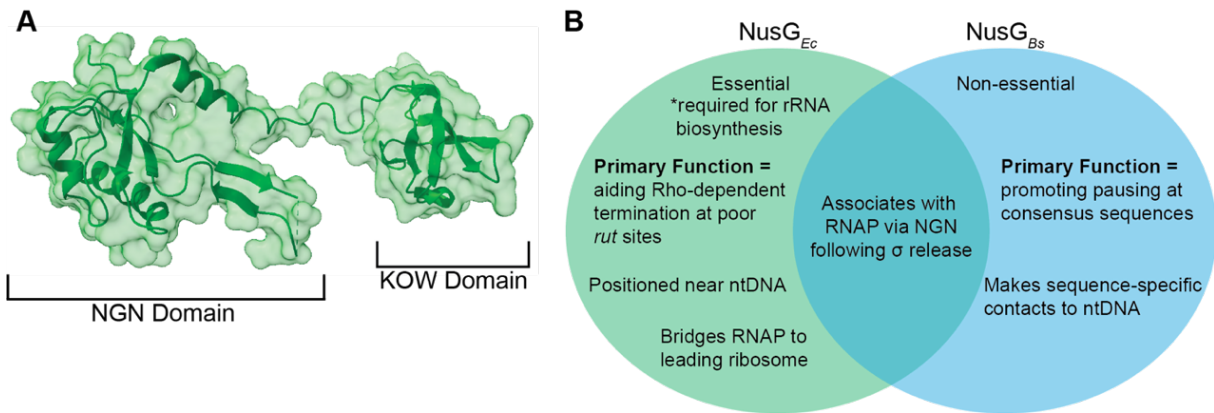


Figure 1-1. Roles of the universal transcription factor NusG differ across organisms (A) A structural model of *E. coli* NusG (PDB: 6GOV) displaying the characteristic NusG N-terminal domain (NGN), which interfaces with RNAP, and the Krypides-Ouzounis-Woese (KOW) domain, which is responsible for binding additional transcription factors – Rho or ribosomal protein S10 (NusE). **(B)** Venn diagram contrasting the functions of bacterial NusG in either *E. coli* (NusG_{Ec}) or *B. subtilis* (NusG_{Bs}), the representative model systems for Gram-negative and Gram-positive organisms, respectively.

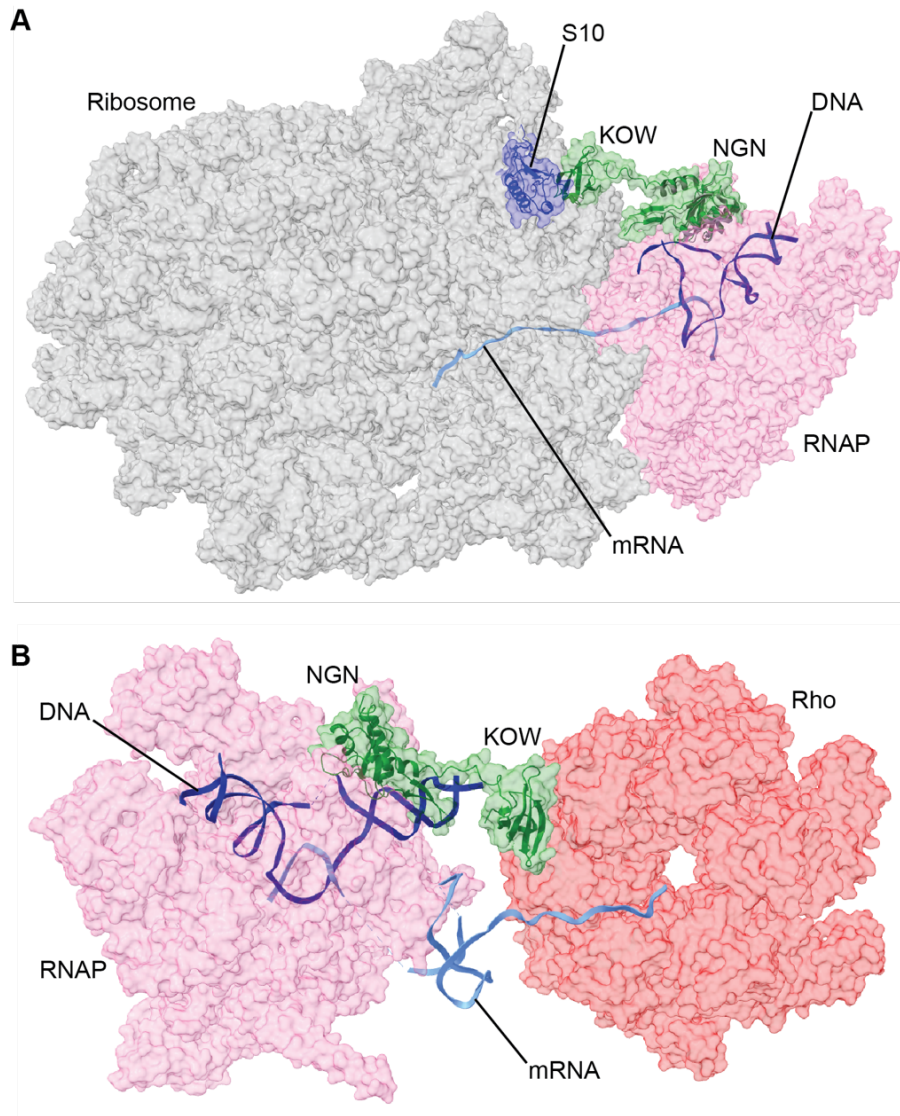


Figure 1-2. NusG bridges interaction between RNAP and the leading ribosome, or Rho Structural models of NusG (green, NGN and KOW domains indicated; PDB: 6GOV) bridging RNAP (pink; PDB: 6VU3) – via NGN-RNAP interactions – with (A) the leading ribosome (grey; PDB: 6VU3) – via KOW-S10 (blue) interactions – or (B) Rho (red; PDB: 6DUQ) – via KOW-Rho interactions. DNA and mRNA molecules are displayed as dark blue and light blue ribbons, respectively.

Modes of Bacterial Transcription Termination

Transcription termination works to destabilize the highly stable TEC at specific sites in the genome, enabling dissociation of the RNAP from the template DNA and nascent RNA

transcript, and is essential to the process of gene expression. Termination events work to (i) release RNAP to be recycled for global gene expression (ii) define the 3' ends of RNAs, either for downstream regulatory purposes or RNA structures, and (iii) prevent the unsuitable expression of downstream genes and antisense transcripts⁶. Bacterial transcription termination occurs via two mechanisms (**Figure 1-3**): intrinsic termination, which occurs based on elements encoded by the DNA template and nascent RNA^{6,23}, or Rho-dependent termination, which relies on Rho, an ATP-dependent helicase that interacts with the TEC and triggers the release of the nascent RNA^{6,40-42}.

Intrinsic (or factor-independent) termination occurs following the formation of a hairpin structure in the nascent RNA. These RNA secondary structures typically comprise a G/C-rich helix followed by a polyuridine tract reminiscent of an elemental pause sequence²³.

Understandably then, the polyuridine sequence causes the TEC to pause, allowing the terminator hairpin structure to form within the RNA exit channel of the RNA polymerase. It is thought that this terminator hairpin serves to prevent backtracking at the weak, (U)-rich DNA/RNA duplex, instead locking the TEC in place over the weak hybrid sequence^{28,29,43}. This sequence loosens protein-nucleic acid interactions within the TEC, triggering an RNAP conformational change, rendering the complex susceptible to dissociation and ultimately, the release of the DNA template and nascent RNA²³. Additionally, intrinsic termination can be enhanced by NusA, an elongation factor that stabilizes interactions between the nascent RNA and RNAP within nearly all TECs in *E. coli*⁴⁴. In doing so, NusA decreases the rate of pause escape, favoring the formation of intrinsic terminator hairpins.

Conversely, Rho requires a less specific sequence to enact termination, which is crucial for its role as a global genome regulator. Instead, Rho associates with the TEC at C-rich, G-poor

sequences across the genome, deemed Rho-utilization (*rut*) sites^{45–49}. Once bound to RNAP, Rho employs helicase activity, which serves to dislodge RNAP and facilitate the dissociation of the TEC from the nascent transcript^{50–56}. As Rho surveils the genome, it plays a critical role in gene silencing, particularly for potentially harmful transcripts such as foreign DNA and antisense transcripts^{42,57,58}. By triggering the premature transcription of aberrant transcripts, Rho ensures the maintenance of cellular resources by reducing transcriptional noise and translational stress, along with reducing collisions between replication and transcription machineries^{6,38}.

Similar to intrinsic termination, Rho-dependent termination is often aided by additional factors, which enhance Rho recruitment and association with the TEC. As previously mentioned, the NusG-CTD, via the KOW motif, makes contacts to Rho. Specifically, NusG interacts with the C-terminal translocase domain of Rho, favoring ring closure and stabilizing the hexameric complex on the nascent RNA³⁴. Although Rho is capable of interfacing directly with RNAP, a requirement for NusG-enhanced Rho binding exists in regions of the genome containing sub-optimal, C-poor sequences, such as antisense transcripts⁵⁹.

Interestingly, NusG also plays a role in preventing termination by Rho. Due to the lack of a physical nuclear membrane in bacteria, the processes of transcription and translation occur in close proximity, and often simultaneous, to one another⁶⁰. In some cases, NusG utilizes the KOW motif in order to bind ribosomal protein S10 (NusE) at the same binding interface employed to bind Rho⁶¹. With this, NusG recruits the pioneering ribosome to the transcription machinery, thus physically coupling the processes of transcription and translation. This transcription-translation coupling not only hinders Rho recruitment due to the mutually exclusive nature of the NusG binding site but also limits access by Rho to the nascent transcript. This

allows NusG, as well as the translating ribosomes, to protect nascent mRNAs from premature, Rho-mediated release.

Transcription-translation coupling harbors additional, important regulatory consequences. For instance, the presence of a leading ribosome at the exit channel of RNAP prevents backtracking events^{62,63} and the formation of R loops⁶⁴, structures formed within the template DNA during transcription which can have detrimental effects on downstream DNA replication. Additionally, coupling promotes TEC processivity past nascent terminator hairpins and shields mRNA transcripts from degradation by RNaseE⁶⁵.

Similar to Rho, ribosomes are capable of interacting with RNAP without additional factors, but recent cryo-EM data suggest the need for additional factors (both NusG and NusA) arises quickly after association of the two machineries⁶⁶. In fact, it has been suggested that these bridging factors serve to prevent collisions between transcription and translation apparatuses^{37,38}. Then, following the dissociation of the leading ribosome, either via stalling or following the recognition of a stop codon, this uncoupling frees RNAP to interface with Rho, resulting in mature transcript release⁶⁷.

As previously stated, NusG proteins from diverse bacteria differ in their regulatory control (**Figure 1-1B**). These differences are most abundantly illustrated when comparing NusG proteins from *E. coli* and *B. subtilis*, the current model systems for Gram-negative and Gram-positive organisms, respectively. RNAP complexes from *E. coli* and *B. subtilis* are only 50% identical⁶⁸; therefore, it stands to reason that the mechanisms employed to regulate each may differ in significant ways. Strikingly, while both *nusG* and *rho* are essential genes in *E. coli*^{42,69}, they are dispensable in *B. subtilis*^{42,70,71}. This can likely be explained by the fact that while *E. coli* relies heavily on Rho for a majority of termination events, Rho has very limited effects in *B.*

subtilis, with intrinsic termination being the more prevalent mechanism instead^{42,44,72}. In *E. coli*, mutations in *nusG* that conferred defects in transcription termination⁷³ or ribosomal interactions³⁵ did not result in a phenotype. These data, along with the limited involvement of Rho in *B. subtilis*, suggest an alternative primary function for NusG, as well as alternative means for global gene silencing.

Additionally, transcription and translation are not thought to be coupled in *B. subtilis*⁷². In fact, RNAP from *B. subtilis* is known to translocate at double the speed of the leading ribosome and pauses less frequently than its *E. coli* counterpart^{72,74}. Therefore, even if NusG were to contact RNAP or the leading ribosome, the two complexes would likely be too far apart for NusG to bridge an interaction. This phenomenon has a noticeable impact on the organization of gene clusters within *B. subtilis*. Unlike in *E. coli*, translational stop codons are often positioned in close proximity to intrinsic terminator regions in *B. subtilis*⁷². Were *B. subtilis* to couple transcription and translation, the position of a leading ribosome would render a majority of the intrinsic termination sites within the genome ineffective. Simply put, this likely suggests that transcription termination events in *B. subtilis* are largely unaffected by translation.

Moreover, *B. subtilis* RNAP exhibits a lower propensity for backtracking³⁹. Instead, NusG is responsible for stimulating TEC pausing at consensus, (T)-rich sequences (TTN_nTTT) required for the successful expression of several operons in *Bacillus*³⁹. This suggests that the primary role of NusG in *B. subtilis* is to operate as a pause-promoting accessory subunit to the TEC, a regulatory mechanism which may be further widespread amongst bacterial species⁷⁵. Compared to its primary function in *E. coli* – assisting Rho activity – the core NusG from *B. subtilis* illustrates the evolutionary pressure for universal transcription factors to tailor their functions in an effort to adapt to the regulatory mechanisms of the cell.

This idea is further exemplified in the diversity exhibited by the C-terminal domains of NusG family proteins. Despite a common evolutionary origin and a conserved binding site on RNAP, the KOW motif within the NusG-CTD confers major mechanistic differences amongst NusG paralogs^{36,76}. Discussed later in this chapter, horizontal gene transfer (HGT) and the need for specialized functions has resulted in a great deal of mechanistic diversity within the NusG protein family. The emergence of NusG paralogs has resulted in specialized transcription regulators, which differ considerably in their regulon targets, interaction networks and overall impact on RNA synthesis, often for the purpose of producing niche-specific macromolecules. These paralogs, which compete with core NusG for space on the TEC⁷⁷, regulate gene expression in both a sequence- and condition-specific manner, and are crucial to the broader topics of bacterial evolution and pathogenesis.

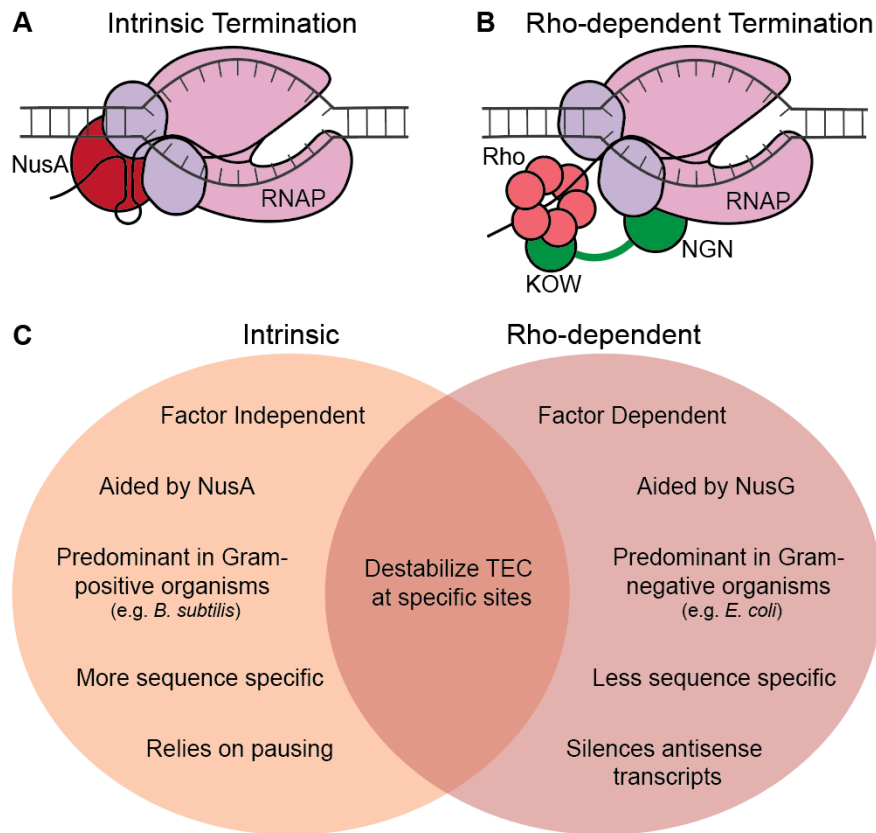


Figure 1-3. Modes of Bacterial Transcription Termination (A) Schematic depiction of intrinsic transcription termination wherein an RNA hairpin structure, often aided by NusA (red), forms within the exit channel of RNAP, forcing the dissociation of the transcription elongation complex. (B) Schematic depiction of Rho-dependent transcription termination, during which Rho helicase, often aided by core NusG (green, NGN and KOW motifs identified), interacts with RNAP and destabilizes the TEC, resulting in dissociation from the nascent RNA. (C) Venn diagram contrasting the methods of transcription termination.

Transcription Attenuation & Processive Antitermination

Transcription termination can be further regulated via two distinct mechanisms: transcription attenuation and processive antitermination, which modulate the efficiency of termination events for the purpose of conditional downstream gene expression. Transcription attenuation mechanisms operate at specific terminator sites and result in the formation of alternative, antiterminator structures that exist in a mutually exclusive manner to termination

sites^{78,79}. Typically, attenuation is observed at *cis*-acting, signal-responsive RNA elements which ablate termination activity at a specific site immediately downstream of the signal-responsive portion. Many classes of *cis*-acting regulatory RNA elements – perhaps most prominently, riboswitches⁸⁰ – have been implicated in transcription attenuation mechanisms controlling cellular response to signals including specific cellular metabolites or metal ions, RNA-binding proteins, environmental temperature fluctuations, uncharged tRNAs and stalled ribosomes^{81–83}. A popularized example of transcription attenuation occurs in *E. coli* in which four RNA elements within the leader region of the *trp* operon control the production of tryptophan in response to current intracellular tryptophan levels⁸⁴ (**Figure 1-4**). More specifically, in a state of high tryptophan availability, translating ribosomes readily occupy the nascent transcript, allowing regions 3 and 4 of the *trp* leader to base pair, forming a terminator hairpin. This results in a shortened or “attenuated” transcript, preventing transcription of the downstream *trp* operon which encodes for the enzymes responsible for the synthesis of tryptophan. When tryptophan levels are low, the scarcity of tryptophan-charged tRNAs causes translating ribosomes to stall at tryptophan codons within region 1. When this happens, region 2 is free to base pair with region 3, forming a hairpin structure (antiterminator hairpin) that prevents formation of a termination hairpin and thereby promotes downstream expression of the *trp* operon and the subsequent biosynthesis of tryptophan.

Unlike attenuation mechanisms, which regulate specific termination events, processive antitermination (PA) mechanisms manipulate active TECs, rendering them broadly resistant to downstream termination signals and less sensitive to pause sites across the length of an entire transcript^{85–87}. Although standard transcription elongation complexes are relatively stable, RNAP is only capable of transcribing roughly 10^4 nucleotides before dissociating from the RNA by

random chance⁶. Given that RNAP is unable to reinitiate transcription along a truncated RNA, the TEC must transcribe entire operons in a single attempt. This means that to express bacterial operons longer than ~10 kilobases (kb), cells may in certain instances enhance the innate processivity of RNAP. Additionally, the expression of many long operons is under strict regulatory control, either via Rho silencing⁴² or the prevalence of intrinsic terminator hairpin sequences⁸⁸, in order to prevent the unnecessary expenditure of cellular resources. To overcome this, bacteria, in accordance with phage genomes, have evolved mechanisms that recruit specialized factors to structurally modify the transcription apparatus in order to bypass impeding termination events⁸⁹. These modifications rely on the concerted effort of multiple RNAP-associated, elongation factors.

In all, bacteria have evolved a complex network of post-initiation regulatory mechanisms that rely on combinations of *cis*-acting signal-responsive RNA elements and accessory proteins (as well as non-coding RNAs and *trans*-encoded small RNAs) in an effort to regulate the transcriptional outcome.

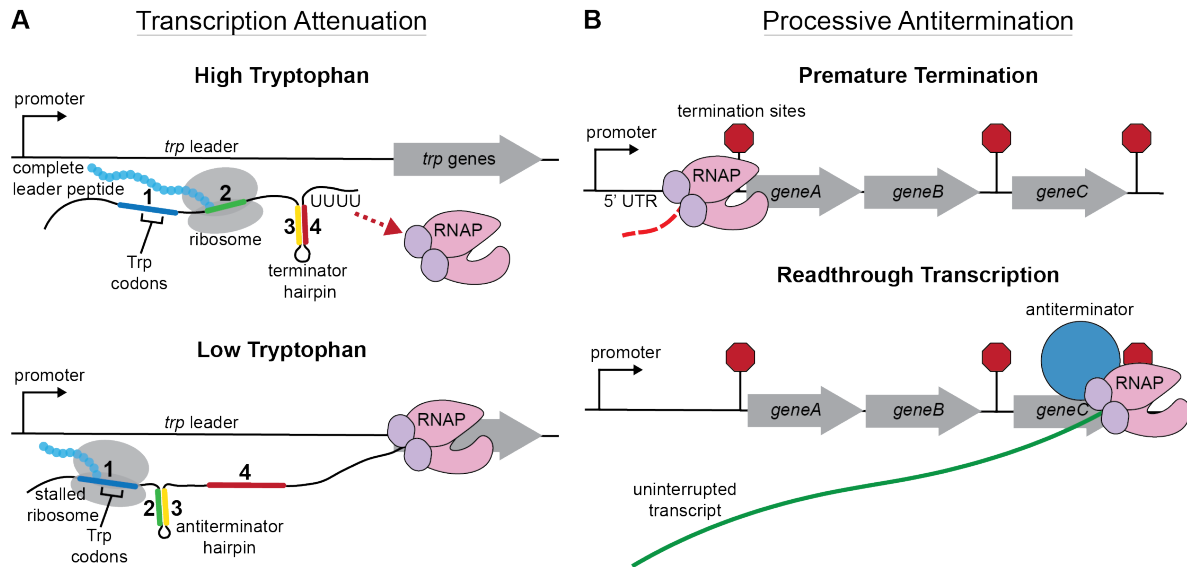


Figure 1-4. Transcription Attenuation and Processive Antitermination exhibit regulatory control over termination events (A) Schematic depiction of transcription attenuation controlling expression of the *E. coli trp* operon. The *trp* leader region possesses four RNA elements. High tryptophan levels (top) allow translation of the leader, resulting in ribosome occupancy of the first two regions and base pairing of regions 3 and 4, forming a terminator hairpin. Low tryptophan levels (bottom), and therefore low levels of tryptophan-charged tRNAs, result in ribosome-stalling at Trp codons at region 1. This allows regions 2 and 3 to base pair, preventing the formation of a terminator hairpin and resulting in the downstream expression of the *trp* operon. **(B)** Schematic depiction of processive antitermination mechanisms. In the absence of an antitermination factor/element (top), termination events trigger the dissociation of the TEC. Following the association of an antiterminator with RNAP (bottom), the TEC is rendered resistant to termination events and downstream expression occurs.

Processive Antitermination Mechanisms

Phage λ N Antitermination

The first processive antitermination mechanism discovered was that of λ phage, a bacteriophage that relies on antitermination during early gene expression⁹⁰. During the lytic cycle, λ phage depends on a protein, λ N, to interact with host factors and promote antitermination of large gene clusters otherwise silenced by Rho^{85,91}. λ N, a small, intrinsically disordered protein, is stabilized via formation of the antitermination complex, which occurs following the synthesis of two RNA elements^{92,93}. The first, *boxB*, is a 15-nucleotide RNA

hairpin sequence resembling a canonical GNRA tetraloop structure and serving as the binding platform for λ N^{92,94-96}. In addition to recruiting λ N, *boxB* also serves as a NusA binding site, producing a λ N:*boxB*:NusA complex that sufficiently prevents the dissociation of the nascent RNA at downstream terminators in vitro^{87,97,98}.

Although capable of antitermination activity, this minimal antitermination complex is bolstered in vivo by the incorporation of additional elongation factors, NusB and NusE (ribosomal protein S10), which are recruited to the second RNA element, *boxA*^{92,98-100}. This short, linear RNA sequence serves to assemble a stable, comprehensive complex of proteins that remains associated with RNAP throughout transcription^{92,97,98}.

Positioned at the RNA exit channel, this ribonucleoprotein complex promotes interactions by λ N to both RNAP and NusA^{92,94,101-103} thereby improving the processivity of λ N antitermination. Within standard transcription elongation complexes, NusA is known to interact with RNAP at the exit tunnel, enhancing intrinsic termination at sites throughout the genome with some terminators considered NusA-dependent^{44,104}. During λ N antitermination, λ N is thought to impede NusA effects on intrinsic hairpin formation by repositioning the RNA-binding C-terminal domains of NusA away from the exit channel^{92,94}. In doing this, λ N essentially reprograms a transcription elongation factor, NusA, into an antitermination factor (**Figure 1-5**).

Additionally, λ N functions to inhibit Rho-dependent termination. During typical transcription elongation, NusG works to recruit Rho to the nascent RNA aiding in Rho processivity and in turn, termination^{57,105}. The λ N antitermination complex restricts Rho activity by promoting the association of the NusG-CTD to additional factors (the NusB:NusE complex) in a way that is mutually exclusive to the binding of Rho⁶¹. In addition to the restricted access to the nascent RNA, it is possible that the enhanced processivity afforded by an antitermination

complex prevents Rho activity in making it more difficult for Rho to establish the necessary interactions with the elongation complex.

Lambdoid phages also encode a second antitermination system, which utilizes a protein, λQ , to promote antitermination of genes expressed later in the phage growth cycle^{86,106}. While similar to λN in function, λQ is mechanistically distinct in that λQ operates during transcription initiation as a DNA-binding protein that interacts with a $\sigma 70$ -associated, paused elongation complex⁸⁶. Without λQ , this elongation complex dissociates at an intrinsic terminator directly downstream of the pause site. λQ , recruited by a λQ -binding element (QBE) within the promoter region, associates with RNAP to render it resistant to downstream termination events¹⁰⁷.

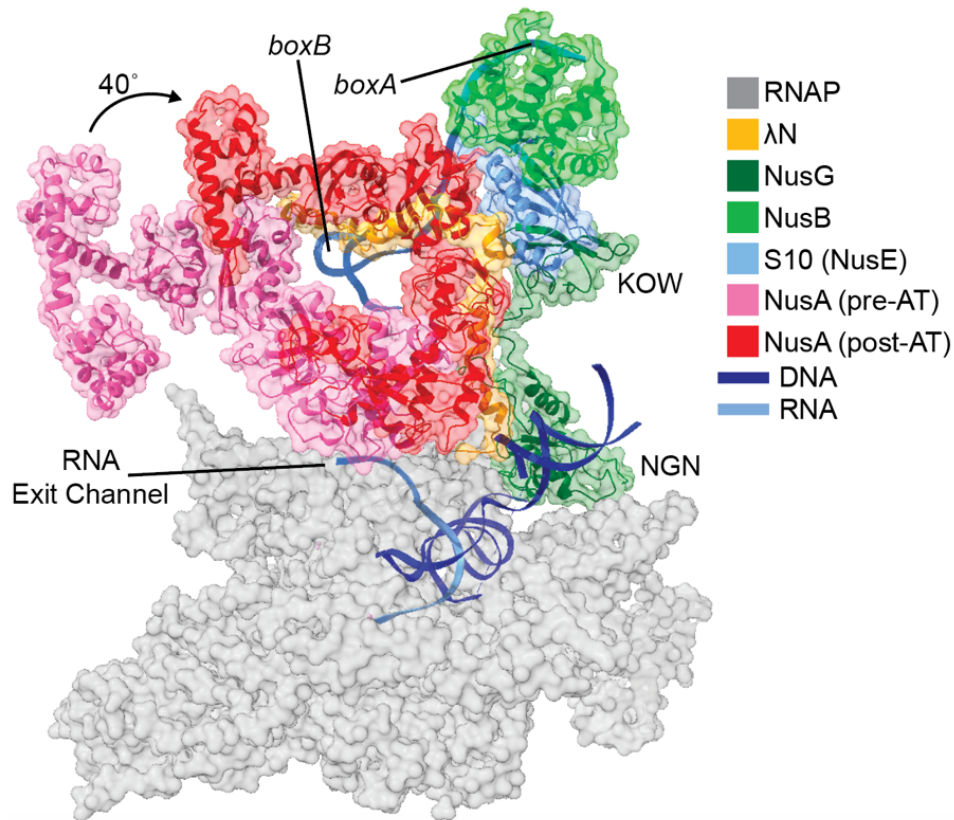


Figure 1-5. Antitermination by Lambda (λ) Phage Structural models generated from cryo-EM data of transcription elongation complexes represent the λ N antitermination complex (PDB: 5MS0) comprised of RNAP (grey), template DNA (dark blue ribbon) and nascent RNA (light blue ribbon), in association with NusG (dark green) and NusA (red). Associations of protein factors to *nut* sequences are shown, wherein NusG interacts with the NusB/S10 heterodimer (light green/light blue) bound to *boxA*, and λ N (yellow) binds the *boxB* hairpin. The λ N antitermination complex effectively repositions NusA roughly 40 degrees away from the RNA exit channel, when overlayed with NusA (pink) in a standard transcription elongation complex (PDB: 6FLQ).

Ribosomal RNA Antitermination

Although initially observed in lambdoid phage operons, additional antitermination mechanisms have since been characterized in bacteria and appear to regulate a diverse range of gene clusters. In bacteria, the expression of long, non-coding RNAs – such as ribosomal RNA (rRNA) operons – are often regulated by premature Rho-dependent transcription termination⁴². Rho, which loads onto the nascent RNA at C-rich *rut* sites, is typically combatted by the

presence of RNA secondary structure or during coupled transcription-translation as the presence of ribosomes prevents Rho loading and translocation^{45,108}. However, given that rRNA operons are not translated, and therefore lack ribosomes, these operons must protect their transcripts from Rho activity via alternative means.

In *E. coli*, the leader regions of all seven rRNA operons contain a highly conserved RNA element resembling the *boxA* sequence employed during λ N antitermination. These rRNA leaders also contain a *boxB*-like hairpin sequence and a third, relatively uncharacterized, linear *boxC* sequence. Despite the presence of three well-conserved RNA elements, only *boxA* appears necessary for rRNA antitermination activity^{98,109}.

Similar to the λ N antitermination complex, *boxA* is responsible for the recruitment of a NusB:NusE heterodimer, which subsequently interacts directly with the NusG-CTD in an effort to prevent Rho recruitment^{98,110}. However, the involvement of host Nus factors alone is not sufficient to promote antitermination activity in vitro. Instead, an additional host protein – an inositol monophosphatase, SuhB – is required to suppress Rho-dependent termination, suggesting an interaction network between SuhB and Nus factors¹¹¹. Specifically, the antitermination activity conferred by the addition of SuhB is dependent on the presence of NusA¹¹², a finding which is further supported by recent cryo-EM data suggesting interactions between SuhB and NusA in vitro¹¹³. These data suggest that the former operates to reposition NusA away from the exit channel in a method analogous to the λ N antitermination mechanism (**Figure 1-6**).

Moreover, some researchers suggest rRNA antitermination activity is required for successful rRNA maturation, and therefore ribosome biogenesis, in vivo¹¹¹. Specifically, it is suggested that Nus factors, in coordination with RNase III, work to promote proper cotranscriptional folding of

16S rRNA, and that SuhB functions to maintain contacts between RNAP and *boxA*-bound Nus factors.

Taken together, the antitermination mechanism present at *E. coli* rRNA operons suggests that λ N antitermination may have arisen as an adaption of host antitermination systems, employing viral proteins to tailor host Nus factors for targeted viral expression.

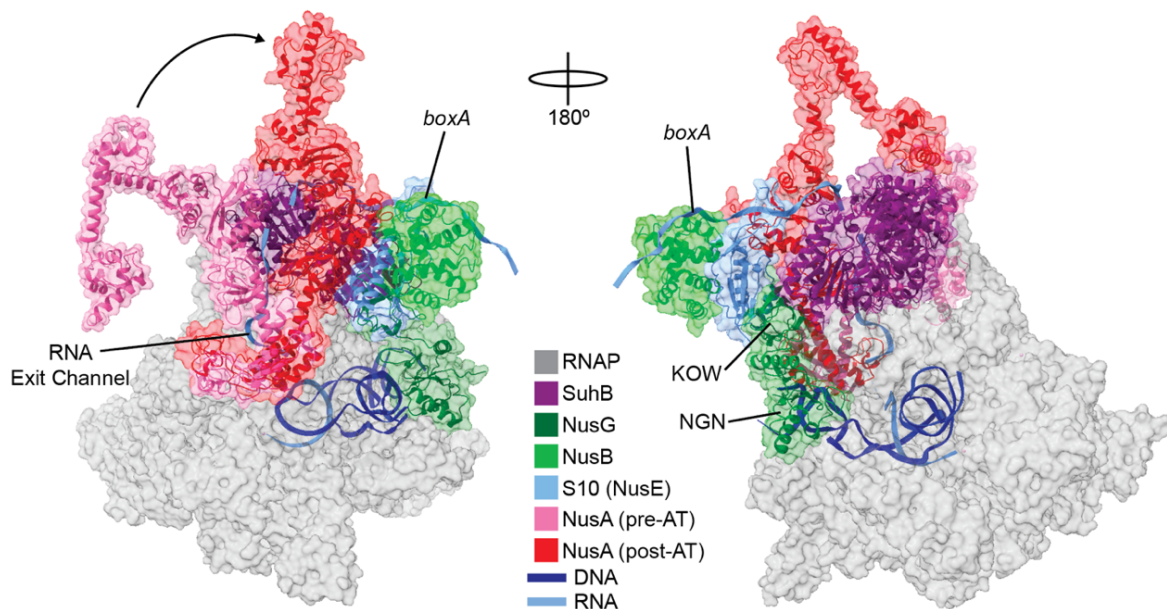


Figure 1-6. Antitermination of *E. coli* rRNA operons Structural models generated from cryo-EM data of transcription elongation complexes represent the antitermination complex responsible for the expression of *E. coli* rRNA operons (PDB: 6TQN). The antitermination complex comprises RNAP (grey), template DNA (dark blue ribbon), and nascent RNA (light blue ribbon), in association with NusG (dark green), NusA (red) and SuhB (purple). A rotated view (right) emphasizes the position of SuhB adjacent to NusA and the association of the NusB/S10 heterodimer (light green/light blue) bound to *boxA*. The rRNA antitermination complex works to reposition NusA away from the RNA exit channel, evidenced by overlaying the rRNA antitermination complex with NusA (pink) in a standard transcription elongation complex (PDB: 6FLQ).

Antitermination via *cis*-acting RNA elements

The discovery of additional processive antitermination mechanisms has revealed a much broader role for RNA elements aside from those that *boxA* and *boxB* play during λ phage and

rRNA antitermination events. For instance, phage HK022 – which encodes for λ Q but not λ N – employs a large, 65-nucleotide RNA element coined *put* (polymerase utilization RNA) for antitermination during early gene expression¹¹⁴ (**Figure 1-7A**). This *put* element, which forms a two-hairpin motif in the leader regions of target operons, modulates transcription elongation via pause suppression, and likely requires no additional factors^{114–116}. Specifically, the secondary structure of *put* is integral to interactions with RNAP, where it remains associated during transcription, allowing *put* to modulate pause signals in the nascent RNA¹¹⁷. Furthermore, phage HK022 possesses a λ N-like protein, Nun, which likely co-evolved with the *put* element. This protein was found to interact with *boxB*, inhibiting RNAP translocation and thereby promoting Nun-dependent transcription arrest^{117–120}. By antagonizing RNAP pausing, *put* decreases the dwell time of the elongating transcription complex, thereby rendering it less susceptible to termination triggered by Nun, Rho or the formation of intrinsic terminators¹²¹. While some details of *put*-mediated antitermination are yet to be elucidated, the discovery of the *put* element expanded the antitermination dogma to include systems driven primarily by RNA elements.

Following the discovery of *put*, an even larger (~125 nucleotides), more structurally complex RNA element was found in *B. subtilis*. Broadly conserved in Bacillales, this RNA element is nearly always associated with gene clusters encoding for the biosynthesis of biofilm or capsule exopolysaccharides (*eps*), and so was coined EAR, for *eps*-associated RNA¹²² (**Figure 1-7B**). Comparative sequence analyses revealed that EAR consists of five helical segments and a characteristic pseudoknot at the 5' terminus, resembling a riboswitch in both size and structural complexity. Deletion of the EAR element, or mutagenesis of conserved residues, resulted in the incomplete transcription of the *B. subtilis eps* operon, instead resulting in transcripts truncated at the sites of intrinsic terminators, which regulate expression of the operon. Additionally,

placement of the EAR element upstream of unrelated intrinsic terminators resulted in downstream expression, demonstrating the ability of EAR to broadly promote antitermination at intrinsic termination sites. While the EAR antitermination mechanism is yet to be reconstituted in vitro, and is thought to require additional, unidentified cellular factor(s), the mechanism implemented by EAR is likely distinct from those utilized during λ N and rRNA antitermination. Both λ N and rRNA systems work to counteract Rho involvement along the nascent RNA, while EAR promotes readthrough at intrinsic terminators. The discovery of EAR illustrates the potential for structurally complex RNAs to participate in antitermination mechanisms and suggests a broader mechanistic diversity for antitermination systems.

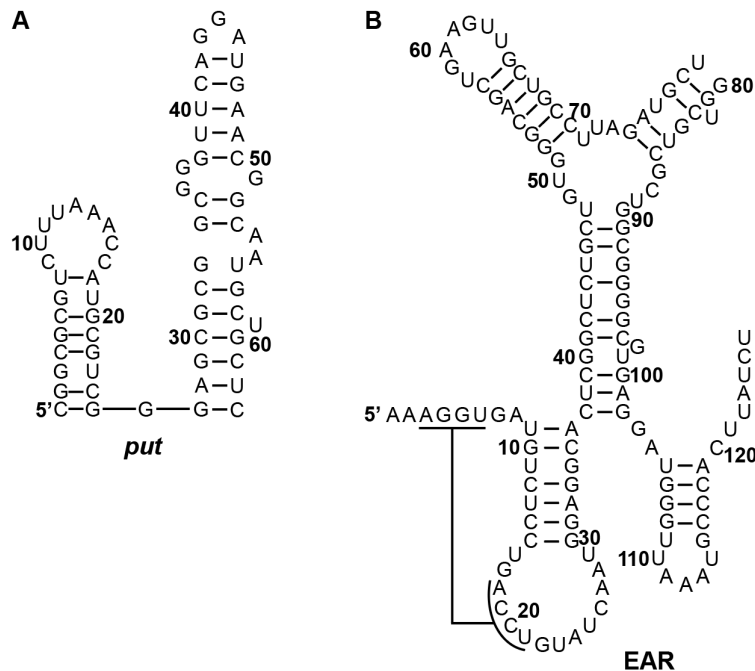


Figure 1-7. Sequence and secondary structure of RNA elements involved in antitermination (A) Dual-hairpin structure of the *put* RNA element involved in HK022 phage antitermination. **(B)** Complex EAR element responsible for antitermination of the *B. subtilis eps* gene cluster. EAR, a G/C-rich RNA element, comprises five conserved helical segments and a characteristic pseudoknot.

Specialized NusG Paralogs

Antitermination by a NusG Specialized Paralog, RfaH

NusG, a universal transcription elongation factor, plays key roles in both λ phage and rRNA antitermination mechanisms. While nearly all bacteria encode for a core NusG, many species also encode additional NusG specialized paralogs (NusG^{SP}) for particular functions^{15,89}. One such specialized paralog, RfaH, was discovered in *E. coli* and is found throughout Gammaproteobacteria¹²³. RfaH, encoded by an essential gene in *E. coli*, is responsible for promoting the expression of four separate virulence-related gene clusters, including genes for outer membrane functions and lipopolysaccharide biosynthesis^{124–127}.

As a NusG paralog, RfaH possesses a conserved, two-domain structure wherein the N-terminal domain constitutes an NGN motif unique to the NusG family of proteins. This domain allows RfaH to interact with RNAP at the same site typically occupied by NusG but with greater affinity, allowing RfaH to outcompete the more abundant (~100:1) NusG for space on the TEC^{76,128,129}. The RfaH-CTD contains a KOW motif similar to NusG, which confers binding to ribosomal protein S10 (NusE) but is incapable of binding Rho¹²³. These key differences permit RfaH to replace core NusG within the TEC, leading to suppression of Rho activity across key regions of the genome.

Given that RfaH and NusG share a high degree of overall sequence and structural similarity^{17,130,131}, it is understandable that they have similar impacts on transcription elongation. Similar to NusG, RfaH suppresses pausing in vitro, likely through its ability to associate with RNAP and prevent the β' subunit from adopting a “swiveled” conformation receptive of both pausing and termination¹⁸. Additionally, both NusG and RfaH form direct contacts to the non-template DNA, further stabilizing the elongation complex and promoting processivity^{18,32,132–134}.

In fact, the ability of RfaH to bind DNA allows for the specific recruitment of RfaH to target operons. Within the untranslated leader regions of RfaH-regulated operons is an 8 bp hairpin element in the non-template DNA strand deemed the “operon polarity suppressor”, or *ops*^{76,135}. Either depletion of RfaH or deletion of *ops* resulted in a significant reduction of downstream gene expression. On the other hand, introduction of *ops* elsewhere in the genome increased downstream expression^{126,127}. Taken together, these data suggest both RfaH and *ops* are necessary for antitermination activity. In fact, RfaH relies on a transcriptional pause which occurs at *ops*, positioning the DNA hairpin within the transcription bubble. This pause favors direct, specific contacts by the RfaH-NTD to solvent exposed bases in the non-template *ops* DNA^{18,129,135}.

Notably, free RfaH occurs in an autoinhibited state in which the RNAP-binding site of the RfaH-NTD is occluded by the KOW motif, which maintains an α -helical conformation in complex with the NTD¹³⁶ (**Figure 1-8A**). Following successful recruitment at *ops*, RfaH transforms into an active state in which the RfaH-NTD, in complex with *ops*, is free to associate with the elongation complex with roughly 1,000-fold higher affinity than free RfaH¹⁸. With this, *ops* ensures RfaH acts only at its target operons rather than interfering with global, core NusG function, while also compensating for the low cellular abundance of RfaH.

Recruitment to *ops* and the subsequent association with the TEC also triggers refolding of the RfaH-CTD into the β -barrel motif shared by NusG¹²³ (**Figure 1-8B**). This active form of the CTD, which grants enhanced binding to ribosomal protein S10 (NusE), allows RfaH to recruit ribosomes to nascent transcripts, bridging transcription and translation processes. This coupling of the transcription apparatus to the leading ribosome not only promotes antitermination by

obstructing RNAP backtracking and Rho access, but aids in the translation of genes controlled by RfaH, which are known to possess both poor ribosome binding sites and rare codons^{76,137}.

In summation, RfaH behaves as a specialized elongation factor, stimulating the expression of select operons typically silenced by core NusG⁷⁶, by promoting the uninterrupted synthesis of target mRNAs and enhancing the subsequent translation of associated genes. RfaH, the NusG paralog that has been studied in greatest molecular detail, expands predictions regarding the cellular functions of NusG proteins and suggests that further investigations into other classes of NusG paralogs is crucial to understanding the full extent by which transcription elongation is regulated in bacteria.

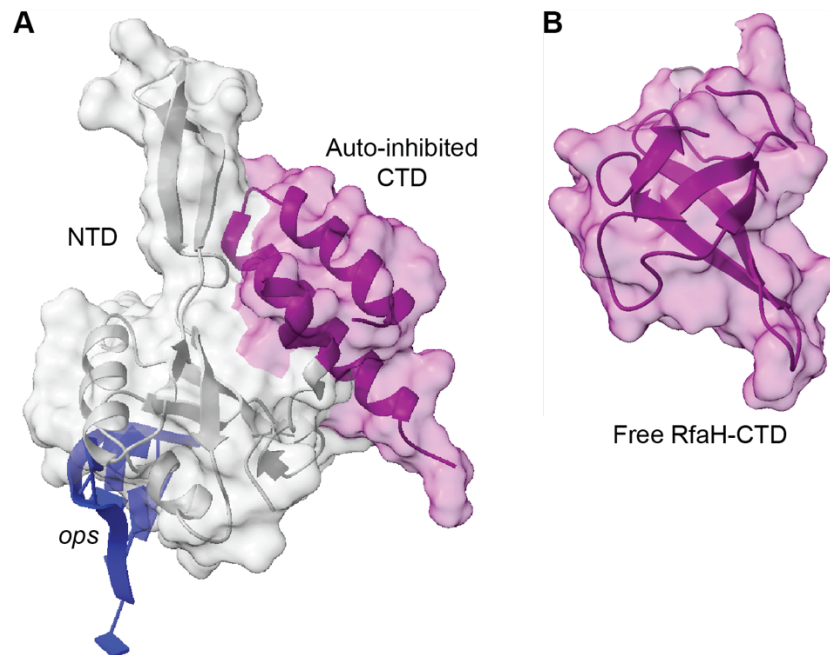


Figure 1-8. The NusG specialized paralog RfaH undergoes a conformational change following recruitment to the ntDNA (A) Structural model of RfaH in an autoinhibited state. The RfaH CTD (purple) maintains an α -helical conformation in complex with the NTD (grey). This autoinhibited protein is recruited by *ops* (blue) on the non-template DNA (PDB: 2OUG). **(B)** Structural model of the active RfaH CTD. Following recruitment to *ops* and association with RNAP, the RfaH CTD undergoes a conformational change and is released from the NTD. The CTD of activated RfaH takes on the β barrel structure typical of NusG KOW domains (PDB: 2LCL).

Others (Up_xY, TaA, ActX)

Aside from RfaH, many other classes of NusG paralogs have been identified and are presumed to also function in processive antitermination. These paralogs, sharing considerable sequence similarity to core NusG – and therefore RfaH – consequently share conserved structural motifs¹³⁸. Unsurprisingly, many of these NusG paralogs have been implicated in the expression of target transcripts, suggesting a broader importance of bacterial NusG paralogs as specialized transcription regulators.

One such paralog, Up_xY, was initially described in *Bacteroides fragilis*, and has since been found to regulate complex polysaccharide biosynthetic pathways across Bacteroidetes^{139,140}. In fact, many Bacteroidetes encode six or more copies of Up_xY proteins, with each being co-transcribed with a different capsular polysaccharide gene cluster¹³⁹. While it is implicated that each Up_xY regulates the transcription of its associated gene cluster, the ability to modulate expression when moved elsewhere in the genome suggests that these paralogs are specifically recruited to their target operons, likely via elements contained in untranslated leader regions^{139,140}. Aside from Up_xY, *B. fragilis* encodes an additional set of proteins, Up_xZ, which serve to inhibit Up_xY proteins *in trans*¹³⁹. It has been suggested that this second set of proteins offers a hierarchical regulatory mechanism to readily control the expression profile of varying capsular polysaccharides¹⁴¹.

In *Myxococcus xanthus*, a Gram-negative soil bacterium, the NusG paralog TaA controls the production of a polyketide antibiotic, TA (or myxovirescin)^{142,143}. This regulatory relationship was suggested given that the first open reading frame of the myxovirescin biosynthesis operon, *taA*, encodes for the NusG paralog. More convincingly, deletion of *taA* ablates myxovirescin production¹⁴³. While the specific role of TaA in the expression of

downstream genes is not presently understood, its close relation to NusG suggests an impact on transcription elongation, likely by processive antitermination.

ActX, also found in a variety of Gammaproteobacteria, is most phylogenetically similar to RfaH^{144,145}. While ActX proteins have not been well characterized yet, their expression from the initial open reading frames of long pilus biosynthesis operons, and their location on conjugative plasmids harboring antibiotic resistance, presents a case for their involvement in the expression of conjugation-related genes^{144,146,147}.

LoaP

An additional class of NusG paralogs was previously discovered by the Winkler laboratory¹⁴⁸, following a bioinformatic analysis comparing bacterial NusG sequences from the NCBI genomic database. The NusG N-terminal domain (NGN) is highly conserved across, and unique to, NusG proteins^{15,36}, and was therefore used as a search model. Given that the specialization of NusG homologs likely emerges from divergence of the C-terminal domain (CTD)^{36,76}, this search method revealed NusG proteins which are similar (but not identical) to core NusG, with implications for specialized functions. This analysis additionally filtered for *nusG* sequences located in close proximity to biosynthetic operons, particularly those responsible for the production of secondary metabolites and polysaccharides. Considering that these operons tend to range between tens to hundreds of kilobases in length, their unusually long transcripts may require uncharacterized antitermination mechanisms for complete, uninterrupted synthesis. Restricting the bioinformatic search further to organisms that encode for at least two NusG homologs served to identify specialized NusG paralogs aside from core NusG. This analysis

identified a monophyletic group of NusG paralogs that were named LoaP for long operon-associated protein¹⁴⁸.

The LoaP subfamily of NusG paralogs is broadly conserved in Gram-positive bacterial species, particularly in Bacillota, Actinomycetota, and some Spirochaetota. Across these organisms, *loaP* is often located adjacent to biosynthetic gene clusters encoding for secondary metabolites, suggestive of an evolved, potentially regulatory relationship between the two. In *Bacillus velezensis*, the *loaP* gene is adjacent to a 70-kilobase operon (*dfn*) encoding for the synthesis of the antibiotic, diffcidin (**Figure 1-9**). Deletion of *loaP* decreased transcript abundance across the length of the *dfn* operon, whereas complementation mostly restored it. Several intrinsic terminator hairpins were found to be interspersed within the *dfn* operon, as well as a large putative terminator hairpin located in the *dfn* 5' leader region. Transcriptomic data revealed that these terminator sites were bypassed in response to *loaP* expression in vivo, supporting the role of LoaP as an antiterminator¹⁴⁸.

While it stands to reason that LoaP requires nucleic acid determinants for recruitment to the TEC, as seen in all other factor-dependent antitermination mechanisms, the 5' leader region preceding the *dfn* gene cluster contains no *boxA*, *boxB* or *ops* sequence. Instead, the *dfn* leader region produces an RNA hairpin structure possessing a UNCG terminal tetraloop, the existence of which is well conserved in other species encoding LoaP¹⁴⁹. In vitro binding studies performed with purified LoaP from multiple species revealed the ability of LoaP to bind the *dfn* hairpin with nanomolar affinity while no binding was observed to *boxB*, the DNA version of the *dfn* hairpin, or versions of the *dfn* hairpin in which conserved residues had been deleted or mutated. This high affinity, highly specific RNA binding activity by a NusG paralog is insofar unique to LoaP.

Returning to the proposed idea that the diversity, and therefore specialized functions, of NusG paralogs is conferred by the less conserved CTD, it was hypothesized that residues enriched in the LoaP-CTD may grant RNA binding abilities. NusG and LoaP sequences from organisms encoding both proteins were used to generate a multiple sequence alignment and the distribution of amino acids at each position was compared by relative entropy to find positions of significant difference¹⁴⁹. In *B. velezensis*, where core NusG and LoaP share 23% sequence identity, three positions – K145, K148 and R149 – particularly stand out given their enrichment for positively charged residues in LoaP relative to the negatively charged residues at corresponding NusG positions. Alanine substitution and subsequent binding analyses confirmed that these residues, and therefore the CTD of LoaP, participate in RNA binding activity.

Fundamentally, LoaP differs significantly from NusG and the only other well characterized NusG paralog, RfaH. Unlike the *E. coli* antitermination mechanisms that act to prevent Rho termination, LoaP antitermination complexes specifically promote readthrough of intrinsic terminator sites¹⁴⁸. Moreover, the prevalence of LoaP proteins amongst Gram-positive organisms implies that LoaP enacts readthrough in the absence of coupled transcription-translation machinery. This, and the novel RNA binding activity exhibited by LoaP, suggests that the molecular mechanism employed in LoaP antitermination differs from the current NusG/RfaH paradigm. Therefore, a mechanistic study of *B. velezensis* LoaP would be broadly relevant to the entire class of LoaP proteins and would provide new and important insights into the mechanisms of transcriptional elongation in Gram-positive bacteria.

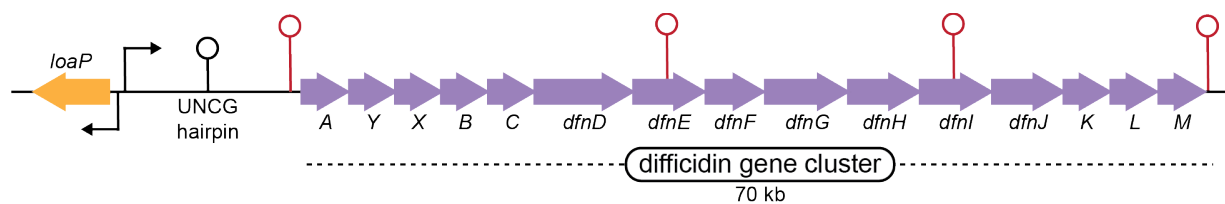


Figure 1-9. Arrangement of the diffcicin operon in *B. velezensis* Schematic depiction of the diffcicin (*dfn*) gene cluster indicating the locations of terminator hairpins (red stem loops) and a UNCG RNA hairpin in the *dfn* leader region (black stem loop). *LoaP* (yellow) is encoded upstream.

Chapter 2: Structure of LoaP bound to RNA hairpin target

Copyright Notice

Some data in this chapter was originally published by Cell Press as: Elghondakly E, Jermain M D, Winkler W, Ferré-D’Amare A R. (2024) Major-groove sequence-specific RNA recognition by LoaP, a paralog of transcription elongation factor NusG. *Structure*, 32, 1488-1497.

Introduction

NusG transcription factors, across all domains of life, possess distinct N- and C-terminal domain structures¹⁵⁰. The highly conserved NusG N-terminal domain (NGN) allows NusG proteins to associate with the transcription elongation complex, via contacts to RNA polymerase and adjacent to the single-stranded, non-template DNA¹⁸. As a functional component of the bacterial TEC, NusG promotes processivity by means of suppressing pausing and backtracking events^{18,151}. The NusG-CTD exists as a binding platform for additional protein factors, forming interactions with Rho to enhance transcription termination events³⁴, or conversely, with ribosomal protein S10 (NusE), bridging the TEC to the leading ribosome in order to prevent Rho-dependent termination^{37,38,151}.

As summarized in Chapter 1, all bacteria encode a core NusG, yet many also encode specialized paralogs which regulate specific gene clusters^{15,89,150}. RfaH, a specialized paralog of NusG known to regulate the expression of virulence operons in Gammaproteobacteria¹²³, is the only other paralog that has been studied in molecular detail. As is true of NusG, the RfaH-NTD forms contacts to RNAP, following specific recruitment to RfaH regulons based on a leader

region hairpin structure, *ops*¹³⁵. The CTD, incapable of binding Rho, forms contacts to S10 in an effort to bridge the transcription and translation complexes and prevent termination^{123,137}.

The NusG specialized paralog LoaP, found commonly in Bacillota, Actinomycetota and some Spirochaetota, is encoded near unusually long operons typically encoding for the biosynthesis of secondary metabolites and polysaccharides¹⁴⁸. In *B. velezensis*, the organism in which LoaP was identified, LoaP regulates the expression of difficidin (*dfn*) and macrolactin (*mln*) polyketide biosynthesis operons. Both gene clusters are preceded by long 5' leader regions which possess a small RNA hairpin structure previously implicated in LoaP antitermination activity¹⁴⁹. A previous study from the Winkler lab¹⁴⁹ confirmed that LoaP from three separate species – *Bacillus velezensis*, *Thermoanaerobacter pseudethanolicus* and *Slackia heliotrinireducens* – all bound the RNA element, herein referred to as the *dfn* hairpin, with high affinity, while neither NusG nor RfaH recognized the hairpin. Strikingly, LoaP is the only known NusG family protein which demonstrates sequence-specific RNA-binding activity. Given this, and the enrichment of positively charged residues within the KOW motif¹⁴⁹, it was predicted that LoaP would rely on its CTD to make contact with an RNA ligand, while the LoaP-NTD, akin to both NusG and RfaH, would enable interactions with RNAP.

In an effort to better characterize the RNA-binding activity displayed by LoaP in the context of antitermination, a co-crystal structure was obtained that resolved the ribonucleoprotein complex between *T. pseudethanolicus* LoaP and the *dfn* hairpin ligand. The resulting structural model revealed that the global folds of the N- and C-terminal domains closely resemble those observed for *E. coli* NusG and RfaH proteins. However, unlike NusG and RfaH, it appears LoaP employs both terminal domains to bind its target nucleotide sequence.

The efforts described herein set out to confirm the molecular basis by which LoaP interfaces with its RNA ligand. More specifically, structure-guided mutagenesis of the *dfn* hairpin was carried out for in vitro binding experiments. This study confirmed that both LoaP domains participate in sequence-specific RNA-binding activity. These data suggest that LoaP employs an antitermination mechanism unique from those previously described for phage gene expression, expression of *E. coli* rRNA operons, or RfaH.

Results

LoaP-*dfn* Crystal Structure

A crystal structure of selenomethionyl *T. pseudethanolicus* LoaP in complex with the 26-nucleotide RNA hairpin derived from the *B. velezensis* *dfn* 5' leader region was determined¹⁵². The established structure, refined at 2.7 angstrom resolution, revealed a stem-loop conformation of the RNA with a terminal UNCG tetraloop, as predicted¹⁴⁹. This RNA hairpin structure also revealed five canonical Watson-Crick base pairs at the base of the hairpin, followed by an underwound stem section which presents two single nucleotide bulges (U6 and C9) (**Figure 2-1**). More specifically, the solvent-exposed conformation of U6 is permitted by the formation of a hydrogen bond between the 2' hydroxyl group of A5 and the non-bridging oxygen (NBO) of C7 (**Table 2-1; Figure 2-2A-B**). The Watson-Crick base pairs flanking this bulge (C7-G21 and G8-C20) display a unique conformation in which G8 is able to make a cross-strand stack on G21. These features result in a local broadening of the major groove. Three additional canonical Watson-Crick base pairs precede the canonical UNCG tetraloop that is located at the apex of the hairpin. This UNCG tetraloop forms a Z-turn, which likely nucleates the folding of the *dfn*

hairpin and maintains stability of the RNA structure¹⁵³. The pair at the base of the stem loop possesses a hydrogen bond between O2 of U13 and N1 of G16, along with a water-mediated hydrogen bond present between N3 of U13 and N2 of G16. The third position of the tetraloop, C15, extends into the major groove, enabling stacking on U13, as well as a hydrogen bond between N4 of C15 and the NBO of U14.

The underwound stem, in conjunction with the tetraloop, forms protein-binding platforms that interface with electropositive regions of the LoaP surface. The phylogenetic conservation of this positively charged RNA-binding interface suggests it is a common feature amongst LoaP proteins. As expected of NusG proteins, LoaP possesses distinct N- and C-terminal domains. However, while it was initially hypothesized that the LoaP CTD was responsible for all RNA interactions, the structural data revealed that the NTD and CTD both associate with the RNA ligand (**Table 2-1; Figure 2-2A-B**).

The LoaP-CTD maintains a five-stranded β -barrel characteristic of the KOW motif possessed by NusG proteins^{17,123,135}. Residues from the LoaP CTD recognize several features about the bulged pyrimidine present in the stem. The Lys140 side chain makes contacts to the NBO of G8 while Lys136 forms a water-mediated hydrogen bond with the NBO of the bulged uridine. The guanidinium of Arg120 stacks onto the bulged U6, and the backbone amides of Ile135 and Ile138 exhibit hydrogen bonds to the N3 and O4' of U6, respectively. The β 6 and β 7 strands of the LoaP CTD are permitted, by the widened major groove of the RNA stem, to make additional contacts about the UNCG tetraloop. The sidechain of Lys 142 interacts with G17 via a hydrogen bond to O6, and Arg 141 contacts both the O6 and N7 of G11. Alternately, Arg141 is capable of hydrogen bonding the NBOs of both U10 and G11. Arg 144 forms multiple hydrogen

bonds at the RNA tetraloop – to the O2 and 2' hydroxyl of C15, as well as to the NBO of G16 – which are bolstered by interactions from the LoaP NGN domain.

The LoaP-NTD adopts the NGN motif characteristic of NusG proteins, which consists of a three-stranded anti-parallel β -sheet surrounded by three helices^{15,17}. In the case of LoaP bound to the *dfn* hairpin, a long β hairpin extends from the core of the LoaP-NTD to interact with the RNA tetraloop. Both NBOs of G16 are recognized by the guanidinium of Arg 36. Lys43 forms a hydrogen bond to the NBO of U13 and Lys51 bridges a water-mediated hydrogen bond to the NBO of C15. Lys51 also makes contact to the Glu40 side chain via a water-mediated hydrogen bond, perhaps stabilizing the stacking interaction between Glu40 and C15, which results in a hydrogen bond to the 2' hydroxyl group of C15. Interactions by Arg41 are thought to provide electrostatic stability, potentially strengthening interactions at the broadened RNA major groove. More specifically, C15 interacts directly with the main-chain carbonyl of Arg41, which maintains an additional, water-mediated hydrogen bond between the Arg41 amide group and C15.

The functional importance of RNA-interactions permitted by LoaP residues were tested by generating site-directed mutations in the protein and investigating the impact of these mutations on binding affinity in vitro. While these results are detailed in earlier work¹⁵², the mutations with the most detrimental effects on binding affinity were those to LoaP residues responsible for contacting the bulged U6 and the tetraloop, particularly at C15.

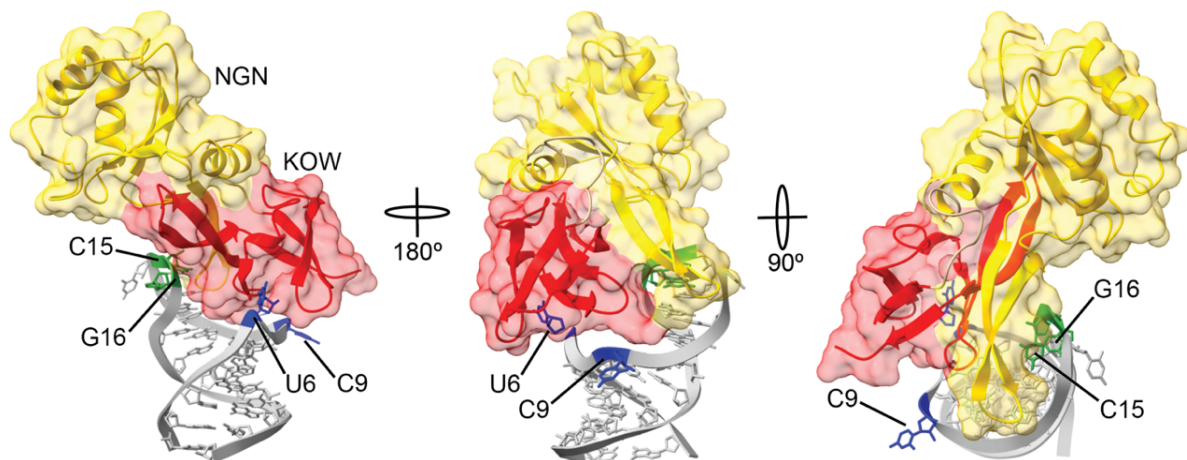


Figure 2-1. Structure of the LoaP-*dfn* hairpin complex Structural representation of LoaP in complex with the *dfn* RNA hairpin (PDB: 8F5G). The LoaP NGN (yellow) and KOW (red) domains, as well as budged nucleotides (blue) and interacting tetraloop residues (green) of the *dfn* hairpin are labeled.

Table 2-1. Summarized interaction sites in the LoaP-*dfn* complex

LoaP-<i>dfn</i> Interactions				
LoaP	<i>dfn</i>	Description	Protein*	RNA*
R36	G16 both NBOs	H-bonds	X	X
E40	C15 2' -OH	H-bond	X	X
E40	C15 NBO	H-bond (H ₂ O mediated)	X	X
R41 main chain carbonyl	C15 N4	H-bond		X
R41 amide group	C15 N3	H-bond (H ₂ O mediated)		
K43	U13 NBO	H-bond		
K51	C15 NBO	H-bond (H ₂ O mediated)	X	X
R120	U6	Stacking	X	X
I135 amide group	U6 N3	H-bond		X
K136	U6 NBO	H-bond (H ₂ O mediated)		X
I138 amide group	U6 O4'	H-bond		X
K140	G8 NBO	H-bond		X
R141	G11 N7 & O6	H-bonds		
<i>R141</i>	<i>U10 & G11 NBOs</i>	<i>H-bonds</i>	X	X
K142	G17 O6	H-bond	X	
R144	C15 O2	H-bond	X	
R144	C15 2' -OH	H-bond	X	X
R144	G16 NBO	H-bond	X	X

LoaP Intramolecular Interactions				
Residue 1	Residue 2	Description	Protein*	RNA*
E40	K51	H-bond (H ₂ O mediated)	X	

<i>dfn</i> Intramolecular Interactions				
Residue 1	Residue 2	Description	Protein*	RNA*
A5 2' -OH	C7 NBPO	H-bond		X
G8	G21	Cross-strand stacking		
U13 O2	G16 N1	H-bond		
U13 N3	G16 N2	H-bond (H ₂ O mediated)		
U13	C15	Base stacking		X
U14 NBO	C15 N4	H-bond		X

Binding measurements of RNA ligands containing substitutions of key functional groups

The high-resolution structure of the LoaP-RNA ribonucleoprotein complex revealed more than a dozen sites of amino acid-RNA contact sites (**Figure 2-1; Table 2-1, Figure 2-2A-B**). However, these predictions are based exclusively on proximity. In other words, amino acids that appeared to be within 1.4 angstroms of RNA nucleotides were assumed to correspond to

sites of protein-RNA interactions. Yet the structural model alone does not determine the relative importance of the different protein-RNA interaction sites. Therefore, the structural model was tested by perturbing some of the putative protein-RNA interactions. To that end, synthetic RNA ligands that contained atomic substitutions at key protein-interacting functional groups were obtained. These RNA molecules were evaluated for their ability to bind LoaP in vitro to test the importance of multiple putative protein-RNA contact sites.

Many of the putative protein-RNA interaction sites involved a non-bridging phosphate oxygen. To investigate some of these sites, commercially produced RNA molecules containing a phosphorothioate linkage rather than a phosphodiester linkage were obtained. The oxygen-to-sulfur substitution is likely to perturb some or most of the amino acid-oxygen interactions that occur. A second theme that emerged from this analysis of putative protein-RNA contacts is that some amino acids appeared to interact with 2' hydroxyl groups. Therefore, RNA molecules containing site-specific 2' deoxynucleotide substitutions were purchased. Alternatively, in some instances, a methyl group was incorporated at the 2' oxygen in order to test putative protein interactions. Finally, a third group of RNA molecules containing other substitutions of functional groups that are likely to interact with LoaP amino acids were obtained. In all instances, the RNA molecules were radiolabeled and assayed for binding to LoaP in vitro using DRaCALA, a modified filter binding assay that can be used to measure the affinity of ribonucleoprotein complexes^{154–156} (**Figure 2-2C-E**).

Each of the functional group substitutions that were tested resulted in significantly decreased RNA-binding activity by LoaP. The most deleterious of these mutations, the substitution of C15 with a 4-thiouridine, was selected for its potential to disrupt the hydrogen bonding potential of the native base. The C15 position was also investigated through a

substitution of deoxyribose for the sugar, and phosphorothioate for the phosphodiester linkage. These changes are predicted to disrupt interactions to residues from both LoaP domains, specifically Glu40, Lys51 and Arg144. This mutant RNA resulted in a significant loss in binding affinity (**Figure 2-2D; Table 2-2**). Similarly, a loss in affinity was observed following phosphorothioate substitutions at non-bridging oxygens (NBO) of U6 and G16, which are predicted to interact with residues from the LoaP-CTD (**Figure 2-2C-D; Table 2-2**).

It should be noted that while these substitutions resulted in the most significant loss in binding affinity, some loss in binding was observed with most of the assayed RNA mutants. These data show that the interactions by LoaP to the UNCG tetraloop and the solvent exposed uridine bulge located at the base of the stem loop are particularly important for successful RNA-binding activity. However, RNA residues that appear to provide structural integrity to these binding sites also serve significant functions. For example, substitutions of RNA functional groups that participate in non-canonical hydrogen bonding interactions between A5 and C7, which maintains the lower uridine bulge, also disrupt LoaP RNA-binding activity (**Figure 2-2C; Table 2-2**). Similarly, atomic substitutions that appear to disrupt G/C base pairs that may stabilize the base of the tetraloop also resulted in significant losses in binding affinity (**Figure 2-2E; Table 2-2**).

In agreement with protein mutagenesis experiments, these data demonstrate that recognition of the UNCG tetraloop, particularly at the conserved cytosine and guanine of the UNCG tetraloop and the bulged uridine in the hairpin stem are critical for binding of LoaP to its target RNA substrate. Not surprisingly, when prior members of the Winkler laboratory identified hairpin sequences that resembled the *dfn* hairpin, the UNCG and lower bulged pyrimidine appeared to be the most conserved features¹⁴⁹, further supporting the role of these residues in the

RNA-binding activity displayed by LoaP proteins. LoaP harbors two distinct domains, which are structurally conserved across the NusG protein family, and utilizes multiple residues in both to recognize regions of the UNCG tetraloop and ribose-phosphate backbone that comprise the *dfn* hairpin element.

Interestingly, the co-crystal structure of LoaP with the *dfn* hairpin is likely representative of its TEC-bound state. Presently, cryo-EM structures have been obtained for both NusG^{18,37} and RfaH¹⁵⁷ bound to the TEC which resolve both proteins interacting with the β subunits of RNAP via their NGN domains. The superimposition of the LoaP crystal structure on either of these reported structures reveals that the RNA-binding surface of LoaP is distinct from the homology-predicted RNAP-binding surface (**Figure 2-3**). This said, a minor clash exists between the β' subunit of RNAP and two β strands of the LoaP-CTD. While it is possible that the *Bacillus* TEC differs significantly from the NusG-bound *E. coli* TECs that LoaP was compared to, it is also possible that the association of the LoaP-*dfn* complex with RNAP triggers conformational changes within the TEC. These changes would likely occur at the RNA exit tunnel, perhaps due to an unknown mechanistic purpose, while also potentially giving LoaP improved access to the nascent RNA.

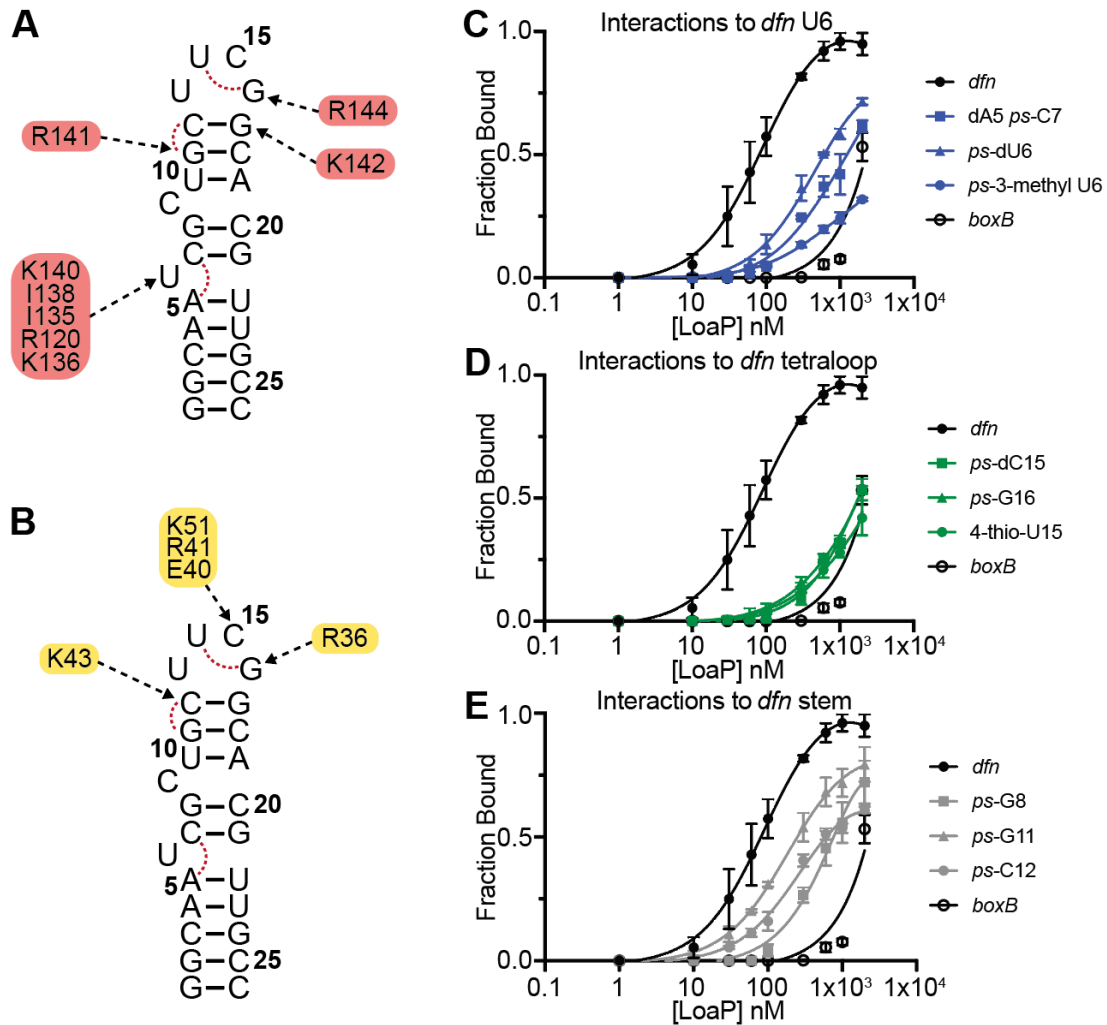


Figure 2-2. Loaf interactions with a bulged pyrimidine and the UNGC tetraloop are critical for RNA-binding Summary of Loaf-*dfn* interactions separated by contacts to the (A) Loaf-CTD (red) or (B) Loaf-NTD (yellow). Dashed red lines indicate hydrogen bonds within the RNA. (C-E) Equilibrium binding curves of Loaf to radiolabeled, synthetic *dfn* RNAs containing atomic substitutions disrupting interactions at the bulged uridine (c), UNGC tetraloop (d) or hairpin stem (e). '*ps*-' indicates phosphorothioate substitution.

Table 2-2. Dissociation constants of LoaP mutant-*dfn* interactions

RNA Ligand	K _d (nM)
<i>dfn</i> hairpin	93 ± 16
dA5, <i>ps</i> -C7	639 ± 506
<i>ps</i> -dU6	489 ± 120
<i>ps</i> -3-methyl U6	501 ± 143
<i>ps</i> -G8	1758 ± 616
<i>ps</i> -G11	188 ± 34
<i>ps</i> -C12	306 ± 51
<i>ps</i> -dC15	NB
4-thio-U15	730 ± 267
<i>ps</i> -G16	686 ± 738
<i>boxB</i> hairpin	NB

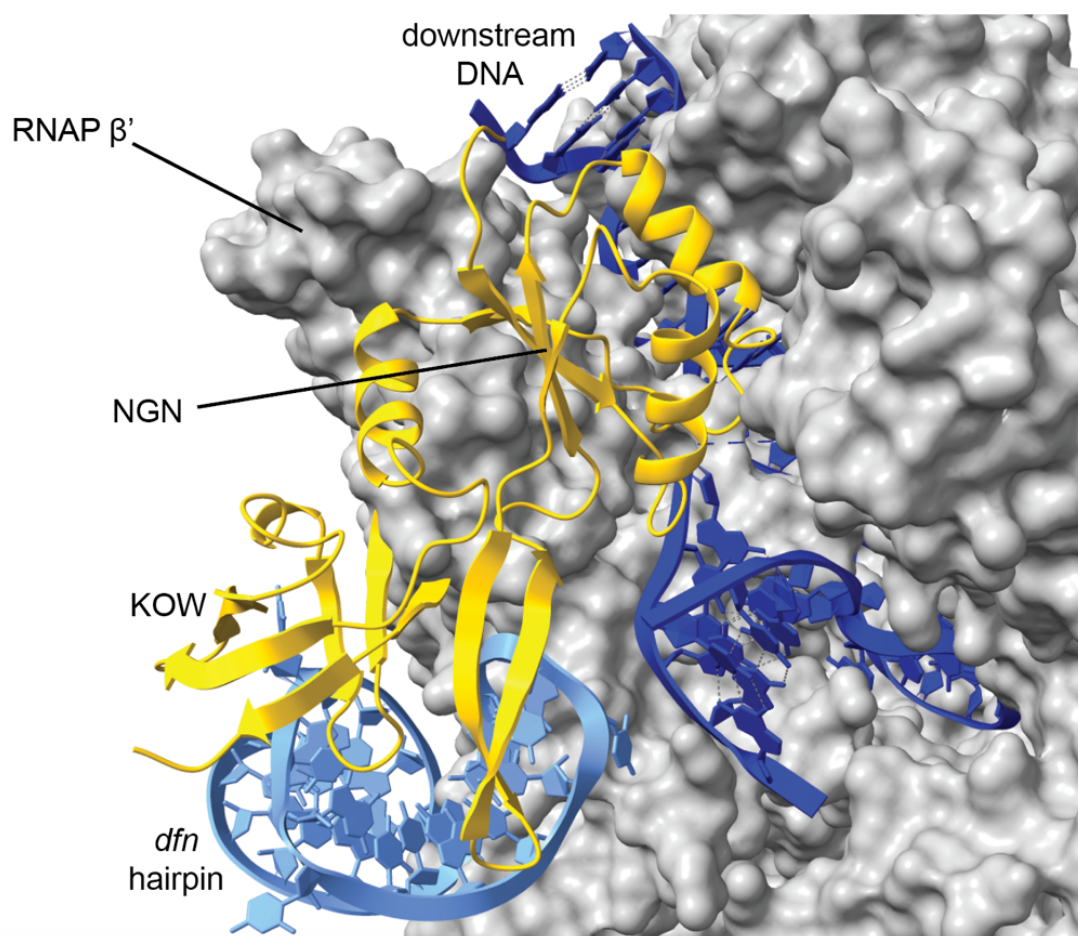


Figure 2-3. Predicted model of LoaP bound to the TEC Structural representation of LoaP (PDB: 8F5G, yellow) bound to the *dfn* hairpin (light blue) and superimposed onto RNAP-bound NusG in a standard *E. coli* TEC (PDB: 6C6U).

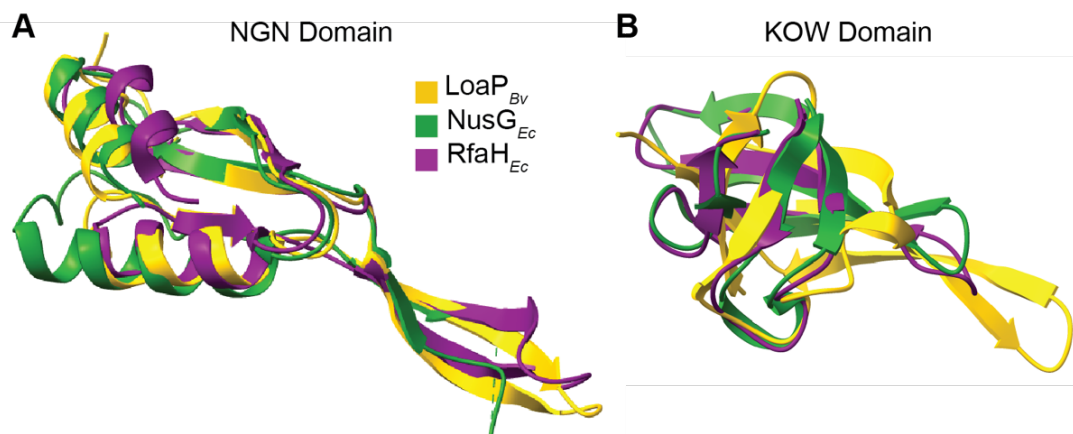


Figure 2-4. Structural comparison of NusG specialized paralogs Superimposition of the (A) NGN and (B) KOW domains of *LoaP_{Tr}* (yellow; PDB: 8F5G), *NusG_{Ec}* (green; PDB: 6GOV) and *RfaH_{Ec}* (purple; PDB:5OND).

Discussion

Presently, *LoaP* is the only *NusG* specialized paralog known to exhibit RNA binding activity, which it carries out by combining strategies characteristic of both DNA- and RNA-binding proteins. Typically, RNA duplexes adopt an A-form helical structure which, unlike the B-form helices exhibited by DNA, offer much narrower major grooves, often limiting access to 3-dimensional protein structures^{158–160}. Instead, proteins that bind to RNA typically interact with double stranded RNA sequences independent of sequence, or interact in a sequence-specific manner only at more exposed regions like bulges, loops or termini^{161–163}. *LoaP* however, recognizes its cognate ligand, the *dfn* RNA hairpin, based on a widened major groove, in addition to recognizing a bulged residue and the UNCG tetraloop itself.

LoaP secures multiple interactions with the bulged U6 residue of the *dfn* hairpin, relying on hydrogen bonding and stacking with the uridine base. The other bulged residue in the *dfn* stem, C9, makes no direct contacts with *LoaP* but rather functions to broaden the major groove

of the *dfn* helix, allowing contacts by the KOW domain and an extended portion of the LoaP-NTD. This sequence-specific major groove binding is reminiscent of DNA-binding proteins but is unprecedented as it pertains to sequence-specific RNA interactions. While UNCG tetraloops are widespread in RNA structures with a propensity for highly stable Z-turn conformations¹⁵³, the co-crystal structure reported herein is only the third known example of a UNCG tetraloop as a protein-binding platform⁹⁵.

Additionally, while the KOW motif has previously been characterized as an RNA-binding domain, particularly when found in ribosomal proteins^{17,33}, this distinct mode of RNA recognition by a KOW domain is novel. The novelty of the RNA-binding activity presented by LoaP is further emphasized by the critical involvement of the NGN domain, which was not previously predicted to be involved in ligand binding but rather association with RNAP only.

The structure of many NusG homologs is thought to reorganize upon recognition of the TEC^{123,128,136,164}. In the case of RfaH, which is recruited in a sequence-specific manner to the non-template DNA, the protein undergoes a structural reorganization of the KOW domain following association with a paused TEC, activating it from an otherwise autoinhibited conformation^{123,136}. This structure switching capability enables the now active, β -barrel structure of the RfaH CTD to engage with ribosomal protein S10, thereby improving transcription-translation coupling¹²³. Moreover, prior bioinformatic analyses predicted that activated structure switching may be a widespread feature of NusG family proteins, suggesting the potential for LoaP to structure switch as well^{165,166}. Having said this, the co-crystal structure of the LoaP-*dfn* complex is consistent with the active forms of previously characterized NusG proteins. When the LoaP-*dfn* structure is superimposed onto previously reported cryo-EM structures of either NusG^{18,37} or RfaH¹⁵⁷ in complex with the TEC, it is evident that both LoaP domains closely

resemble those of TEC-bound NusG proteins (**Figure 2-4**). This suggests that the co-crystal structure is representative of how LoaP may interface with TEC, given that LoaP is capable of occupying a similar binding interface on RNAP with minimal steric clash. However, given that the unbound structure of LoaP is still unknown, it remains a possibility that LoaP could adopt an autoinhibited conformational state, which may be switched into the activated form upon binding the appropriate RNA hairpin, resulting in “active” LoaP. It also remains possible that other NusG-like proteins may recognize RNA elements in a sequence-specific manner while engaging with the TEC. Therefore, structural data on the LoaP-*dfn* complex demonstrates mechanistic features for the regulation of transcription elongation that may be relevant for other NusG homologs.

Materials and Methods

Crytallography

Crystals of a LoaP-RNA complex were obtained, and the structure was determined as described in Elghondakly et al., 2024¹⁵².

Differential Radial Capillary Action of Ligand Assay (DRaCALA)

All RNA and DNA species used in the in vitro binding assays are summarized in the Appendices, and were purchased from IDT and 5'-radiolabeled with [γ -³²P] ATP using T4 Polynucleotide Kinase (T4PNK; New England Biolabs). Longer nucleic acid species (>17 bp) were purified using a Zymo RNA Concentrator and Clean-up Kit. Binding reactions were set up

in 96-well plates by incubating 2.5 nM radiolabeled RNA with increasing amounts of LoaP in Binding Buffer (50 mM NaH₂PO₄ pH 7.2, 100 mM NaCl, 0.1 mM EDTA, 1 mM MgCl₂, 100 µg/mL BSA, 25 µg/mL total yeast RNA) for 45 min at room temperature in the dark. 2 µL aliquots were spotted onto nitrocellulose membranes using a fixed replicator pin tool and allowed to air dry for 20 min. Nitrocellulose membranes were exposed to a phosphor screen for 20 min, visualized using an Amersham Typhoon phosphorimager, and quantified using ImageJ and GraphPad Prism software. All binding data were collected from at least three technical replicates.

Chapter 3: Reconstitution of a *Bacillus subtilis* elongation complex in vitro

Introduction

NusG is used by nearly all organisms as a key factor within the TEC and is therefore a universal feature of DNA-dependent RNA polymerase complexes¹⁶⁷. Bacterial NusG proteins facilitate interactions with other elongation factors and plays crucial roles in regulating the processivity and termination of the transcription machinery^{6,15,18}. As discussed in Chapter 1, all organisms rely on a core, housekeeping NusG, yet many also encode for one or more specialized paralogs of NusG^{15,89,150}. Such paralogs are believed to interact with the TEC akin to, and in place of, core NusG at specific operons, although their mechanistic influences over transcription elongation are largely unknown. In fact, although NusG specialized paralogs are predicted to be widespread across bacterial genomes, the only NusG paralog that has been mechanistically investigated is RfaH, which is widespread amongst Proteobacteria for the regulation of outer membrane proteins^{76,123}. Presently, the antitermination mechanisms enacted by NusG and RfaH have only been characterized within the context of *E. coli*, yet their analyses have established a framework for how NusG paralogs are thought to affect gene expression. Prior studies demonstrate that RfaH is recruited to a paused TEC via interactions with a conserved hairpin structure within the non-template DNA^{18,76,135}. Once associated with the TEC, RfaH bridges the leading ribosome to RNAP, thereby improving transcription-translation coupling¹²³. Genes that are regulated by RfaH feature poor ribosome binding sites, non-standard start codons, and a high degree of rare codons⁷⁶. These poor translation determinants render the targeted operons dependent on RfaH for expression. Unlike the core NusG, the RfaH CTD is incapable of interacting with Rho¹³⁷, thereby conferring antitermination activity to RfaH-associated TECs.

There exists, however, a gap in knowledge pertaining to antitermination complexes employed in Gram-positive bacteria. Given that RNAP from *E. coli* shares only roughly 50% sequence identity to its *B. subtilis* counterpart⁶⁸, it can be inferred that the transcription process differs amongst these model systems. This is certainly true when considering the differences (summarized in Chapter 1) amongst regulatory events such as pausing^{39,75,168,169}, backtracking^{39,170}, and particularly, termination^{6,72}. Most importantly, Gram-positive bacteria are not thought to couple transcription and translation; therefore, antitermination complexes present in these organisms are unlikely to rely on contacts to the leading ribosome⁷². Additionally, Rho, an essential termination factor in Gram-negative bacteria is either absent or largely dispensable in Gram-positive bacteria, which rely more heavily on intrinsic termination instead^{42,72}. These discrepancies highlight the major differences that exist between these organisms and are likely representative of the distinct transcription paradigms in Gram-positives and Gram-negatives. In turn, the transcription elongation complex at large, including associated accessory factors likely exhibits fundamental differences across phyla.

As described in Chapter 1, prior studies by the Winkler lab identified the NusG specialized paralog, LoaP, which is widespread amongst Gram-positive bacteria¹⁴⁸. These studies characterized key molecular features of LoaP, encoded by *B. velezensis*, which differ from NusG and RfaH. The most striking difference is that *B. velezensis* LoaP exhibits antitermination activity against intrinsic termination events rather than conferring anti-Rho activity. Additionally, prior data suggest the LoaP antitermination mechanism likely involves the novel binding of a UNCG RNA hairpin structure within the 5' UTRs of gene clusters regulated by LoaP¹⁴⁹. However, at the onset of the work detailed in this dissertation, the molecular determinants for LoaP antitermination activity, including recruitment signals and any additional

factors, were unidentified. Additionally, it was not known whether the RNA hairpin that binds directly to LoaP was functionally important in vivo, in which it may be an important determinant for the expression of LoaP-regulated genes. Therefore, an investigation of the LoaP molecular mechanism would reveal how Gram-positive bacteria can use antitermination complexes to actively subvert intrinsic termination events, which, by definition, would reveal a mechanism that deviates from the RfaH paradigm for NusG paralogs. For these reasons, efforts described herein sought to reconstitute transcription elongation of *Bacillus subtilis* RNA polymerase in vitro. To employ in vitro transcription strategies (further detailed at the end of this chapter and throughout the next), universal transcription factors – NusA, NusG, NusB, S10 (NusE) and Rho – encoded by *B. velezensis*, were purified for use in concert with a His-tagged *B. subtilis* RNAP previously purified in the Winkler lab. These mixtures were investigated for activity of DNA templates corresponding to the *B. velezensis* *dfn* leader region.

These efforts were significantly hampered by the purification strategy and specific activity of *B. subtilis* RNAP that contained a hexahistidine-tagged RpoC. Although the overall research community relies on this purification strategy for *B. subtilis* RNAP, our laboratory found it to require an unusually high amount of cellular material, and, more importantly, to result in impure protein preparations that exhibited prep-to-prep variation in specific activity. Indeed, our laboratory found that increased purity of the hexa-histidine-tagged RNAP resulted in lowered activity overall. Therefore, a more reliable approach is needed for purification and investigation of *B. subtilis* RNA polymerase. To that end, a new protocol was established for purification of *B. subtilis* RNAP, which resulted in protein preparations that exhibited greater purity while retaining transcription activity in vitro.

Given that the aforementioned transcription factors from *B. velezensis* share remarkable sequence similarity to their *B. subtilis* counterparts (96% between NusAs, 99% between NusGs, 87% between NusBs, 100% between NusE/S10s, and 98% between Rho subunits) it is assumed that transcription by RNAP_{BS} with factors encoded by *B. velezensis* will be representative of broader *Bacillus* TECs. Critical to the formation of effective transcription complexes in vitro is the reliable preparation of highly pure accessory transcription factors, the purification protocols for which are detailed below, exemplified with representative images from the purification of NusA_{BV}.

Results

Overexpression of His-tagged Transcription Factors

The successful purification and increased yield of a target protein begins with the effective overexpression of said protein. How well a protein is expressed can be influenced by various factors, including the expression background, promoter strength, effective protein folding, protein solubility and protein toxicity. To begin optimizing intracellular protein production, the coding sequence of each intended transcription factor (*nusA_{BV}*, *nusB_{BS/BV}*, *nusG_{BS/BV}*, *rpsJ(S10)_{EC/BS/BV}*, *rho_{BV}*) was codon optimized for expression in *E. coli*, meaning codons that may be readily used by the source organism, but rarely in *E. coli*, were swapped for those that are more prevalent in *E. coli*. Synthetic sequences were then cloned into a multicopy vector possessing an IPTG-inducible N-terminal decahistidine tag for downstream immobilized metal affinity chromatography (IMAC). This histidine tag sequence is additionally fused to a

bdSENP-cleavable SUMO tag, the hydrophilic nature of which increases protein solubility and stability, simultaneously decreasing degradation and increasing protein yield¹⁷¹. Expression constructs were then used to transform *E. coli* T7 Express cells (New England Biolabs) engineered for protein overexpression, and expression conditions were tested.

All overexpression strains were grown in 2xYT media at 37 °C, shaking, until reaching mid-exponential phase ($OD_{600} = \sim 0.5$). At this point, target protein expression was induced via the addition of isopropyl- β -D-thiogalactoside (IPTG) at a final concentration of 0.5 mM. Expression was then carried out under one of two conditions: an additional 3 hours at 37 °C, or at 25 °C overnight (~18-20 hours) (**Figure 3-1**). These specific conditions were chosen given that shorter inductions at higher temperatures tend to yield the robust expression of relatively stable proteins, while longer, slower expression at lower temperatures tends to favor proper protein folding and solubility for less stable, or more complex, proteins (New England Biolabs). Following expression, cultures were lysed by sonication and clarified at 13,000 x g for 15 minutes. The resulting supernatants were resolved via SDS-PAGE. Using the purification of NusA_{Bv} as an example, the greatest yield of soluble protein was observed following a 3-hour induction at 37 °C.

When expression was not observed under either of these conditions, the concentration of IPTG was optimized. While higher amounts of IPTG often result in increased promoter expression and downstream protein production, this can be detrimental in the case of unstable or potentially toxic proteins. This was the case when purifying ribosomal protein S10 (NusE), which only resulted in soluble protein that could be isolated from the cytoplasmic fraction when relatively low amounts of IPTG (0.1 mM final) were used for gene induction.

While successful expression was determined by resolving clarified lysates, it is common for a significant fraction of expressed protein to remain in the cell pellet. This can occur when proteins are misfolded, aggregated or associated with additional cellular factors like the cell membrane or nucleic acids (New England Biolabs). Relevant to this project, transcription factors are often highly charged, ensuring crucial contacts to the TEC or directly to genetic elements. Given this, a significant fraction of these factors can be pulled out of cellular lysates upon centrifugation. To combat this, a series of precipitation techniques were employed in order to selectively pellet, or keep soluble, these target proteins.

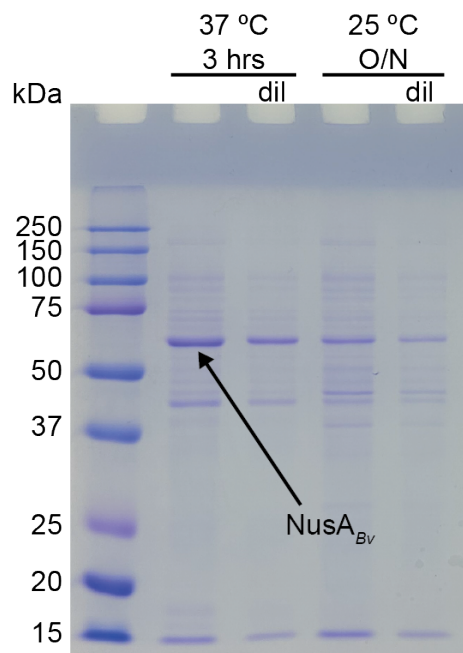


Figure 3-1. The yield of overexpressed NusA_{Bv} benefits from a shorter induction at a higher temperature
 Expression of NusA_{Bv} was induced during mid-exponential growth with 0.5 mM IPTG at either 37 °C for 3 hours (lanes 1 and 2) or 25 °C overnight (lanes 3 and 4). Cells were lysed via sonication and clarified by centrifugation. Lysates were diluted five-fold (or ten-fold, labeled ‘dil’) and resolved via SDS-PAGE. Band corresponding to His-tagged NusA_{Bv} is indicated.

Precipitation techniques enhance recovery of transcription factors

As active components of the TEC, transcription factors often engage with or near the DNA duplex and/or nascent RNA. This nucleic acid-binding capability often results in some level of non-specific binding, which can be detrimental during purification efforts and downstream assays. Persistent nucleic acids can alter the solubility of target proteins and contaminate nucleic acid species of interest used during in vitro experiments (e.g. DNA templates during in vitro transcription). To avoid this, protein precipitation techniques can be employed to selectively enrich fractions for a target protein while precipitating nucleic acids and other proteins.

Polyethylenimine (PEI) is an insoluble, cationic polymer often used for the precipitation of negatively charged species^{172–174}. More specifically, the imino groups present in a PEI molecule exhibit a pK_a value of $\sim 10-11$, conferring a positive charge at neutral pH. Interacting with nucleic acids (and acidic proteins), PEI forms a network which readily precipitates. Analogous to ion-exchange chromatography, the ability of PEI to bind to and precipitate acidic species is dependent on the salt concentration, or ionic strength, of the solution¹⁷⁴. While interactions with nucleic acids are less discriminate, given that the negative charge of the phosphodiester backbone is constant regardless of length or nucleotide composition, the ability of PEI to interact with acidic proteins is less consistent¹⁷⁴. At lower salt concentrations ($\sim < 200$ mM NaCl), most acidic proteins will interact with PEI and precipitate, while higher salt concentrations will elute the protein from the polymer relative to the ionic strength of the protein itself.

Given that most transcription factors themselves are acidic, they can be expected to interact with PEI, along with nucleic acids, at low salt concentrations. In an effort to separate *B*.

velezensis transcription factors from contaminating nucleic acids, a method to pellet only the nucleic acids, while keeping target proteins soluble, is required. To do this, cells overexpressing transcription factors of interest were lysed in high salt (0.5 M NaCl) under the assumption that mildly acidic proteins would remain soluble and nucleic acids would not. PEI was then added to lysates at increasing concentrations and lysates were clarified via centrifugation at 13,000 x g at 4 °C for 15 minutes. Supernatants were retrieved and evaluated via SDS PAGE (**Figure 3-2A**). Using the purification of NusA_{Bv} as an example, PEI was added to lysates at a final concentration ranging from 0.1-0.8% (v/v), yet the high salt concentration of the lysis buffer maintained the solubility of NusA up to a PEI concentration of 0.4%. To optimize the precipitation of nucleic acids while keeping NusA soluble, the remainder of the NusA purification was carried out following PEI treatment at 0.4%.

Alternatively, NusA could have been precipitated along with nucleic acids under low salt conditions and eluted from the polymer using higher salt. However, eluting a protein from PEI also solubilizes the PEI, which would hinder downstream transcription efforts and therefore require an additional PEI removal step.

Another precipitation technique uses the salt ammonium sulfate. Ammonium sulfate, which is very soluble in aqueous solutions, readily stabilizes protein structure while decreasing protein solubility¹⁷⁴⁻¹⁷⁶. More precisely, the highly charged ions produced by ammonium sulfate in solution compete with proteins for interactions with water molecules. As the concentration of ammonium sulfate ions increases, proteins of greater ionic strength are forced out of solution or “salted out”.

This is an effective method of protein enrichment within lysates as ammonium sulfate can be employed to remove less soluble proteins while consequently stabilizing and selectively

precipitating target proteins. In order to carry out a successful ammonium sulfate precipitation, one must determine a percent saturation of ammonium sulfate in which the protein of interest does not precipitate yet other contaminating proteins do. This step can also be employed to ensure that all remaining PEI is removed from the lysate. Following this, a saturation level of ammonium sulfate that does result in the precipitation of a target protein can be used to further concentrate and enrich said protein, while removing contaminating proteins that do not pellet.

Using the purification of NusA_{BV} again as an example, ammonium sulfate was added (in powder form for rapid dissolution) to lysates, following precipitation with 0.4% PEI, in increasing concentrations ranging from 20% to 80% (w/v) saturation. Precipitations were carried out on ice for 20 minutes before centrifuging lysates at 10,000 x g for 15 minutes.

The resulting pellets were resuspended and their contents resolved via SDS PAGE at which point the presence of NusA was determined (**Figure 3-2B**). NusA remained soluble until ammonium sulfate reached 40% saturation. By 50% ammonium sulfate saturation, nearly all NusA precipitated and with less contamination by additional proteins that precipitated at higher levels of ammonium sulfate. This suggests that a two-step ammonium sulfate precipitation at 30% and 50% saturation (a 30-50% cut) would be sufficient for enriching NusA for downstream purification steps.

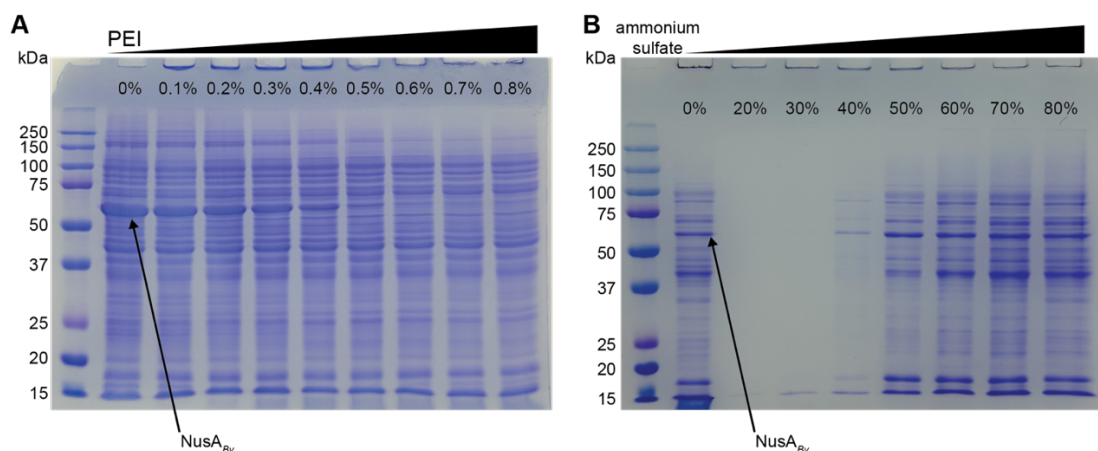


Figure 3-2. Precipitation techniques enhance the recovery of NusA_{BV} (A) Polyethyleneimine (PEI) precipitation test in which PEI was added to cell lysates generated in high salt (500 nM NaCl). PEI was added in amounts ranging from 0.1-0.8% (v/v) and selectively pellets nucleic acids while leaving NusA soluble in most lysates. Resulting supernatants resolved via SDS-PAGE. Band corresponding to His-tagged NusA_{BV} is indicated. (B) Ammonium sulfate precipitation test carried out in lysates already clarified in the presence of 0.4% PEI. Ammonium sulfate was added in amounts ranging from 20-80% saturation (w/v). Protein composition of the resulting pellets resolved via SDS-PAGE. Band corresponding to His-tagged NusA_{BV} is indicated.

Nickel Affinity Chromatography

All target proteins were cloned in frame with an N-terminal decahistidine tag fused to a bdSENP-cleavable SUMO solubility tag. Following ammonium sulfate precipitation, the resulting pellets were resuspended in respective buffers and incubated with Ni-NTA resin in a gravity column. Imidazole, which interacts with nickel ions, was added in increasing concentrations to compete for binding to the nickel resin, eluting lysate proteins based on nickel affinity. Resuspensions often include low amounts of imidazole in an effort to remove proteins which interact weakly with the nickel resin in the flowthrough. Nickel columns were then washed with a slightly greater concentration of imidazole before eluting target proteins with high amounts of imidazole. However, given its ability to interact with histidine, long exposure to high concentrations of imidazole can disrupt both intra- and intermolecular interactions, destabilizing

– and potentially denaturing – target proteins and interrupting protein interactions in downstream applications. Given this, eluted proteins are quickly dialyzed into low-imidazole, low-salt cleavage buffers. Additionally, bdSENP protease was added (1 $\mu\text{L}/\text{mL}$ elution) alongside DTT and MgCl_2 (1 mM and 2 mM final, respectively) at which point protein cleavage was carried out, on ice, for 1 hour (**Figure 3-3A**).

Ion Exchange Chromatography

Conventionally, cleaved histidine tags can be removed via a “reverse” nickel column, in which the cleaved elution is added back to Ni-NTA resin, which readily binds the cleaved affinity tag whilst the now tag-less protein is collected in the flowthrough. Unfortunately, this tactic is not sufficient for separating nickel-affinity tags from many transcription factors. Many proteins interact non-specifically with the nickel resin, often due to the presence of negatively charged regions, exposed histidine residues or a capacity for metal binding, resulting in the substantial loss of eluted protein to reverse nickel columns. Instead, ion exchange columns can be employed in order to separate cleaved histidine tags, which are often uncharged in solutions of neutral pH, from highly charged transcription factors.

In the context of NusA, which exhibits an isoelectric point (pI) of 4.78 and is therefore fairly acidic, or negatively charged, at neutral pH, anion exchange was utilized. During anion exchange chromatography, a positively charged resin is used to attract and bind negatively charged species, which are selectively eluted from the resin based on ionic strength following the addition of charged ions, or salt. During the purification of NusA, the now tag-less protein was fixed to the positively charged ammonium ions of a Q-column. Buffers containing increasing amounts of NaCl were used to wash the column, removing any contaminating protein and

eventually NusA. All eluted fractions were resolved via SDS-PAGE and it was determined that NusA eluted from the Q-column, with substantial purity, following the addition of 400 mM NaCl (**Figure 3-3**). This fraction alone was collected and dialyzed into storage buffer (20 mM Tris-HCl pH 8, 10 % glycerol, 100 mM NaCl, 1 mM DTT, 0.1 mM EDTA), after which the protein concentration was determined, and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

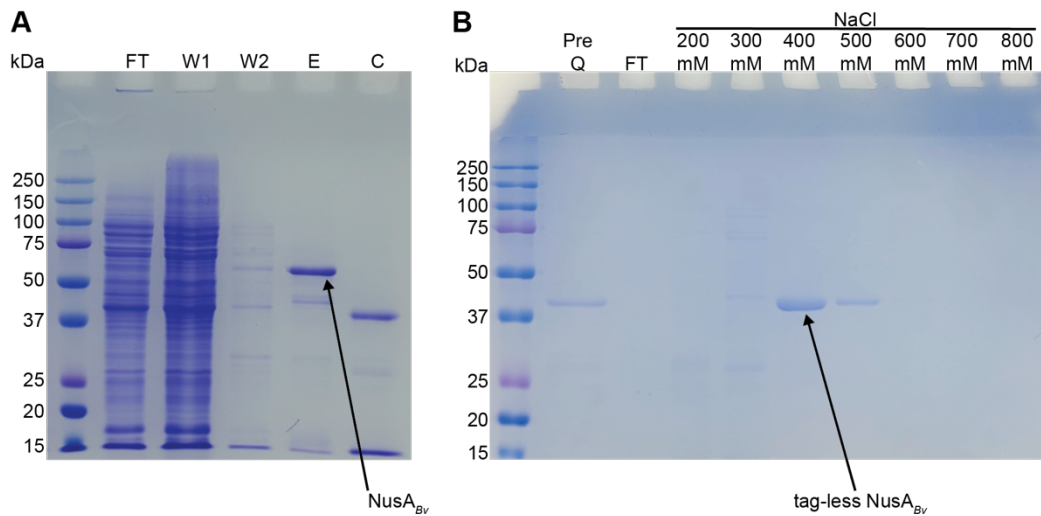


Figure 3-3. SDS-PAGE gels resolving the purification steps for 10xHis-SUMO-NusA_{Bv} (A) Samples following Ni²⁺-affinity chromatography (FT, flowthrough; W1, first wash; W2, second wash; E, elution; C, post-cleavage). Band corresponding to His-tagged NusA_{Bv} is indicated. (B) Elution fractions during anion exchange chromatography. Protein fractions were eluted via the stepwise addition of NaCl in concentrations ranging from 200-800 mM. Band corresponding to untagged NusA_{Bv} is indicated.

Twin-Strep-tagged *rpoC* increases purity of active RNAP_{Bs}

Previous attempts by the Winkler laboratory to purify His-tagged RNAP from *B. subtilis* often resulted in inactive, or poorly active, enzyme. These attempts utilized a cleavable 6xHis-tag translationally fused to the C-terminus of the native *rpoC* gene (encoding the RNAP β’

subunit). These purification attempts were conducted in a $\Delta nusG$ genomic background, as to prevent downstream NusG involvement during the in vitro investigation of LoaP, a NusG paralog. Following purification by both nickel and heparin affinity columns, the resulting protein prep, although of reasonably high purity, elicited no RNA production in vitro. Instead, RNAP activity could only be obtained from purification protocols that omitted the heparin affinity column and resulted in preps of less-than-ideal purity. This likely indicates that the process of obtaining an RNAP_{BS} prep of high purity inactivates/denatures the enzyme or removes a vital cofactor. Additionally, an elevated ratio between the absorbances at 260 nm and 280 nm ($A_{260}/280 = 1.4$) indicated significant nucleic acid contamination that was present in the preparations. While this lower-purity RNAP_{BS} mixture can be successfully employed for downstream in vitro transcription assays (detailed further in Chapter 4), and contaminating nucleic acids were shown not to contribute to downstream RNA production, both contaminating proteins and nucleic acids have the potential to interfere with downstream applications. This is particularly true when investigating the functions of transcription factors, which readily interact with both nucleic acids and other protein factors to form highly stable nucleoprotein complexes.

In an effort to obtain RNAP_{BS} that is both active and highly pure, the nickel-affinity tag was swapped for a Twin-Strep tag, which exhibits greater affinity (in the pM range) to Strep-Tactin resin, an engineered version of biotin-binding streptavidin (IBA Lifesciences). The Strep-tag consists of a short peptide sequence (WSHPQFEK) that binds to streptavidin or streptavidin derivatives (**Figure 3-4A-B**). The Twin-Strep tag contains two of these peptide sequences connected by a short linker sequence. Twin-strep-tagged RNAP was purified via affinity chromatography wherein the resin was washed repeatedly with lysis buffer prior to the addition of excess biotin for elution. The resulting elution fraction was highly pure. An SDS-PAGE gel

resolving the purification steps shows only the α , β , and β' subunits of the RNAP apoenzyme present in the elution fraction (**Figure 3-4C**). Additionally, an A260/280 ratio of 0.8 indicated minimal nucleic acid contamination. Therefore, a single affinity purification of the Twin-Strep-tagged RNAP resulted in a protein preparation that showed greater protein purity than the His-tagged RNAP mixtures.

Strikingly, this highly pure, Twin-Strep-tagged RNAP also displayed transcription activity *in vitro*. Transcription reactions, which were carried out as described previously¹⁷⁷, were performed in both the presence and absence of template DNA possessing a near consensus bacterial promoter sequence. In doing this, any contribution by nucleic acids present in the prepped RNAP to downstream RNA production could be determined. The minimal nucleic acid contamination determined via absorbance readings was evidenced by the absence of a radiolabeled RNA product in reactions that lacked the addition of a DNA template (**Figure 3-5A**).

Perhaps more cogently, Twin-Strep-tagged RNAP was also shown to interact with universal transcription factors as predicted. Within the *B. subtilis trp* leader region, a canonical pause occurs at U107, which is influenced by NusA and NusG, both individually and concurrently¹⁷⁸. To ensure that RNAP operates as intended at a mechanistic pause, this pause site was placed in a DNA template for synchronized, *in vitro* transcription. More specifically, this template possesses a constitutive promoter that lacks a CTP incorporation site until position 29¹⁷⁹. By withholding CTP in the initiation nucleotide mixture, halted complexes were formed at position 29, which, by the addition of all four nucleotides, could be released as synchronized transcription elongation complexes¹⁸⁰. Using this time-resolved assay, the length of the U107

pause could be monitored via urea PAGE in the presence of accessory factors. As previously characterized, the presence of NusA and/or NusG extended the life of the pause (**Figure 3-5B**).

Taken together, these data indicate that employing affinity chromatography wherein the affinity tag present on the protein of interest exhibits a greater binding affinity for its cognate resin, results in a protein elution of significant purity and activity. This method of purification combats the elution of, and downstream contamination by, additional proteins by keeping the protein of interest bound both more selectively and tighter. Additionally, fewer purification steps likely favor active protein by presenting less instances for denaturation, or protein aggregation, or loss of accessory proteins.

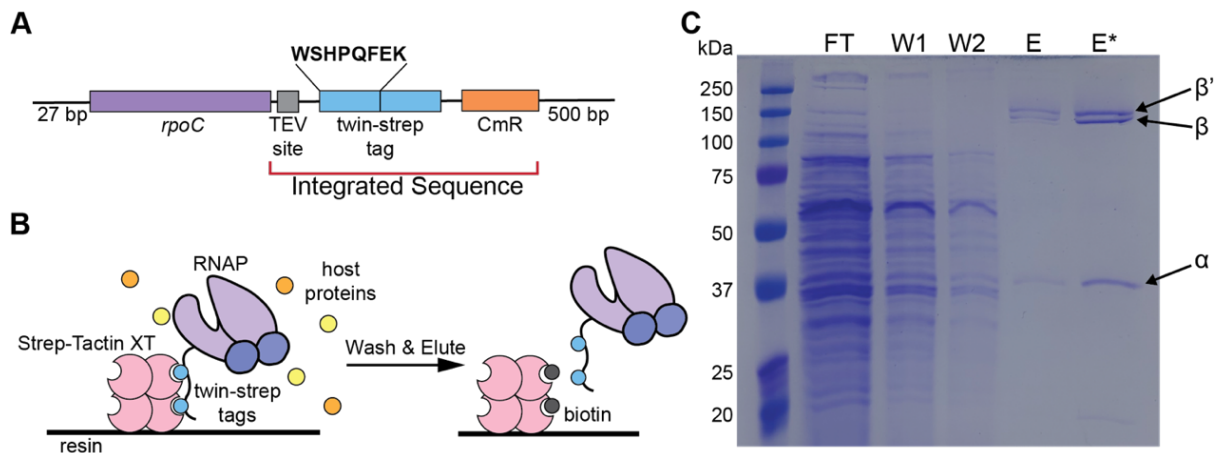


Figure 3-4. Purification of a Twin-Strep-tagged *B. subtilis* RNAP (A) Schematic depiction of the cloning strategy employed to produce a native integration of a TEV-cleavable Twin-Strep tag translationally fused to *rpoC*. (B) Cartoon representation of affinity chromatography method employing Strep-Tactin XT resin for the selective purification of a Twin-Strep-tagged RNAP. (C) SDS-PAGE gel resolving the purification progress of Twin-Strep-tagged RNAP apoenzyme. Bands corresponding to α , β , and β' subunits are indicated. FT, flowthrough; W1, first wash; W2, second wash; E, elution; E*, elution after incubating resin at 90 °C for 5 minutes.

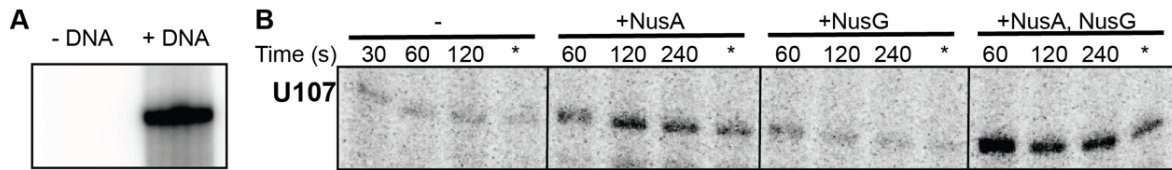


Figure 3-5. Twin-Strep-tagged *B. subtilis* RNAP possesses transcription activity in vitro Urea PAGE gels resolving radiolabeled RNA products of in vitro transcription. **(A)** Reactions were carried out in the absence (- DNA) or presence (+ DNA) of DNA template to ensure that any nucleic acids present in the RNAP preparation did not contribute to downstream RNA production. **(B)** Twin-Strep-tagged RNAP was incubated with a synchronized transcription template possessing the U107 pause characterized in the *trp* leader, which is known to be influenced by NusA and NusG, both independently and together. Reactions were incubated at 37 °C for 15 minutes prior to the addition of NTPs (200 μM), after which aliquots were removed at the specified time points and quenched in Stop buffer. (*) represents the addition of 250 μM NTPs 5 minutes after the last time point.

Discussion

The universal transcription elongation factor, NusG is a crucial part of nearly all bacterial TECs with functions regulating both the processivity and termination efficiency of the transcription apparatus^{6,15,18,167}. That said, NusG has been most abundantly characterized in *E. coli*, with findings strengthened by the extensive study of RfaH, a NusG paralog prevalent in Gammaproteobacteria^{15,89}. Together, information regarding NusG and RfaH has set a precedent for antitermination by NusG family proteins. More specifically, data suggest that NusG paralogs are recruited to target regulons via interactions by their NTD with the non-template DNA and RNAP^{18,31,32,76,135}. These interactions are thought to anchor NusG-like proteins to the TEC, while the CTD forms contacts with the leading ribosome, preventing Rho-dependent termination by improving the coupling of transcription and translation^{35,123}. It is accepted however, that transcription-translation coupling, a crucial component to all presently characterized antitermination mechanisms, does not occur in *B. subtilis*⁷². This phenomenon may have broader implications for other Gram-positive organisms both in the antitermination mechanisms they employ and for transcription elongation as a whole.

Previous work from the Winkler lab discovered LoaP, a NusG paralog present across Gram-positive organisms¹⁴⁸, which are thought to rely more heavily on intrinsic termination and likely do not rely on the leading ribosome to prevent termination events. This presents a lack of knowledge regarding antitermination – and more broadly, transcription elongation – in bacterial systems not closely related to *E. coli*.

LoaP, involved in the production of polyketide antibiotics in *B. velezensis*, enacts antitermination activity against intrinsic termination events¹⁴⁸. Additionally, LoaP interacts directly with a UNCG RNA hairpin in the 5' leader region of target operons with high affinity and high selectivity, using amino acid residues from both LoaP terminal domains^{149,152}. These findings suggest that the antitermination mechanism employed by LoaP greatly differs from those of NusG and RfaH and warrants further investigation into the molecular mechanism by which LoaP results in terminator readthrough. At the onset of the work detailed in this dissertation, the molecular determinants for LoaP antitermination activity, including recruitment signals and any additional factors, were unidentified. While the work done in regard to LoaP implies that there may be regulatory mechanisms employed by NusG family proteins that are yet to be discovered, it also suggests that regulatory mechanisms employed during transcription elongation may differ across model systems.

The work detailed herein set out to obtain a functional *B. subtilis* transcription complex for the in vitro investigation of LoaP antitermination and more broadly, transcription elongation by Gram-positive organisms. Universal transcription factors encoded by *B. velezensis* were purified for use in synchronized transcription assays further detailed in the next chapter. The purification of highly pure accessory transcription factors presents some challenges, however. With a propensity for associating with nucleic acids and additional protein factors, the successful

purification and activity of such proteins hinges upon the ability to effectively remove any co-elutants which may impact downstream applications, while maintaining the solubility of target factors.

Steps to optimize protein expression were done in an effort to favor solubility of target proteins for downstream purification steps. The abundances of target proteins in clarified supernatants were compared following expression with varied amounts of IPTG, incubation temperature and length of expression. Conditions which maximized the amount of soluble protein were used for downstream purification steps in an effort to increase protein yield. Sequentially, the use of PEI and ammonium sulfate precipitation techniques greatly enrich the presence of target proteins, while selectively discarding contamination nucleic acids and proteins from the working sample. Due to the acidic nature of many transcription factors, lysis buffers were supplemented with high salt (500-1000 mM NaCl) before PEI addition, maintaining the target protein in the supernatant while selectively pelleting nucleic acids contaminants. Following ammonium sulfate enrichment steps, nickel affinity chromatography was employed and the resulting elutions were subjected to bdSENP1 protease cleavage. Elutions were then further purified by either “reverse” nickel affinity chromatography or ion exchange based on the propensity of cleaved target protein to interact non-specifically with Ni-NTA resin.

Many of the same issues which afflict the purification of transcription factors also emerge, understandably, during the purification of RNAP. Previous work in the Winkler laboratory to obtain purified RNAP from *B. subtilis* resulted in either inactive enzyme, or an active protein complex teeming with both protein and nucleic acid contaminants. To circumvent these issues, the nickel affinity tag employed in previous purification attempts was replaced with a streptavidin affinity tag. The resulting purification yielded both active and highly pure enzyme

with potential for more robust downstream applications in the absence of contaminants which may otherwise interfere with in vitro assay attempts. With the potential for the more robust purification of transcription factors as well, this highly effective purification technique for a bacterial polymerase also presents potential for investigating macromolecular interactions in additional pull-down or roadblocking methods. With these applications in mind, a Twin-Strep tag was fused to *B. subtilis rpoC* in $\Delta nusG$, $\Delta nusB$, and Δrho backgrounds, for the future interrogation of the *B. subtilis* TEC interaction network.

Materials and Methods

Strain Construction

DNA sequences of desired coding regions were obtained from the National Center for Biotechnology Information (NCBI) genomes database, codon optimized for expression in *E. coli* and purchased from Integrated DNA Technologies (IDT). Synthetic g-blocks were subcloned into a vector possessing an IPTG-inducible N-terminal decahistidine tag fused to a bdSENP-cleavable SUMO tag. All plasmids were assembled using Gibson Assembly techniques and used to transform both *E. coli* XL10-Gold cells (Agilent) for plasmid replication and *E. coli* T7 Express cells (New England Biolabs) for protein overexpression. The composition of each plasmid was confirmed using Sanger sequencing. All sequences summarized in Appendices.

NusA_{Bv} purification protocol

Cultures were grown in 2xYT media at 37 °C, shaking, until reaching mid-exponential phase (OD₆₀₀ = ~ 0.5). IPTG was added (0.5 mM final) and cultures were incubated for an

additional 3 hours. Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 8, 0.5 mM dithiothreitol (DTT), 500 mM NaCl, 5% glycerol, 10 mM imidazole, 0.5 mM phenylmethylsulfonyl fluoride (PMSF)) at 10 mL/g wet cell pellet. Suspensions were lysed via sonication at 40% amplitude for 4 cycles of 20 seconds with 2 minutes on ice between each. Polyethylenimine (PEI) was added slowly, with gentle mixing, to a final concentration of 0.4% (v/v) and lysates were incubated on ice for 5 minutes or until a white precipitate formed. The lysate was clarified at 13,000 x g at 4 °C for 15 minutes in order to remove cell debris and PEI-nucleic acid precipitate. Clarified supernatant was collected and ammonium sulfate was added, in powder form, with gentle mixing to a final concentration of 20% (w/v). Precipitation was carried out on ice for 20 minutes before centrifuging lysate at 10,000 x g at 4 °C for 15 minutes. The supernatant was retrieved, and additional ammonium sulfate was added to increase the final concentration to 70% (w/v). Precipitation was again carried out on ice for 20 minutes prior to centrifuging the lysate once more at 10,000 x g at 4 °C for 15 minutes. The supernatant was discarded, and the resulting protein pellet was resuspended in 30 mL resuspension buffer (50 mM Tris-HCl pH 8, 0.5 mM DTT, 500 mM NaCl, 5% glycerol, 10 mM imidazole). Resuspension was incubated with cOmplete™ His-Tag Purification Resin (Roche) in a gravity column at 4 °C, with gentle rocking, for 1 hour. Flowthrough was drained at ~ 0.5 mL/min after which the nickel resin was washed with 10 column volumes of resuspension buffer, followed by 10 column volumes of resuspension buffer containing 50 mM imidazole. NusA was eluted from Ni-NTA resin into a sterile tube with 5 column volumes of resuspension buffer containing 250 mM imidazole. DTT and MgCl₂ were added to elution at a final concentration of 1 mM and 2 mM, respectively, along with bdSENP1 protease (1 µL/mL elution) and incubated, on ice, for 1 hour. Cleaved elution was dialyzed overnight at 4 °C in Q-

loading buffer (20 mM Tris-HCl pH 8, 5 % glycerol, 100 mM NaCl, 1 mM DTT). The following day, a Strong Anion Exchange Spin Column (Roche) was equilibrated with 5 mL Q-loading buffer and centrifuged at 500 x g for 5 min. The anion exchange column was then loaded with the cleaved nickel elution and centrifuged again. The column was washed with 5 mL aliquots of Q-loading buffer containing increasing salt concentrations (200-800 mM NaCl) in a stepwise fashion, spinning samples at 500 x g for 5 min following each buffer addition. Fractions were evaluated via SDS-PAGE and the NusA-containing fraction was dialyzed into storage buffer (20 mM Tris-HCl pH 8, 10 % glycerol, 100 mM NaCl, 1 mM DTT, 0.1 mM EDTA) at 4 °C for 2 hours, or overnight. The concentration of purified sample was determined, and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

NusB_{Bs/Bv} purification protocol

Cultures were grown in 2xYT media at 37 °C, shaking, until reaching mid-exponential phase ($OD_{600} = \sim 0.5$). IPTG was added (1 mM final) and cultures were incubated at room temperature (25 °C) overnight (18-20 hours). Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 10% glycerol, 500 mM NaCl, 0.5 mM PMSF) at 10 mL/g wet cell pellet. Suspensions were lysed via sonication at 40% amplitude for 4 cycles of 20 seconds with 2 minutes on ice between each. The lysate was clarified at 13,000 x g at 4 °C for 15 minutes. Clarified supernatant was incubated with cOmplete™ His-Tag Purification Resin (Roche) in a gravity column at 4 °C, with gentle rocking, for 1 hour. Flowthrough was drained at ~ 0.5 mL/min after which the nickel resin was washed with 10 column volumes of resuspension buffer (50 mM Tris-HCl pH 8, 5 % glycerol, 500 mM NaCl, 10 mM imidazole, 0.5 mM DTT) followed by 10 column volumes of resuspension buffer containing

50 mM imidazole. NusB was eluted from Ni-NTA resin into a sterile tube with 5 column volumes of resuspension buffer containing 250 mM imidazole. DTT and MgCl_2 were added to elution at a final concentration of 1 mM and 2 mM, respectively, along with bdSENP1 protease (1 $\mu\text{L}/\text{mL}$ elution) and incubated, on ice, for 1 hour. Cleaved elution was added again to fresh nickel resin and incubated on ice, rocking, for 1 hour. Flowthrough was drained at $\sim 0.5 \text{ mL}/\text{min}$ and collected. Cleaved elution was analyzed via SDS-PAGE and dialyzed for 2 hours, or overnight, at 4 °C in storage buffer (50 mM Tris-HCl pH 8, 5 % glycerol, 200 mM NaCl, 0.5 mM DTT). The final concentration of purified sample was determined, and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

NusG *Ec/Bs/Bv* purification protocol

Cultures were grown in 2xYT media at 37 °C, shaking, until reaching mid-exponential phase ($\text{OD}_{600} = \sim 0.5$). IPTG was added (0.5 mM final) and cultures were incubated for an additional 3 hours. Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 8, 1mM DTT, 1 M NaCl, 0.5 mM PMSF) at 10 mL/g wet cell pellet. Suspensions were lysed via sonication at 40% amplitude for 4 cycles of 20 seconds with 2 minutes on ice between each. PEI was added slowly, with gentle mixing, to a final concentration of 0.5% (v/v) and lysates were incubated on ice for 5 minutes or until a white precipitate formed. The lysate was clarified at 13,000 x g at 4 °C for 15 minutes in order to remove cell debris and PEI-nucleic acid precipitate. Clarified supernatant was collected and ammonium sulfate was added, in powder form, with gentle mixing to a final concentration of 20% (w/v). Precipitation was carried out on ice for 20 minutes before centrifuging lysate at 10,000 x g at 4 °C for 15 minutes. The supernatant was retrieved, and additional ammonium sulfate was added to increase

the final concentration to 70% (w/v). Precipitation was again carried out on ice for 20 minutes prior to centrifuging the lysate once more at 10,000 x g at 4 °C for 15 minutes. The supernatant was discarded, and the resulting protein pellet was resuspended in 30 mL resuspension buffer (50 mM Tris-HCl pH 8, 1mM DTT, 1 M NaCl). Resuspension was incubated with cOmplete™ His-Tag Purification Resin (Roche) in a gravity column at 4 °C, with gentle rocking, for 1 hour. Flowthrough was drained at ~ 0.5 mL/min after which the nickel resin was washed with 10 column volumes of resuspension buffer, followed by 10 column volumes of wash buffer (20 mM Tris-HCl pH 8, 300 mM NaCl, 30 mM imidazole, 1 mM DTT). NusG was eluted from Ni-NTA resin into a sterile tube with 5 column volumes of elution buffer (20 mM Tris-HCl pH 8, 300 mM NaCl, 250 mM imidazole, 1 mM DTT). MgCl₂ was added to elution at a final concentration of 2 mM along with bdSENP1 protease (1 µL/mL elution) and incubated, on ice, for 1 hour. Cleaved elution was added again to fresh nickel resin and incubated on ice, rocking, for 1 hour. Flowthrough was drained at ~ 0.5 mL/min and collected. Cleaved elution was analyzed via SDS-PAGE and dialyzed for 2 hours, or overnight, at 4 °C in storage buffer (50 mM Tris-HCl pH 8, 5 % glycerol, 100 mM NaCl, 1 mM DTT). The final concentration of purified sample was determined, and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

S10 (NusE) *Ec/Bs/Bv* purification protocol

Cultures were grown in 2xYT media at 37 °C, shaking, until reaching mid-exponential phase (OD₆₀₀ = ~ 0.5). IPTG was added (0.1 mM final) and cultures were incubated at room temperature (25 °C) overnight (18-20 hours). Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 8, 1 M NaCl, 10 mM imidazole, 0.5 mM

PMSF) at 10 mL/g wet cell pellet. Suspensions were lysed via sonication at 40% amplitude for 4 cycles of 20 seconds with 2 minutes on ice between each. PEI was added slowly, with gentle mixing, to a final concentration of 0.5% (v/v) and lysates were incubated on ice for 5 minutes or until a white precipitate formed. The lysate was clarified at 13,000 x g at 4 °C for 20 minutes in order to remove cell debris and PEI-nucleic acid precipitate. Clarified supernatant was collected and ammonium sulfate was added, in powder form, with gentle mixing to a final concentration of 20% (w/v). Precipitation was carried out on ice for 20 minutes before centrifuging lysate at 10,000 x g at 4 °C for 15 minutes. The supernatant was retrieved, and additional ammonium sulfate was added to increase the final concentration to 70% (w/v). Precipitation was again carried out on ice for 20 minutes prior to centrifuging the lysate once more at 10,000 x g at 4 °C for 15 minutes. The supernatant was discarded, and the resulting protein pellet was resuspended in 30 mL resuspension buffer (50 mM Tris-HCl pH 8, 1 M NaCl, 10 mM imidazole). Resuspension was incubated with cOmplete™ His-Tag Purification Resin (Roche) in a gravity column at 4 °C, with gentle rocking, for 1 hour. Flowthrough was drained at ~ 0.5 mL/min after which the nickel resin was washed with 10 column volumes of resuspension buffer, followed by 10 column volumes of resuspension buffer containing 50 mM imidazole. S10 was eluted from Ni-NTA resin into a sterile tube with 5 column volumes of resuspension buffer containing 250 mM imidazole. DTT and MgCl₂ were added to elution at a final concentration of 1 mM and 2 mM, respectively, along with bdSENp1 protease (1 µL/mL elution) and incubated 4 °C for 12 hours, or overnight. Cleaved elution was dialyzed at 4 °C for 2 hours in S-loading buffer (20 mM Tris-HCl pH 8, 5 % glycerol, 150 mM NaCl, 1 mM DTT). A Strong Cation Exchange Spin Column (Roche) was equilibrated with 5 mL S-loading buffer and centrifuged at 500 x g for 5 min. The cation exchange column was then loaded with the cleaved nickel elution and

centrifuged again. The column was washed with 5 mL aliquots of S-loading buffer containing increasing salt concentrations (300-1000 mM NaCl) in a stepwise fashion, spinning samples at 500 x g for 5 min following each buffer addition. Fractions were evaluated via SDS-PAGE and the S10-containing fraction was dialyzed into storage buffer (50 mM Tris-HCl pH 8, 10% glycerol, 300 mM NaCl) at 4 °C for 2 hours, or overnight. The concentration of purified sample was determined, and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

Rho_{Bv} purification protocol

Cultures were grown in 2xYT media at 37 °C, shaking, until reaching mid-exponential phase ($OD_{600} = \sim 0.5$). IPTG was added (1 mM final) and cultures were incubated at room temperature (25 °C) overnight (18-20 hours). Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 5% glycerol, 1 M NaCl, 10 mM imidazole, 0.5 mM PMSF) at 10 mL/g wet cell pellet. Suspensions were lysed via sonication at 40% amplitude for 4 cycles of 20 seconds with 2 minutes on ice between each. PEI was added slowly, with gentle mixing, to a final concentration of 0.5% (v/v) and lysates were incubated on ice for 5 minutes or until a white precipitate formed. The lysate was clarified at 13,000 x g at 4 °C for 20 minutes in order to remove cell debris and PEI-nucleic acid precipitate. Clarified supernatant was collected and ammonium sulfate was added, in powder form, with gentle mixing to a final concentration of 20% (w/v). Precipitation was carried out on ice for 20 minutes before centrifuging lysate at 10,000 x g at 4 °C for 15 minutes. The supernatant was retrieved, and additional ammonium sulfate was added to increase the final concentration to 70% (w/v). Precipitation was again carried out on ice for 20 minutes prior to centrifuging the lysate once more at 10,000 x g at 4 °C for 15 minutes. The supernatant was discarded, and the resulting

protein pellet was resuspended in 30 mL resuspension buffer (50 mM Tris-HCl pH 7.4, 5% glycerol, 1 M NaCl, 10 mM imidazole). Resuspension was incubated with cOmplete™ His-Tag Purification Resin (Roche) in a gravity column at 4 °C, with gentle rocking, for 1 hour. Flowthrough was drained at ~ 0.5 mL/min after which the nickel resin was washed with 10 column volumes of resuspension buffer, followed by 10 column volumes of resuspension buffer containing 50 mM imidazole and 10 column volumes of resuspension buffer containing 100 mM imidazole. Rho was eluted from Ni-NTA resin with 5 column volumes of elution buffer (50 mM Tris-HCl pH 7.4, 5% glycerol, 1 M NaCl, 250 mM imidazole), then dialyzed for 2 hours, or overnight, at 4 °C in storage buffer (50 mM Tris-HCl pH 7.4, 10% glycerol, 300 mM NaCl). The final concentration of purified sample was determined, and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

Twin-Strep-tagged RNAP_{BS} purification protocol

Cultures were grown in 2xYT at 37 °C, shaking, until reaching late-exponential phase ($OD_{600} = 1$). Cells were harvested by centrifugation and resuspended in strep buffer (100 mM NaH_2PO_4 pH 7.8, 150 mM NaCl, 1 mM EDTA, 1 mM PMSF) at 10 mL/g wet cell pellet. Suspensions were lysed via sonication at 40% amplitude for 4 cycles of 20 seconds with 2 minutes on ice between each. Lysates were clarified at 10,000 x g at 4 °C for 15 minutes. Supernatants incubated with 2 mL Strep-Tactin XT (IBA Lifesciences) in a gravity column at 4 °C, with gentle rocking, for 1 hour. Flowthrough was drained at ~ 0.5 mL/min after which the resin was washed twice, with 5 column volumes each, of strep buffer. RNAP was eluted by incubating resin with 3 column volumes Strep-Tactin Elution Buffer (Zymo) on ice for 1 hour

prior to collecting eluate. RNAP elution was dialyzed overnight at 4 °C into storage buffer (20 mM NaH₂PO₄ pH 7.2, 150 mM NaCl, 10 mM MgCl₂, 10% glycerol) after which the concentration was determined, and aliquots were flash frozen in liquid nitrogen and stored at -80 °C.

Synchronized In Vitro Transcription

Transcription elongation complexes were established by combining 100 nM linear DNA templates with 500 nM *B. subtilis* RNAP apoenzyme in transcription buffer (20 mM Tris-HCl pH 8.0, 40 mM KCl, 5 mM MgCl₂, and 1 mM β-mercaptoethanol). Transcription reactions were supplemented with 500 nM purified *B. subtilis* sigma factor A (σA) and 100 μM ApG to assist in initiation. These transcription mixtures included ATP and GTP at 5 μM and UTP at 2.5 μM, but lacked CTP. They were incubated for 15 min at 37 °C, resulting in elongation complexes halted on the templates at C29. Elongation was stimulated by the addition of nucleotides (ATP, CTP, GTP, and UTP at 200 μM) and heparin (100 μg/mL) to limit re-initiation of free polymerases. Transcription progress was tracked via ³²P derived from the incorporation of [α-³²P]UTP (3,000 Ci/mmol). After 5 min, additional NTPs (Chase) were added (250 μM each). When included, transcription factors NusA, and NusG were added during the formation of the halted complex to final concentrations indicated in the figures. Samples were removed in the time frames specified in the figures and quenched in Stop Buffer (10 M Urea, 20 mM EDTA, 45 mM Tris-borate pH 8.3) prior to phenol extraction. The samples were heated for 2 min at 90 °C and resolved by 6% denaturing PAGE containing 1× TBE and 4 M urea. RNA products were visualized and using a Cytiva Amersham Typhoon PhosphorImaging System, and ImageJ Software. All synchronized transcription data were collected from at least three technical replicates.

Chapter 4: *Bacillus velezensis* LoaP promotes antitermination of antibiotic synthesis operons by antagonizing NusA

Introduction

During transcription, each nucleotide incorporation cycle can be viewed as a decision between synthesis, pausing, arrest and termination. Termination signals are typically located at the ends of operons⁶. However, when they occur within gene clusters they often function as key regulatory checkpoints¹⁸¹. Transcription termination often occurs when the TEC is forced to dissociate from the DNA template and nascent RNA after encountering Rho termination factor^{6,40,41,182}, which is found widely in bacterial genomes. In Gammaproteobacteria such as *Escherichia coli*, the gene encoding Rho is essential¹⁸². In contrast, many Bacillota, such as *Bacillus subtilis*, rely more heavily on intrinsic terminators, showing only minor or moderate phenotypes after deletion of the gene encoding Rho^{71,182}. In addition to Rho, transcription termination can also occur when the TEC encounters an intrinsic terminator hairpin^{6,23}. These RNA elements comprise a G/C-rich helix followed by a polyuridine tract. The polyuridine sequence causes the TEC to pause, which allows the terminator hairpin to form within the RNA exit tunnel of the RNA polymerase, ultimately triggering release of the DNA template and nascent RNA²³. Intrinsic terminators are ubiquitous in the *B. subtilis* genome, typically located within leader regions in addition to the ends of operons⁸⁸, and are often enhanced by the universal transcription factor, NusA^{44,104}.

The fact that *E. coli* relies heavily on Rho for termination, while *B. subtilis* depends more on intrinsic terminator hairpins, suggests that there may be key differences in their mechanisms of transcription elongation. Both organisms rely on similar RNA polymerase proteins, including elongation factors NusA and NusG, yet there are striking differences in the roles of these

elongation factors across species. NusG proteins, which are encoded by all bacteria, associate with both the *E. coli* and *B. subtilis* TEC, yet impact the TEC in different ways. NusG consists of an N-terminal domain which associates with the TEC near the upstream DNA helix^{18,183}. The NusG_{Ec} protein reduces pausing of the *E. coli* TEC, thereby increasing the overall transcription rate¹⁸⁴. In contrast, NusG_{Bs} appears to exhibit a more complicated relationship with the TEC. Similar to the *E. coli* protein, it can prevent transcription pauses from occurring, although it has also been shown to enhance pause sites across the *B. subtilis* genome in a sequence-specific manner^{31,39,75}.

The NusG_{Ec}-CTD, thought to extend from the polymerase surface^{15,33,36}, was previously demonstrated to act as a protein-protein interaction hub, associating in a mutually exclusive manner with either ribosomal protein S10 or Rho^{33,36,57,105}. The protein-protein complex formed by NusG_{Ec} and S10_{Ec} has been previously argued to act as a bridge between the TEC and a leading ribosome, helping to couple transcription with translation while preventing Rho from accessing the NusG_{Ec}-CTD, thereby reducing the overall degree of Rho termination.³⁵

This differs from *B. subtilis*, which appears to employ Rho mostly as a means of preventing antisense transcription and which routinely places intrinsic terminators near stop codons⁷². Additionally, and unlike *E. coli*, the global in vivo transcription rates in *B. subtilis* are double that of the translation rate⁷². Based in part on these data, it has been argued that transcription and translation are uncoupled in *B. subtilis*, in striking contrast to that in *E. coli*. This implies that these organisms are likely to use fundamentally different mechanisms for the genetic regulation of transcription elongation.

Under certain conditions, the *E. coli* TEC can be rendered resistant to Rho, even as the polymerase traverses unusually long DNA sequences⁸⁶. Unlike transcription attenuation

regulatory mechanisms, such as riboswitches, where individual transcription termination regions are bypassed in response to selective cellular conditions⁸⁰, processive antitermination occurs when the TEC is modified to a terminator-resistant form^{85,86}. As described in Chapter 1, several classes of *E. coli* antitermination complexes have been discovered and investigated. These complexes are often recruited to the TEC by nucleic acid determinants, and remain associated with the TEC during transcript synthesis, often blocking Rho access and in turn, termination⁸⁹. Additionally, structural analyses of λ N and *rrn* antitermination have revealed that the resulting ribonucleoprotein complexes often shift the placement of NusA such that it exhibits reduced access to the RNA exit tunnel and prevents Rho access to NusG^{92,94}.

These specialized NusG paralogs are assumed to act as regulatory factors that act on the TEC as it transcribes specific operons within a targeted regulon. Therefore, each of these NusG paralogs can be assumed to associate with the TEC to direct the formation of their own unique antitermination complexes. Among the NusG-specialized paralogs that have been discovered, only RfaH has been investigated in molecular detail. *E. coli* RfaH is recruited to a paused TEC complex by binding to a specific DNA sequence ('*ops*'), displayed in the non-template DNA strand of the transcription bubble^{135,185}. Once bound, RfaH_{Ec} interacts with the TEC through its NTD, similar to NusG_{Ec}¹⁸. The RfaH_{Ec}-CTD is thought to improve coupling between the TEC and the leading ribosome, even more than core NusG_{Ec}, thereby enhancing translation of the targeted genes, which generally feature poor ribosome binding sites and alternative start codons^{76,123}. Crucially, the RfaH_{Ec}-CTD is incapable of binding to Rho¹³⁷. Therefore, this promotes the antitermination of RfaH-targeted operons, which include genes for outer membrane functions and lipopolysaccharide biosynthesis.

The *loaP* gene is frequently found adjacent to biosynthetic gene clusters encoding for the synthesis of secondary metabolites and polysaccharides. In *Bacillus velezensis*, the *loaP* gene is adjacent to a 70-kilobase operon (*dfn*) encoding for the synthesis of the antibiotic, diffcadin¹⁴⁸. Deletion of *loaP* decreased transcript abundance across the length of the *dfn* operon, whereas complementation mostly restored it. Several intrinsic terminator hairpins were found to be interspersed within the *dfn* operon, as well as a large putative terminator hairpin located in the *dfn* 5' leader region. Transcriptomic data revealed that these terminator sites were bypassed in response to *loaP* expression in vivo. This demonstrated that, unlike the *E. coli* antitermination mechanisms that act to prevent Rho termination, LoaP antitermination complexes specifically promote readthrough of intrinsic terminator sites¹⁴⁸.

The antitermination complexes governed by NusG specialized paralogs are undoubtedly different from one another, both in their structural and mechanistic requirements. In particular, the LoaP mechanism must in principle differ from that of RfaH, given that it is primarily found in Gram-positive bacteria¹⁴⁸ which rely more on intrinsic terminator hairpins than Rho termination. More specifically, while the *E. coli* antitermination complexes act primarily to block access to Rho and, in the case of RfaH, improve transcription-translation coupling, *Bacillus* species neither exhibit a high reliance on Rho nor couple transcription with translation; therefore, the LoaP mechanism, which is likely to target intrinsic terminators exclusively, should differ entirely from all known antitermination complexes. A mechanistic study of *B. velezensis* LoaP would be broadly relevant to the entire class of LoaP proteins and would provide new and important insights into the mechanisms of transcriptional elongation in Gram-positive bacteria.

In this study, proteomic and metabolomic analyses were employed to investigate the *B. velezensis* LoaP regulon. This provided conclusive evidence that LoaP selectively regulates the

production of polyketide synthase enzymes for the antibiotics difficidin and macrolactin, while having no effect on the third *B. velezensis* polyketide biosynthesis pathway, which produces bacillaene. These data also demonstrated that modest overexpression of *loaP* led to increased production of these secondary metabolite synthesis enzymes. These data suggest that all determinants for LoaP regulation are contained within the leader regions of the difficidin and macrolactin synthesis operons. In this study, *B. subtilis* RNA polymerase and the housekeeping sigma factor, as well as the elongation factors NusA, NusG, and LoaP, were purified, and transcription elongation across the *dfn* leader region was reconstituted in vitro. These assays identified both a transcription pause site and a large terminator region within the *dfn* leader, with the terminator site being significantly enhanced by NusA. More importantly, the ability of LoaP to bind a small RNA hairpin was demonstrated, potentially as part of a recruitment mechanism, and was shown to promote readthrough of the terminator region in vitro. These data show that LoaP alone can associate with the TEC to promote antitermination, indicating that the LoaP antitermination complex requires only the LoaP-hairpin ribonucleoprotein complex, and additional transcription factors, including a leading ribosome, are not necessary. Furthermore, disruption of the LoaP-hairpin ribonucleoprotein complex in the native *B. velezensis* host resulted in significantly decreased production of difficidin and macrolactin antibiotics. Together, these findings reveal the molecular determinants for LoaP antitermination, which are broadly relevant to the overall widespread class of LoaP regulatory proteins. This knowledge will aid in the design of new synthetic biology tools that can be used to improve the discovery and heterologous expression of natural products.

Results

LoaP is selectively responsible for production of polyketide secondary metabolites in *Bacillus velezensis*

Genes encoding LoaP are often located adjacent to biosynthetic operons responsible for the production of secondary metabolites or polysaccharides. *B. velezensis*, the organism in which *loaP* was first investigated¹⁴⁸, possesses three known polyketide synthesis (PKS) operons, *dfn*, *mln*, and *bae*, which are responsible for the production of the antibiotics difficidin, macrolactin, and bacillaene, respectively¹⁸⁶ (**Figure 4-1**). In a previous study¹⁴⁸, transcriptomic data showed a moderate potential increase in transcript abundance of *dfn* and *mln* gene clusters following *loaP* induction, suggesting that LoaP may directly or indirectly regulate PKS expression in *B. velezensis*.

In this study, the LoaP regulon was directly investigated using proteomic analysis comparing four strains: wild-type, $\Delta lo a P$, a *loaP* complementation strain and an “overexpression” strain that contains the native *loaP* gene as well as an IPTG-inducible, second copy of *loaP*. Cells were grown to mid-exponential growth phase before IPTG was added to all cultures and incubated with the cells for 2 hours. Cells were then collected for trypsin digestion and mass spectrometry analysis to identify proteins which showed significant changes in abundance.

A total of 290 proteins showed a greater than two-fold increase in abundance in the *loaP* complementation strain, upon induction of *loaP*, as compared to the $\Delta lo a P$ strain. Of these candidates, 67 proteins exhibited at least five peptide spectral matches, and more importantly, also showed an increase in the *loaP* overexpression strain compared to $\Delta lo a P$ (**Figure 4-2A**). Approximately one-third of the LoaP-upregulated proteins (23 total) were derived from genes located within two separate polyketide synthase gene clusters, the difficidin and macrolactin

polyketide synthesis pathways (**Figure 4-2C-D**). However, bacillaene pathway proteins were not influenced by the presence or absence of the *loaP* gene (**Figure 4-2B**).

Polyketide synthases are unusually large proteins with a high degree of sequence similarity between one another. This suggests that some of the peptides identified by these proteomic analyses may exhibit redundancy between different polyketide synthases. Therefore, a unique peptide was identified for each gene located throughout the diffcidin, macrolactin, and bacillaene biosynthesis gene clusters, which again revealed LoaP-dependent expression of the *dfn* and *mln* genes (**Figure 4-3**).

To directly investigate whether the proteomic data belied meaningful outcomes for production of the polyketide antibiotics, mass spectrometry was used to quantify the relative abundance of the three polyketide metabolites (oxydiffcidin, macrolactin and bacillaene) within the different *loaP* expression strains. This analysis demonstrated that the production of diffcidin and macrolactin was entirely dependent on LoaP, whereas bacillaene was produced irrespective of LoaP (**Figure 4-2E-F**). While prior transcriptomic data have shown that LoaP triggers modest changes in transcript abundance of the *dfn* and *mln* operons¹⁴⁸, the data herein demonstrate that the presence of LoaP is absolutely required for natural product formation; therefore, LoaP regulation is meaningful at the metabolomic level.

Many of the remaining proteins that were enhanced by LoaP have implications in iron and oxidative stress responses, while seven of these proteins are encoded by a single gene cluster involved in nitrate reductase production (*nar*). This locus also includes a putative transcriptional regulator enhanced by LoaP. Manual examination of the genes encoding these proteins did not reveal any known determinants of LoaP antitermination, making it difficult to determine whether

these genes were direct targets for LoaP regulation. Some of the proteomic changes observed in this study may have been due to indirect regulatory effects instead.

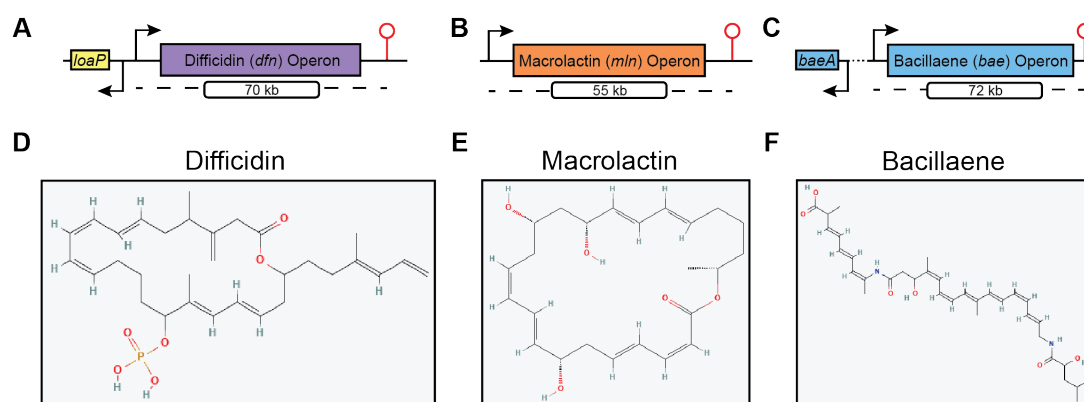


Figure 4-1. Polyketide antibiotics produced by *B. velezensis* (A-C) Gene clusters encoded by *B. velezensis* responsible for the biosynthesis of polyketide secondary metabolites (D) difficidin, (E) macrolactin, and (F) bacillaene, respectively.

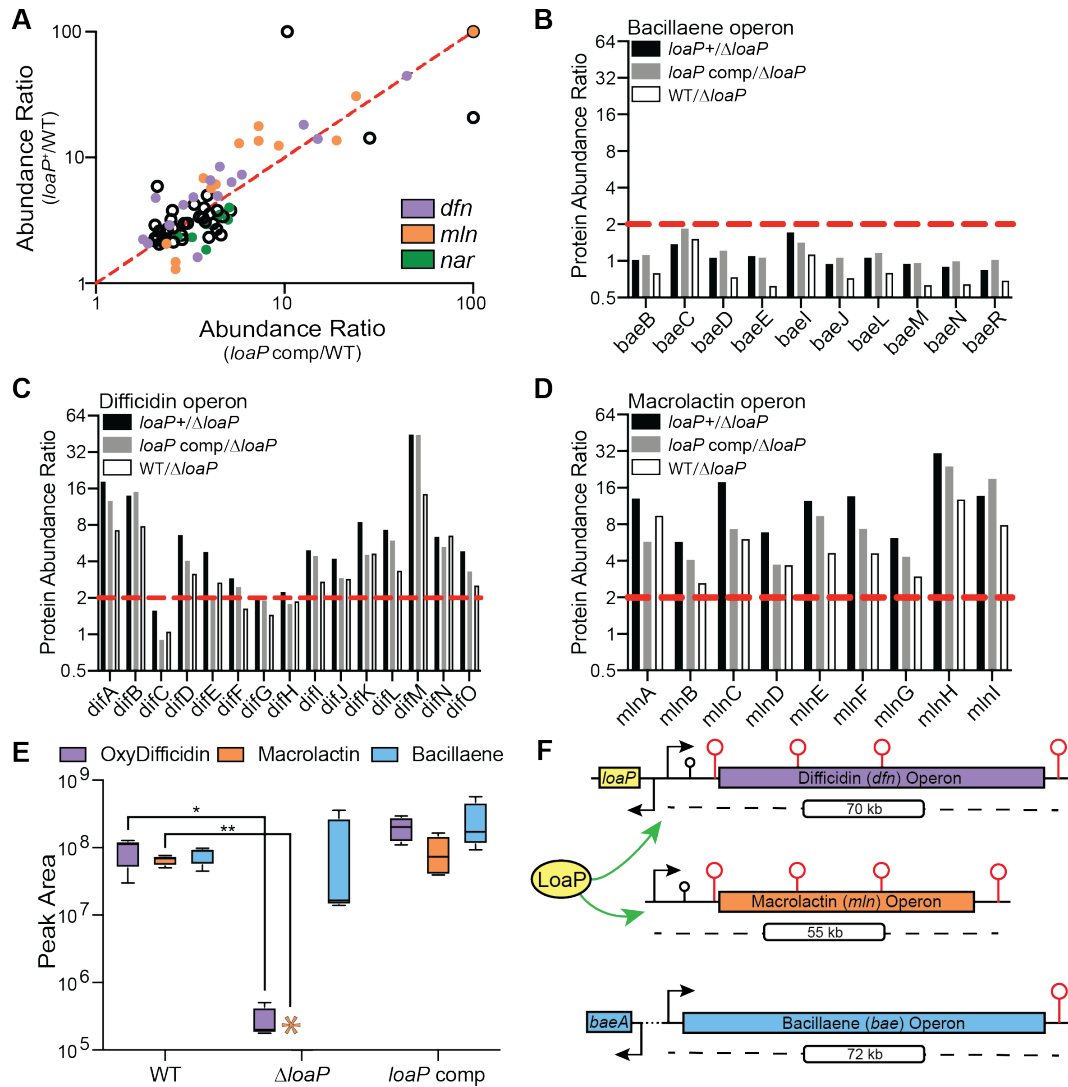


Figure 4-2. LoAP is selectively responsible for the production of polyketide secondary metabolites in *Bacillus velezensis* (A) Plot comparing protein abundance ratios determined by mass spectrometry of a *loaP* complementation strain (*loaP comp*) relative to wild-type (x-axis), and a *loaP* overexpression strain (*loaP⁺*) relative to wild-type (y-axis). Red dotted line at $y=x$ represents threshold for a 1:1 prevalence of peptides in the presence of LoAP (B-D) Abundance ratios relative to a $\Delta loap$ background for proteins corresponding to the bacillaene (b), diffidicin (c), or macrolactin (d) gene clusters determined by mass spectrometry. LoAP overexpression strain denoted “*loaP⁺*”, LoAP complementation strain denoted “*loaP comp*”. Two-fold change in abundance threshold represented by red dotted line. (E) Metabolic detection of oxydiffidicin, macrolactin and bacillaene in *B. velezensis* determined by mass spectrometry. Orange asterisk denotes the absence of macrolactin in a $\Delta loap$ background. Statistical significance was determined following unpaired t tests. One asterisk (*) denotes a p-value < 0.05, two asterisks (**) denotes a p-value < 0.01. (F) Schematic of LoAP-targeted operons. Green arrows identify gene clusters regulated by LoAP. Black stem loops represent the locations of UNCG tetraloop hairpins. Red stem loops represent the locations of intrinsic terminator hairpins.

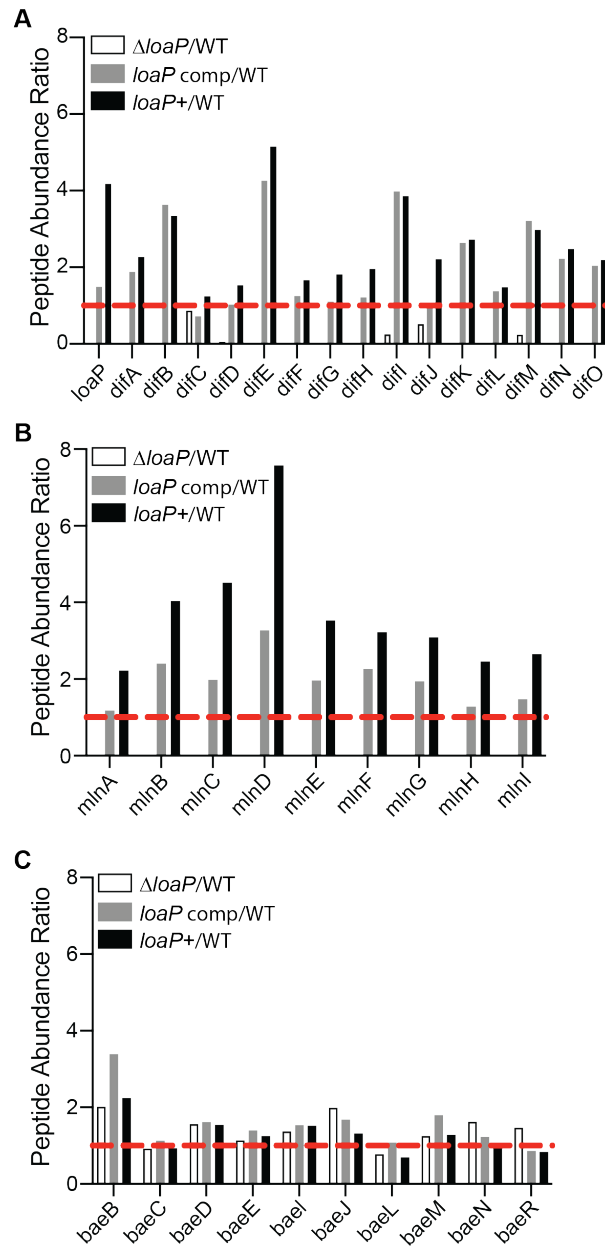


Figure 4-3. Abundance of peptides unique to (A) *LoaP* or proteins encoded by the *dfn* (B) *mln* and (C) *bae* operons determined by mass spectrometry analysis of wild-type *B. velezensis* (WT), $\Delta lo a P$, a *LoaP* complementation strain ($lo a P$ comp) or a *LoaP* overexpression strain ($lo a P^{+}$). Dotted red line denotes abundance ratios > 1.

In vitro transcription of the *dfn* leader region

LoaP likely regulates its direct gene targets using a specific set of molecular determinants that recruit LoaP, and that orchestrate the molecular mechanism of antitermination. To investigate these molecular requirements, transcription of the unusually long *dfn* leader region was examined in vitro. The *Bacillus velezensis* genome is strikingly similar to that of *Bacillus subtilis*, which is consistent with its former assignment as a subspecies of *B. subtilis*^{148,187}. In fact, the universal transcription factors NusA and NusG share 96% and 99% sequence identity, respectively, across these organisms. Therefore, it was reasoned that the *B. subtilis* transcription complex would sufficiently resemble that of *B. velezensis*, to allow the formation of the LoaP antitermination complex. To this end, an affinity-tagged *B. subtilis* RNA polymerase and the key transcription elongation factors NusG and NusA were purified. DNA templates were prepared, wherein the *dfn* promoter was swapped with a constitutive promoter that lacked a CTP incorporation site until position 29¹⁷⁹. By withholding CTP in the initiation nucleotide mixture, halted complexes were formed at position 29 (**Figure 4-4A-C**), which, by the addition of the missing nucleotide, could be released as synchronized transcription elongation complexes¹⁸⁰. Using this time-resolved assay, three major products were observed across the length of a DNA template encompassing the 417 nucleotide *dfn* leader region: a prominent pause site (P), a termination region (T), and a band corresponding to run-off transcription (RO) (**Figure 4-4C-D**). The *dfnA* terminator appeared as two bands that were not adequately resolved during standard urea PAGE and are herein referred to as the *dfnA* “terminator region.” When *loaP* is expressed, the TEC reads through this initial intrinsic termination site, allowing expression of the first gene in the gene cluster, *dfnA*.

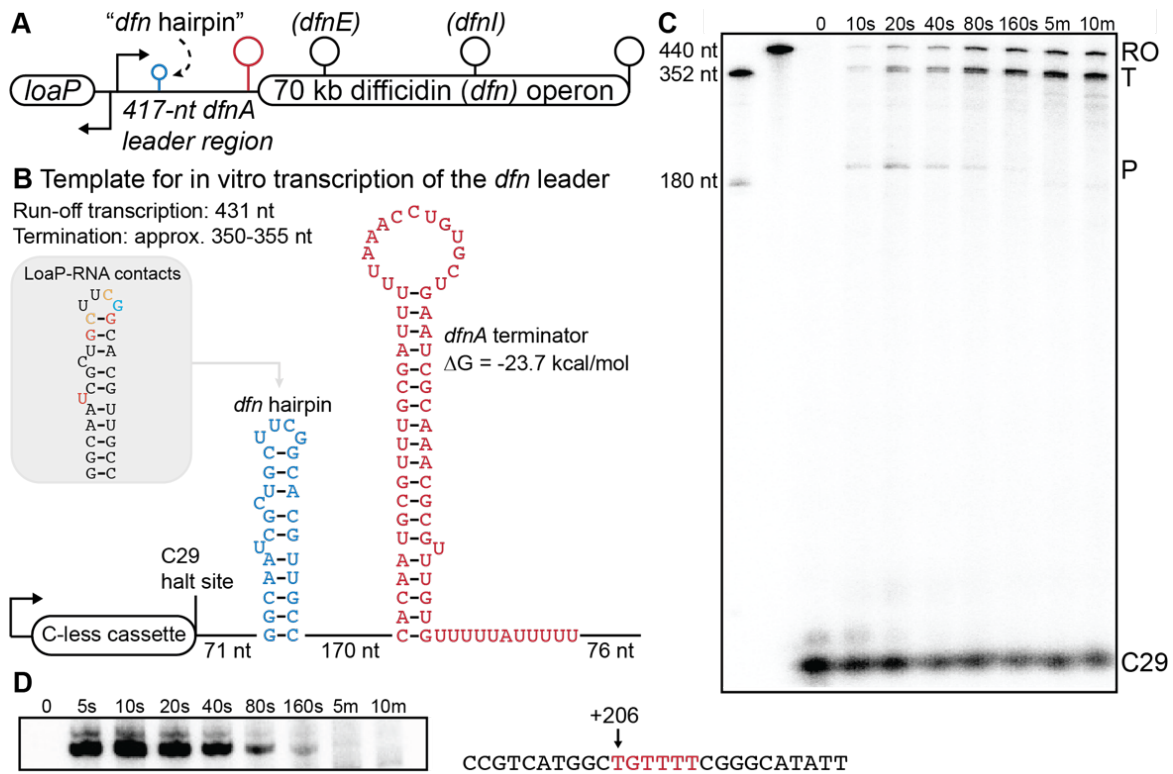


Figure 4-4. Synchronized transcription of the *dfn* 5' UTR reveals a RNAP_{BS} pause site (A) Schematic of the *dfn* gene cluster. The *dfn* hairpin is represented in blue, the *dfnA* terminator hairpin in red and additional *dfn* terminators as black stem loops. (B) DNA template of the *dfn* UTR for in vitro synchronized transcription assays. The *dfn* hairpin is represented in blue and the *dfnA* terminator hairpin in red. Nucleotide lengths between elements and of termination and run-of products are specified. Inset: Loap-RNA contacts are specified as interaction by the Loap-NTD (gold), Loap-CTD (red) or both terminal domains (blue). (C) Urea-PAGE of a synchronized transcription reaction on the *dfn* UTR template. Reactions were incubated at 37 °C for 15 minutes prior to the addition of NTPs (200 μM), after which aliquots were removed at the specified time points and quenched in Stop buffer. Size markers transcribed from PCR products denote termination (T), runoff (RO) and pause (P) products. C29 halt site is indicated. (D) RNAP_{BS} pause site (shown in red) resolved via urea-PAGE (left) and characterized via size markers and dideoxy sequencing ladders (not shown).

NusA enhances termination efficiency of the *dfnA* intrinsic terminator

NusA, an essential protein in *B. subtilis*⁷⁰, has been shown to enhance termination at intrinsic terminator sites, both in vivo and in vitro, by enhancing TEC pausing and modulating the RNAP exit tunnel to assist the formation of terminator hairpins^{25,104,188,189}. NusA_{BV} was titrated into transcription elongation reactions to investigate whether the *dfn* leader terminator

region is affected by NusA. The incremental addition of NusA_{Bv} led to a marked, dose-responsive increase in *dfnA* termination efficiency (**Figure 4-5A**). Given that the cytoplasmic concentration of NusA_{Bs} has been determined to be approximately 1-4 μM ¹⁹⁰, and that *nusA* is essential in *B. subtilis*, these results are likely physiologically relevant.

NusA forms contacts with the β -flap of RNAP, near the nascent RNA exit tunnel, and TEC-associated NusA_{Ec} exhibits RNA-binding activity^{191,192}. Therefore, whether purified NusA associates directly with the *dfnA* terminator in vitro was tested using a modified filter-binding assay (DRaCALA)¹⁵⁴⁻¹⁵⁶. For this experiment, a radiolabeled RNA representative of the *dfnA* terminator sequence was mixed with increasing concentrations of purified NusA_{Bv}, spotted onto nitrocellulose membranes, and then imaged and quantified. In this assay, unbound RNA radially diffuses further than protein-bound RNA, thereby permitting the calculation of the bound fraction. In parallel with the *dfnA* terminator, a weakened terminator (*dfnA**), which lacked several G:C bonds within the terminator stem, and a random unstructured 44-mer were also tested. DRaCALA data indicated that NusA bound to all RNA substrates with an equilibrium binding affinity in the low micromolar range (**Figure 4-5B-C**). Given that NusA did not appear to exhibit any specificity for the terminator sequence, the binding activity observed herein is likely non-specific. From this, it was concluded that *B. subtilis* NusA exhibits non-specific but detectable RNA-binding affinity in vitro.

To assess the influence of NusA in vivo, *nusA* was cloned under xylose-inducible control onto a multicopy plasmid, and the *dfn* leader region was transcriptionally fused to *yfp* and integrated single-copy into the *B. subtilis* genome at a non-essential locus. Upon induction of *nusA*, the fluorescence of individual cells decreased significantly (**Figure 4-5D-E**). In contrast, NusA had no effect on a reporter system that lacked the *dfn* leader terminator region. Taken

together with the in vitro analyses, NusA specifically enhances the formation of the terminator hairpin preceding the first gene in the *dfn* gene cluster, operating as a NusA-dependent, initial regulatory checkpoint for *dfn* expression.

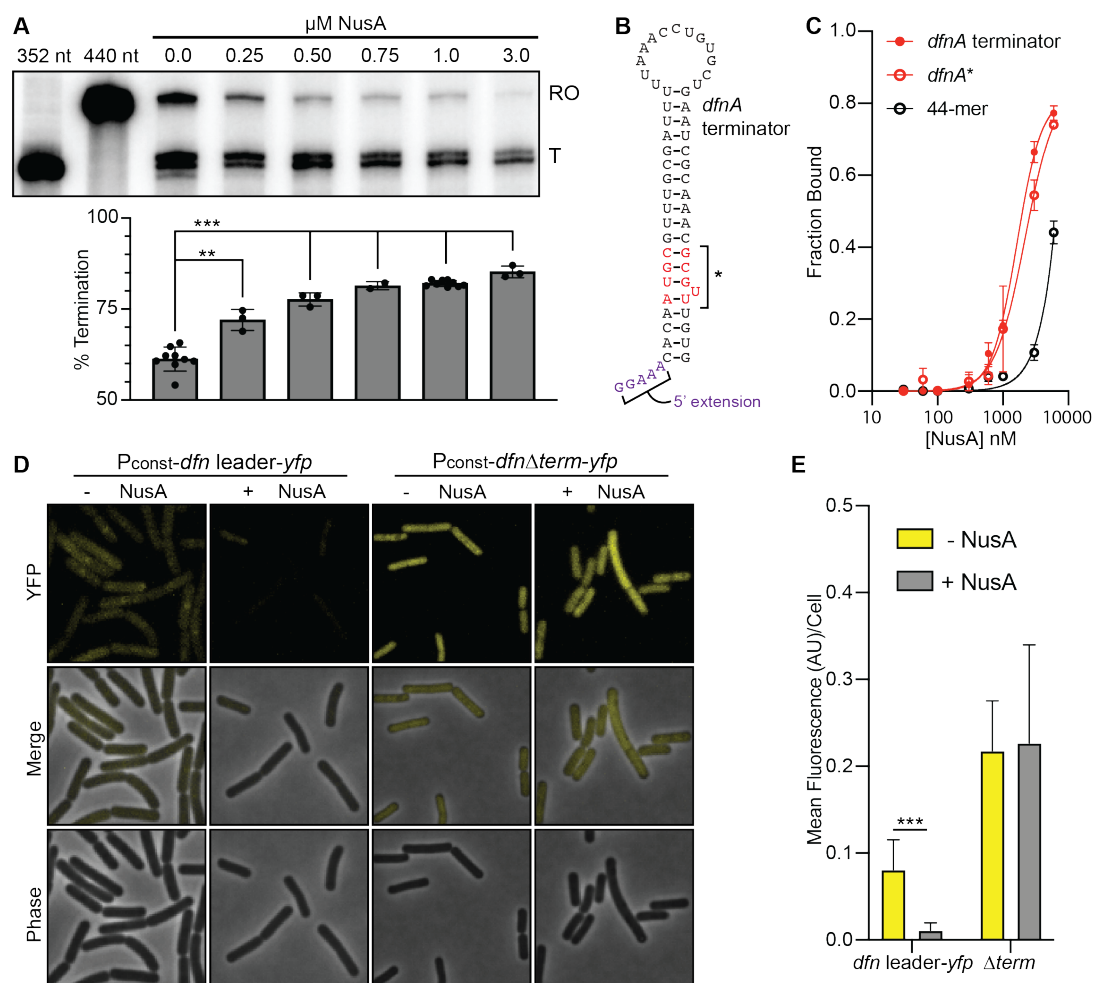


Figure 4-5. NusA enhances termination efficiency of the *dfnA* intrinsic terminator (A) Top: In vitro synchronized transcriptions of *dfn* UTR template with addition of NusA at the specified concentrations. Reactions were incubated at 37 °C for 15 minutes prior to the addition of NTPs (200 μM), then quenched in Stop Buffer after 5 minutes. Size markers transcribed from PCR products denote termination (T) and runoff (RO) products. Bottom: Quantification of transcription reactions measuring termination products as a normalized ratio of overall lane intensity. Statistical significance was determined following unpaired t tests. Two asterisks (**) denotes a p-value < 0.01. Three asterisks (***) denotes a p-value < 0.001. **(B)** Schematic of the *dfnA* terminator. Residues in red indicate those which were removed from a *dfnA** mutant terminator. Residues in purple indicate non-native sequence added to the 5' end for radiolabeling purposes. **(C)** Equilibrium binding curves obtained via DRaCALA wherein increasing amounts of NusA were combined with radiolabeled synthetic RNAs: *dfnA* terminator (red), a mutant *dfnA* terminator (red open) or a randomized, unstructured 44-mer (black open). The normalized fraction bound is shown with error bars representing the standard deviation from the mean. Binding affinities were calculated from at least three experimental replicates. **(D)** Representative microscopy images of *B. subtilis* expressing either a constitutive P_{const} -*dfn* leader-*yfp* reporter or P_{const} -*dfn* leader Δ term-*yfp* reporter integrated into the non-essential *amyE* locus, as well as a xylose-inducible copy of *nusA* on a multicopy plasmid. **(E)** Bar graph representing the average fluorescence intensity per cell measured from (d). Statistical significance was determined following unpaired t tests. Three asterisks (***) denotes a p-value < 0.001.

LoaP antagonizes NusA in the *dfn* leader region

In previously characterized *E. coli* antitermination mechanisms, it was found that additional elongation factors bound to the TEC to reposition NusA away from the exit tunnel, thereby decreasing termination efficiency^{92,94,113}. To test whether LoaP could promote antitermination of a NusA-dependent terminator in vitro, synchronized transcription assays of the *dfn* leader region were conducted (**Figure 4-6A**). These reactions contained physiological amounts of NusA_{Bv} (1 μM) and LoaP_{Bv} at concentrations ranging from 0 to 1 μM. A significant decrease in termination efficiency was observed when 1 μM NusA was combined with 100 nM LoaP (10:1) or higher, with LoaP reverting the percentage of terminated transcripts to near that observed on the template alone (~60%). In contrast, the presence of NusG_{Bv} had no effect on termination in the presence or absence of NusA. Additionally, in vitro binding experiments surveying direct interactions between LoaP and the *dfnA* terminator revealed only non-specific RNA-binding activity by LoaP (**Figure 4-6B**). Thus, LoaP, without the aid of ribosomes or additional elongation factors, can trigger antitermination activity by exerting a direct influence on NusA at the exit tunnel.

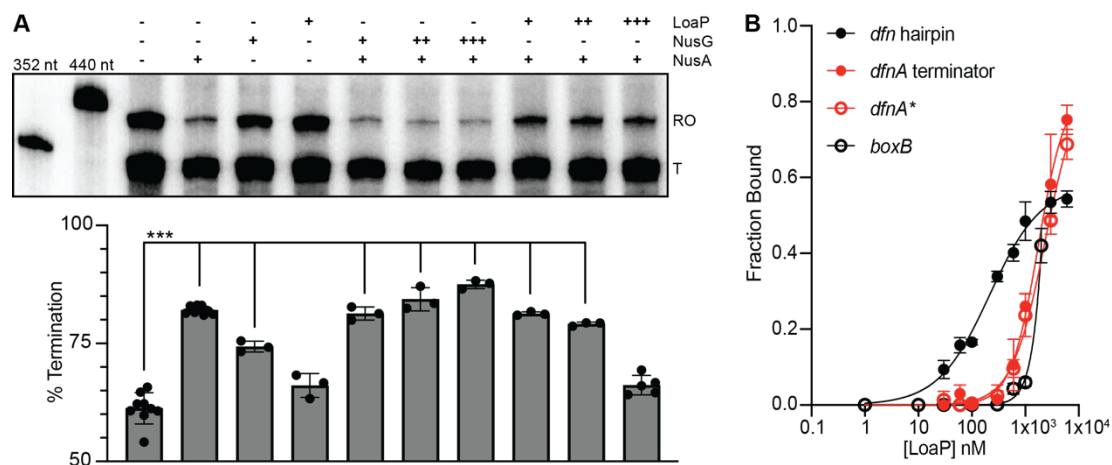


Figure 4-6. LoAP antagonizes NusA-mediated intrinsic termination in vitro (A) Top: In vitro synchronized transcriptions of *dfn* UTR template with addition of NusA, NusG and/or LoAP indicated. NusA added to a final concentration of 1 μ M. For NusG and LoAP, final concentrations of 0.1, 0.5 and 1 μ M are indicated by +, ++ and +++, respectively. Reactions were incubated at 37 $^{\circ}$ C for 15 minutes prior to the addition of NTPs (200 μ M), then quenched in Stop Buffer after 5 minutes. Bottom: Quantification of transcription reactions measuring termination products as a normalized ratio of overall lane intensity. Statistical significance was determined following unpaired t tests. Three asterisks (***) denotes a p-value < 0.001. **(B)** Equilibrium binding curves obtained via DRaCALA wherein increasing amounts of LoAP were combined with radiolabeled synthetic RNAs: *dfn* hairpin (black), *dfnA* terminator (red), a mutant *dfnA* terminator (red open) or the *boxB* hairpin (black open). The normalized fraction bound is shown with error bars representing the standard deviation from the mean. Binding affinities were calculated from at least three experimental replicates.

LoaP antitermination activity is dependent on an RNA hairpin located within the *dfn* leader region

A previous study revealed that a characteristic UNCG-type RNA hairpin (**Figure 4-4B**) is present in both *dfn* and *mln* leader regions and is likely to play an important role in antitermination in vivo¹⁴⁹. It was also shown that LoAP proteins from several bacterial species could bind the UNCG hairpin from the *B. velezensis* *dfn* leader region ('*dfn* hairpin') with sub-micromolar equilibrium binding affinity¹⁴⁹. As described in Chapter 2, the high-resolution three-dimensional structure of the LoAP-RNA ribonucleoprotein complex was determined using X-ray crystallography¹⁵². However, the role of this RNA element is unknown, and its general importance has not yet been established. To investigate whether the hairpin is specifically

important for antitermination activity in vivo, single-cell fluorescence was quantified for a *B. subtilis* strain that contained a transcriptional reporter fusion wherein *yfp* was placed downstream of the *dfn* leader region. This strain also contained a xylose-inducible *B. velezensis* *loaP* gene integrated single-copy into the genome. Induction of *LoaP_{Bv}* led to a significant increase in YFP signal for a wild-type *dfn* leader region; however, the *LoaP*-dependent increase in YFP signal was lost after deletion of the hairpin sequence (**Figure 4-7A-B**). Therefore, the hairpin sequence is required for antitermination activity in vivo.

To test whether the hairpin affected *LoaP* antitermination activity in vitro, DNA templates specifically lacking the hairpin sequence were used for synchronized transcription assays. Halted TECs were incubated on this template with either NusA alone or a combination of NusA and *LoaP*. Although the wild-type template showed a *LoaP*-dependent decrease in NusA-mediated termination (**Figure 4-6**), deletion of the hairpin sequence abrogated this effect (**Figure 4-7C**). After deletion of the hairpin, NusA promoted transcription termination, regardless of the presence or absence of *LoaP*. Therefore, the antitermination activity observed through in vitro experimentation is likely to closely resemble the antitermination mechanism in vivo.

The functional outcome of *LoaP* antitermination is, in theory, the production of the difficidin and macrolactin antibiotics. Therefore, to investigate whether the *LoaP*-RNA hairpin interaction is functionally meaningful, beyond proof-in-principal reporter fusions, *LoaP* proteins that contained site-directed mutations of RNA-binding residues, as well as a non-binding residue K128, were surveyed for alterations in the synthesis of difficidin and macrolactin. Specifically, the mutated *loaP* genes were integrated into a *loaP* deletion strain and all three polyketide metabolites were quantified by mass spectrometry after the induction of *loaP* expression (**Figure 4-7D**). Each of the site-directed mutations suggested to be involved in RNA binding activity led

to a complete or significant decrease in diffcicidin and macrolactin levels. Together, these data demonstrate that the hairpin element is a critically important determinant of *LoaP* antitermination and is specifically required for *LoaP* to interfere with *NusA* activity.

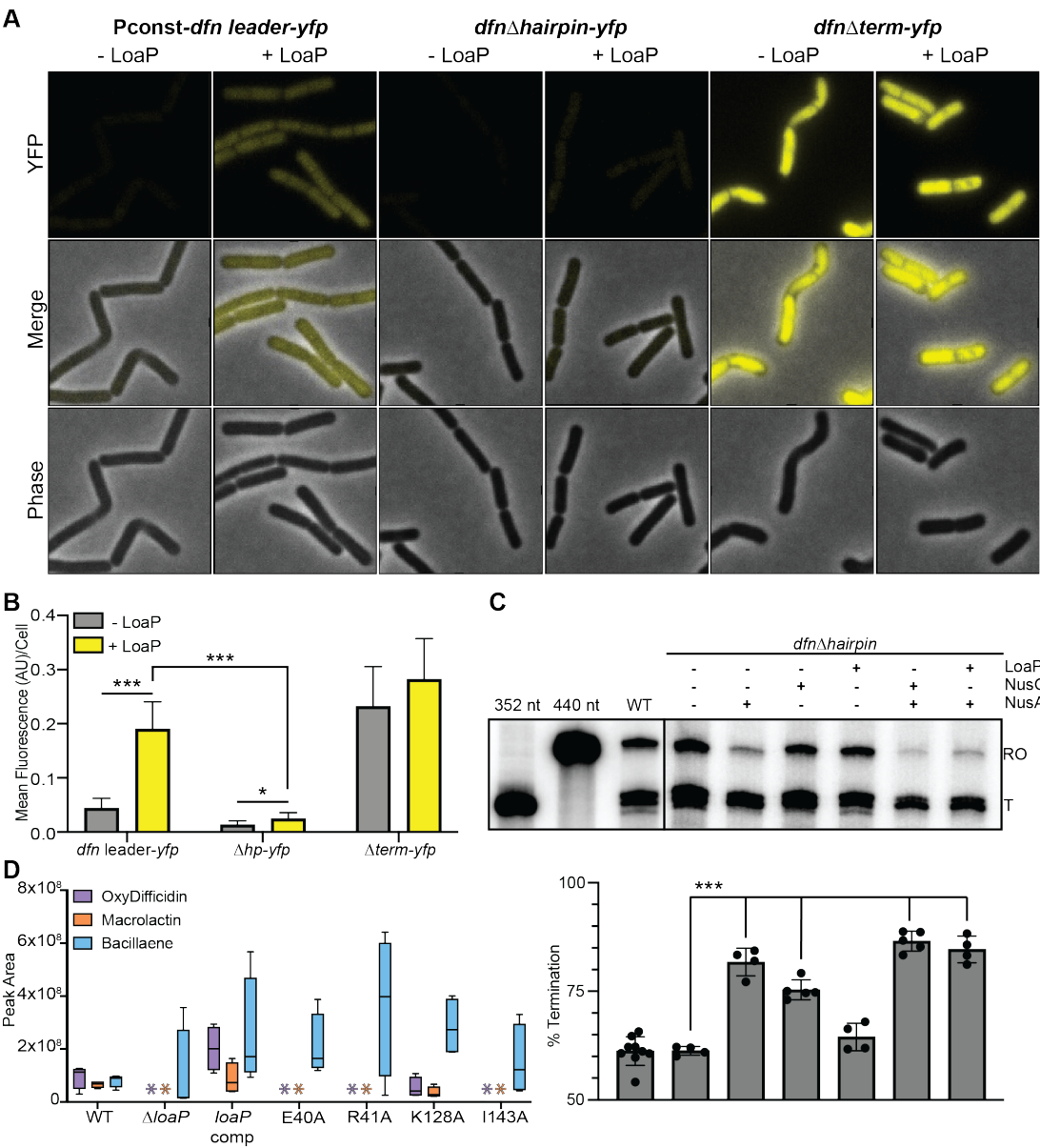


Figure 4-7. LoaP antitermination activity is dependent on an RNA hairpin located within the *dfnA* leader region (A) Representative microscopy images of *B. subtilis* expressing either a constitutive P_{const} -*dfn* leader-*yfp* reporter, a P_{const} -*dfn* leader Δ hairpin-*yfp* reporter, or a P_{const} -*dfn* leader Δ term-*yfp* reporter each integrated into the non-essential *amyE* locus, as well as a xylose-inducible copy of *loaP* integrated into the non-essential *thrC* locus. (B) Bar graph representing the average fluorescence intensity per cell measured from (a). Statistical significance was determined following unpaired t tests. One asterisk (*) denotes a p-value < 0.05. Three asterisks (***) denotes a p-value < 0.001. (C) Top: In vitro synchronized transcriptions of *dfn* UTR template with addition of NusA, NusG and/or LoaP (each at 1 μ M final) indicated. Reactions were incubated at 37 °C for 15 minutes prior to the addition of NTPs (200 μ M), then quenched in Stop Buffer after 5 minutes. Bottom: Quantification of transcription reactions measuring termination products as a normalized ratio of overall lane intensity. Statistical significance was determined following unpaired t tests. Three asterisks (***) denotes a p-value < 0.001. (D) Metabolic detection of oxydifficidin, macrolactin and bacillaene in *B. velezensis*, determined by mass spectrometry. Orange and purple asterisks denote the absence of difficidin or macrolactin, respectively, in a Δ *loaP* background as well as strains expressing LoaP mutants incapable of recognizing the *dfn* hairpin (E40A, R41A, and I143A).

A transcriptional pause within the *dfn* leader region moderately affects LoaP antitermination activity

Interestingly, synchronized transcription reactions of the *dfn* leader region differed modestly when the reactions contained either *B. subtilis* or *E. coli* RNAP. Specifically, while bands corresponding to termination and run-off transcription were observed with both polymerase complexes, RNAP_{Ec} did not generate a pause site that was observed with RNAP_{Bs} (Figure 4-8). This was not entirely unexpected as the two polymerases have been shown to differ in their recognition of elongation regulatory signals⁷⁴. The approximate location of the RNAP_{Bs} pause site was mapped through a combination of size markers and a dideoxy sequencing ladder and found that the pause site resides approximately half of the distance between the *dfn* leader hairpin and the terminator region (Figure 4-4C-D; Figure 4-9A). Because a strong transcriptional pause is essential for the recruitment of RfaH, this study investigated whether the *dfn* leader pause site is also important for LoaP activity in vivo. Specifically, the pause site and surrounding sequence were removed from the *dfn* leader-*yfp* reporter system. When *loaP* was left uninduced, there was no change in readthrough activity compared to the full-length *dfn* leader reporter. However, induction of *loaP* in the absence of the pause sequence resulted in a moderate

reduction of antitermination activity when compared to the full-length *dfn* leader control (**Figure 4-9B**).

To test whether the pause site was influenced by transcription factors, individual proteins or more specifically, combinations of purified NusA_{Bv}, NusG_{Bv}, and LoaP_{Bv} were added in vitro. The *dfn* pause was observed in the absence of purified proteins, suggesting it does not strictly require additional factors. The addition of NusA increased the half-life of the pause site, whereas the addition of NusG decreased the pause length (**Figure 4-9C**). These data add to previously published data showing these elongation factors regulate *B. subtilis* TEC pausing^{31,39,75,189,193}. Unlike its NusG counterpart, LoaP alone did not alter the pause half-life. However, when NusA and LoaP were added together, the NusA enhancement was reduced overall (**Figure 4-9C**), suggesting that LoaP can antagonize NusA-mediated pausing. Moreover, these data suggest that this transcription pause may be important for the association of additional protein factors.

Additionally, while deletion of the pause site did seemingly hinder LoaP antitermination activity in vitro (**Figure 4-10**), this effect is much more striking in vivo. It is possible that the in vitro transcription assay, which relies on synchronized TECs, incompletely reflects the transcription elongation environment observed in vivo. For example, if the pause site functions in vivo to assist in the recruitment of additional factors, this requirement may be reduced or eliminated in a synchronized transcription assay format, where saturating levels of proteins are added during the formation of the halted complex. Taken together, these data suggest that the pause site located within the *dfn* leader is not strictly required for antitermination activity but may significantly enhance it. More so, this pause site may be enhanced by NusA as a way to favor the recruitment of NusG proteins for downstream regulation at *dfnA*.

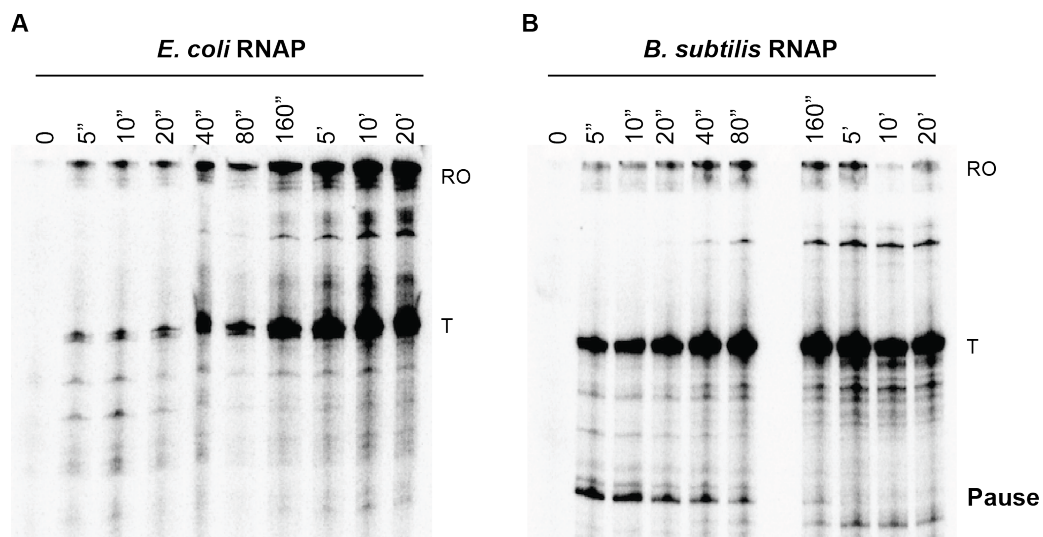


Figure 4-8. A pause in the *dfn* leader region is specific to RNAP_{Bs}. In vitro synchronized transcription of the *dfn* leader with (A) *E. coli* RNAP or (B) *B. subtilis* RNAP. Aliquots were removed and quenched in STOP buffer at the specified time points. Bands representing the pause site, termination (T) and run-off (RO) products are specified.

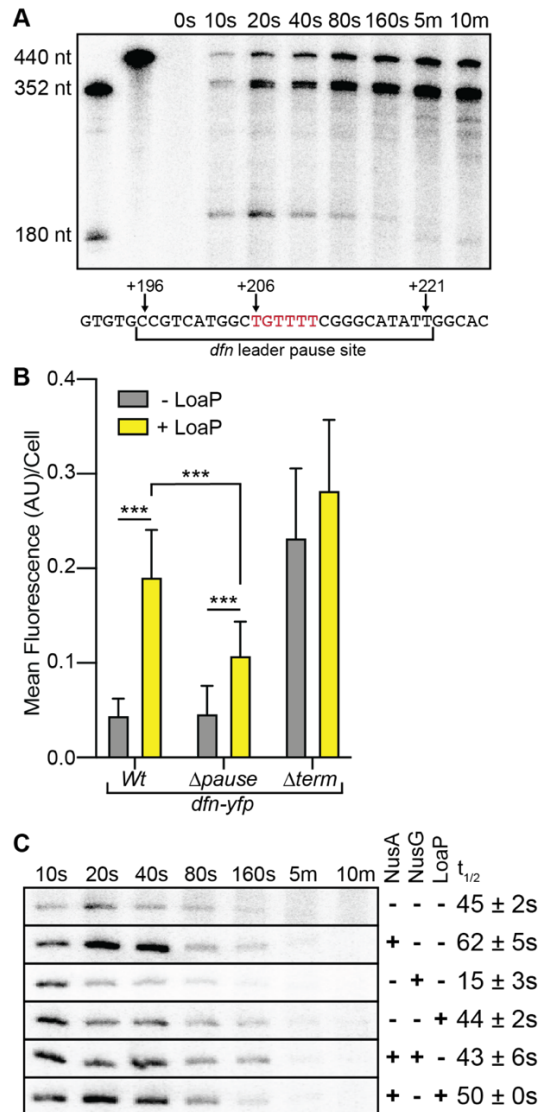


Figure 4-9. A transcriptional pause within the *dfnA* leader region moderately affects antitermination activity (A) Location of a RNAP_{BS}-specific pause in the *dfn* leader region. Top: In vitro synchronized transcription on the *dfn* UTR template. Reactions were incubated at 37 °C for 15 minutes prior to the addition of NTPs (200 μ M), after which aliquots were removed at the specified time points and quenched in Stop buffer. Size markers transcribed from PCR products specified. Bottom: Schematic of the *dfn* leader pause with transcript positions specified. (B) Bar graph representing the average fluorescence intensity per cell measured from microscopy of *B. subtilis* expressing either a constitutive P_{const}-*dfn* leader-*yfp* reporter, a P_{const}-*dfn* leader Δ pause-*yfp* reporter, or a P_{const}-*dfn* leader Δ term-*yfp* reporter each integrated into the non-essential *amyE* locus, as well as a xylose-inducible copy of *loaP* integrated into the non-essential *thrC* locus. Statistical significance was determined following unpaired t tests. Three asterisks (***) denotes a p-value < 0.001. (C) *dfn* pause site resolved via urea-PAGE in the presence of NusA, NusG, and/or LoaP (each at 1 μ M final), as specified. Reactions were incubated at 37 °C for 15 minutes prior to the addition of NTPs (200 μ M), after which aliquots were removed at the specified time points and quenched in Stop buffer. Pause half-lives ($t_{1/2}$) were determined via non-linear regression analysis.

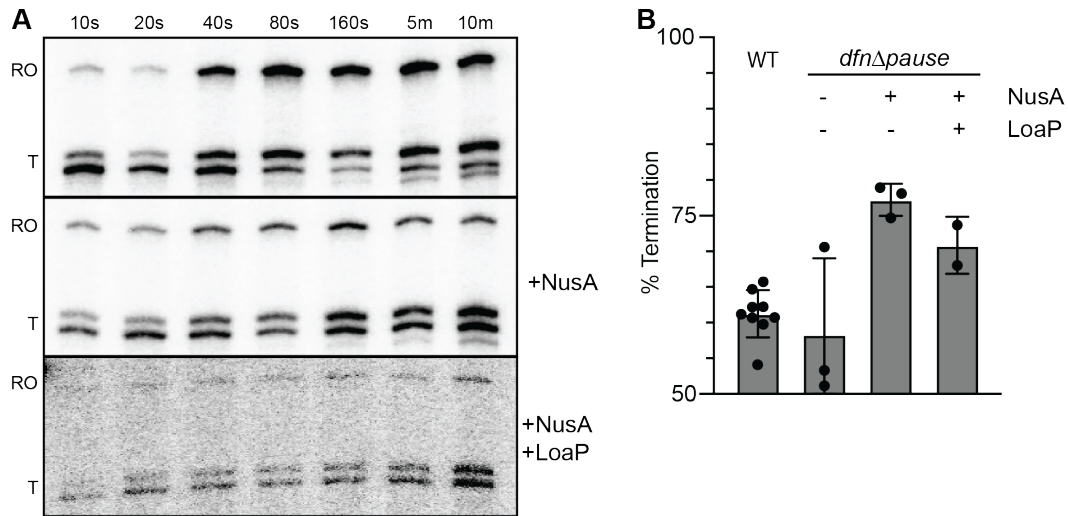


Figure 4-10. Deletion of the *dfn* leader pause sequence in vitro (A) Time course in vitro transcription of $\Delta pause$ templates following the addition of NusA and/or LoP (at 1 μ M final), as specified. Reactions were incubated at 37 °C for 15 minutes prior to the addition of NTPs (200 μ M), after which aliquots were removed at the specified time points and quenched in Stop buffer. **(B)** Quantification of transcription reactions shown in (a), measuring termination products as a normalized ratio of overall lane intensity and compared to a wild-type template.

Recruitment of LoP to the TEC is novel

NusG_{Ec} and RfaH_{Ec} are both recruited to the TEC via their N-terminal NGN domain^{18,129}, which situates each protein immediately adjacent to the non-template DNA sequence. Indeed, RfaH is selectively targeted to operons that contain a characteristic non-template DNA sequence called *ops*, to which RfaH directly binds^{129,135}. Similarly, it has been proposed that *B. subtilis* NusG associates with non-template DNA when bound to the TEC³¹. In all instances, it is assumed that the NusG NTD-TEC interaction leaves the NusG CTD free to interact with transcription and translation factors.

To determine whether LoP might utilize a non-template DNA recruitment sequence, the aforementioned in vivo reporter system was modified to introduce 10 base pair deletions

upstream of the *dfn* hairpin sequence (**Figure 4-11A-B**). The first four deletions had no effect on LoaP antitermination activity in vivo; however, deletion of the 10 base pairs immediately preceding the *dfn* hairpin sequence led to complete loss of antitermination activity. It was then investigated whether purified LoaP associates directly with single-stranded DNA species corresponding to these sequence regions. This analysis revealed only non-specific, low-affinity interactions between LoaP and ssDNA (**Figure 4-11C**). However, this experiment is inconclusive, as it does not rule out the possibility that a potential interaction between LoaP and non-template DNA sequences might also require an active TEC, or that LoaP may bridge interactions to the upstream DNA duplex instead. A comparison of the DNA sequence upstream of the hairpin sequence in the *mln* leader region did not reveal an identical sequence. Therefore, the role of this sequence remains inconclusive. Although it is still possible that a non-template sequence immediately upstream of the hairpin is involved in the recruitment of LoaP, it is also possible that mutational deletion of this sequence interferes with hairpin function in vivo, having nothing to do with LoaP recruitment.

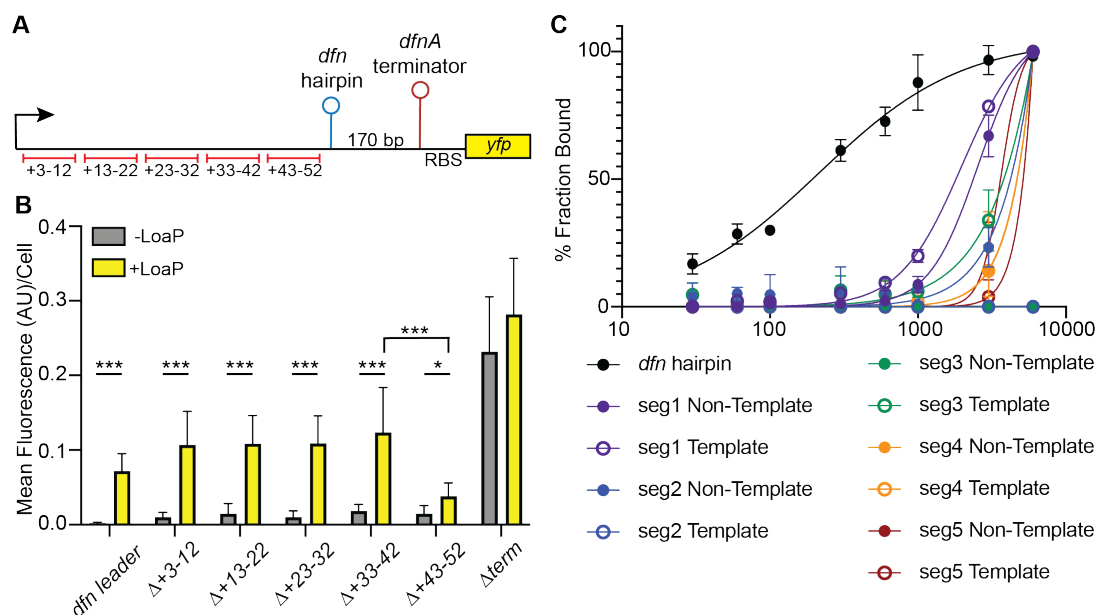


Figure 4-11. Investigation of LoaP recruitment to the TEC (A) Schematic of the deletions made to the *dfn*-*yfp* reporter construct. The *dfn* hairpin is represented in blue and the *dfnA* terminator in red. Deletions of 10 bp DNA segments are denoted by red bars. (B) Bar graph representing the average fluorescence intensity per cell measured from *B. subtilis* expressing a constitutive P_{const} -*dfn* leader-*yfp* reporter possessing the deletions specified, as well as a xylose-inducible copy of *loaP* integrated into the non-essential *thrC* locus. (C) Equilibrium binding curves obtained via DRaCALA wherein increasing amounts of LoaP were combined with radiolabeled synthetic DNAs. The normalized fraction bound is shown with error bars representing the standard deviation from the mean. Binding affinities were calculated from at least three experimental replicates.

LoaP enacts readthrough at additional *dfn* terminators

Prior transcriptomic data¹⁴⁸ suggest the LoaP-dependent readthrough of intrinsic terminators located within coding regions of the *dfn* gene cluster. In an effort to characterize this antitermination activity in vivo, these terminator sequences (*dfnE* and *dfnI*) were substituted in place of the *dfnA* terminator hairpin in the aforementioned *dfn* leader-*yfp* reporter construct. As done previously, *loaP* expression was induced from a non-essential locus, and the resulting fluorescence intensity was averaged across the cell population. In support of prior findings, LoaP was shown to enact readthrough at each of the intrinsic terminators present in the *dfn* gene cluster – *dfnA*, *dfnE* and *dfnI* – in vivo (Figure 4-12A-B). However, this LoaP-dependent

readthrough effect is notably weaker at the final terminator of the gene cluster, *dfnI*, which permits higher background fluorescence in the absence of LoaP. When these same terminator hairpins were substituted in DNA templates for in vitro transcription, an increase in termination efficiency was observed at both *dfnA* and *dfnE* following the addition of purified NusA, while templates possessing the *dfnI* terminator did not demonstrate this effect (**Figure 4-12C**).

Moreover, LoaP is unable to enact readthrough activity at either *dfnA* or *dfnE* when added alone in vitro (data not shown). Rather, LoaP requires the presence of NusA at these first two terminators to promote downstream readthrough in vitro, further supporting the importance of repositioning NusA at the RNA exit channel for antitermination activity. However, LoaP alone decreases the termination efficiency of – or allows readthrough at – *dfnI* in vitro (**Figure 4-12D**).

Taken together, these data suggest that strict regulatory control over more distal genes in the *dfn* gene cluster may not be as pertinent. Perhaps by the time the TEC has reached the *dfnI* hairpin, the cellular commitment to express the full length *dfn* mRNA has already been made and the weaker *dfnI* terminator acts instead to tailor the relative amount of full-length transcript.

Regardless, in vivo data also suggest that LoaP-dependent readthrough at *dfnI* still requires the presence of the *dfn* hairpin within the 5' UTR (**Figure 4-13**), further supporting the role of the RNA element in antitermination activity, potentially as a LoaP recruitment element or a functional component of the molecular antitermination mechanism. The ability of LoaP to enact readthrough outside of the antagonization of NusA presents the possibility of a yet simpler antitermination mechanism in which the association of RNA-bound LoaP with RNAP is enough to enact downstream readthrough at certain terminators, yet further investigation is required.

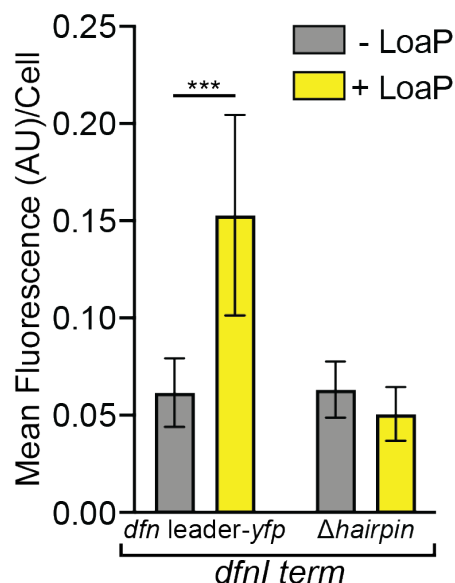


Figure 4-13. LoaP antitermination at the *dfnI* terminator requires the *dfn* hairpin Bar graph representing the average fluorescence intensity per cell measured from *B. subtilis* expressing a constitutive P_{const} -*dfn* leader-*yfp* reporter, or P_{const} -*dfn* leader Δ hairpin -*yfp*, containing the *dfnI* terminator sequence directly upstream of *yfp*, as well as a xylose-inducible copy of *loaP* integrated into the non-essential *thrC* locus. Statistical significance was determined following unpaired t tests. Three asterisks (***) denotes a p-value < 0.001.

Discussion

Although regulatory events occur at every stage of the bacterial transcription cycle, knowledge of regulation during elongation and termination is particularly lacking. Directly following the elongation of a nascent transcript, termination occurs when the transcription elongation complex (TEC) is forced to dissociate after encountering either RNA-bound Rho or an intrinsic terminator hairpin, both of which encourage separation of RNAP from the DNA template and nascent RNA^{6,23,40}. Rho-dependent termination is responsible for most transcription termination events in Gram-negative organisms, in which *rho* also tends to be an essential gene^{71,182}. Conversely, Gram-positive organisms, such as *B. subtilis*, seldom use Rho-dependent

termination and instead favor the utilization of intrinsic terminators, which are often enhanced by additional transcription factors⁸⁸. One such factor, NusA, enhances both transcription pausing and termination at specific sites in the genome^{44,104,188,194}. Once the TEC escapes the promoter region, NusA forms transient interactions at the RNA exit tunnel through contact with the α CTD and β flap, as well as nascent RNA, promoting pausing and terminator formation^{44,188,194,195}.

However, transcription termination can be overcome by changes to the TEC. Specifically, processive antitermination mechanisms occur when the TEC is modified to resist termination⁸⁹. Only a small number of antitermination mechanisms have been well characterized and the extent of antitermination usage across bacterial genomes is not known. The antitermination systems that have been investigated are mostly *E. coli* regulatory mechanisms that act primarily to prevent Rho termination, where the *E. coli* NusG protein is a key component. The *E. coli* NusG N-terminal domain uses a unique, highly conserved NGN domain to make contacts to RNAP near the non-template DNA, while using a conserved beta sheet motif within the CTD to mediate protein-protein interactions with either Rho or NusE (S10)³⁶. The *E. coli* NusG-NusE (S10) complex participates in antitermination complexes, which blocks the mutually exclusive interaction of NusG with Rho^{33,36,57,105}. Alternatively, *E. coli* TEC-bound NusG can associate with ribosome-bound NusE (S10) as a key part of the bridge that couples the *E. coli* TEC with a leading ribosome^{35,37,38,61}. Rho cannot access NusG when transcription and translation are coupled; therefore, transcription-translation coupling acts as an antitermination mechanism as the TEC synthesizes open reading frames. Once the TEC-ribosome encounters a stop codon, translation is uncoupled, and Rho can gain access to NusG^{36,57,105}.

In this manner, the *E. coli* NusG paralog RfaH also promotes transcription-translation coupling^{76,123} although the RfaH CTD contains mutations that render it unable to associate with Rho, thereby specializing the protein for antitermination activity¹³⁷. The investigations of antitermination mechanisms in Gram-negative bacteria leads to the conclusion that they have evolved as a means of avoiding Rho termination. However, this leaves a gap in knowledge pertaining to antitermination mechanisms in bacteria that do not rely heavily on Rho or that lack it entirely.

It has been recently suggested that most Gram-positive bacteria are unlikely to couple transcription with translation⁷². Furthermore, the Rho termination factor is absent in some Gram-positive bacteria, such as streptococci, and is non-essential in other bacteria, such as *B. subtilis*¹⁸². These observations directly challenge the assumptions that have been derived from investigations of antitermination mechanisms in Gram-negative bacteria.

In lieu of global Rho termination, Gram-positive bacteria appear to rely more heavily on intrinsic terminator hairpin elements. Therefore, it can be assumed that Gram-positive antitermination complexes have evolved molecular strategies that specifically interfere with the termination mechanisms of intrinsic terminator hairpins. Indeed, of the three antitermination mechanisms discovered in Gram-positive bacteria, all have been shown to promote the readthrough of intrinsic terminator hairpins^{122,148,196}, yet the molecular basis of their antitermination activity has not yet been discovered. To address this gap in knowledge, this study investigates the antitermination mechanism employed by *Bacillus velezensis* LoaP, a NusG paralog encoded by many Gram-positive bacteria.

In a previous study¹⁴⁸, transcriptomic data suggested that *B. velezensis* LoaP influences two polyketide synthesis (PKS) gene clusters. These gene clusters, which produce the secondary

metabolites difficidin (*dfn*) and macrolactin (*mln*), are unusually long and adorned with several intrinsic terminator hairpins that likely evolved as an additional layer of regulatory control against the expression of energetically costly operons. Additional analyses revealed a UNCG-tetraloop RNA hairpin element (*dfn*) present in the leader regions of both gene clusters¹⁴⁹. However, the hairpin structure was not identified within the leader region of the third PKS gene cluster expressed by *B. velezensis*, which is responsible for the production of bacilleane (*bae*)¹⁸⁶. In this study, proteomic analyses were used to search for proteins that were increased upon LoaP expression and that decreased in a *loaP* deletion strain. This revealed that peptides unique to proteins from the *dfn* and *mln* gene clusters were highly dependent on *loaP* expression, with most being completely undetectable in the absence of LoaP. However, this was not the case with peptides unique to the *bae* gene cluster, whose abundance did not change with respect to *loaP* expression. These data demonstrated that *B. velezensis* LoaP acts primarily on the *dfn* and *mln* operons. The fact that both operons contain a nearly identical UNCG-tetraloop hairpin sequence within their 5' leader regions suggests a mechanistic role for this RNA determinant. Furthermore, each 5' leader region contains a putative intrinsic terminator hairpin, suggesting that these terminator sites are likely to be key points of regulatory control. This observation also indicates that all molecular determinants required for LoaP antitermination are contained within the leader regions of these target regulons.

It was reasoned that a DNA template encompassing the *dfn* leader region could be used to reconstitute LoaP antitermination in vitro. Using synchronized transcription reactions, in which only a single TEC could operate on a given DNA template, it was found that the *dfn* leader region contained a functional terminator site that was enhanced by the presence of NusA. Concurrent addition of NusG had no effect on NusA-mediated termination; however, the

addition of LoaP significantly decreased NusA-mediated termination. From this, it was concluded that LoaP antitermination activity derives from its ability to directly antagonize NusA, a function which is likely to be broadly relevant for the subclass of LoaP regulatory proteins. From these data, it is speculated that NusA manipulates the exit tunnel of RNAP to better support the formation of the nascent hairpin, a function which has been previously characterized of NusA_{Ec}^{44,104}. It is proposed that LoaP disrupts NusA function by potentially repositioning NusA away from the exit tunnel, without the need for additional factors.

Earlier binding assays found that LoaP proteins purified from multiple species all bind the *dfn* UNCG-type hairpin with sub-micromolar affinity and high specificity¹⁴⁹, using both N- and C-terminal domains¹⁵². To investigate the importance of this ribonucleoprotein complex, the hairpin was deleted from a *dfn* leader region-*yfp* reporter fusion and found that it resulted in a loss of antitermination activity as compared to a wild-type *dfn* sequence. Moreover, site-directed mutagenesis of RNA-binding residues¹⁵² resulted in a significant decrease in the cellular production of difficidin and macrolactin, demonstrating that the RNA-binding activity of LoaP is essential for antibiotic production.

LoaP is the only known bacterial NusG-like protein that specifically associates with an RNA element. Although high-resolution structural data of the LoaP antitermination complex are needed to unambiguously determine the role of the ribonucleoprotein complex, it is speculated that it may play a role in the recruitment of LoaP to the TEC. Both NusG and RfaH rely on their NTDs to interact near or to sequences of the non-template DNA strand^{18,129}. More specifically, RfaH takes advantage of an RNAP pause brought about by a DNA hairpin in the non-template strand, '*ops*'^{129,135}. Upon forming in the transcription bubble, this sequence stalls the TEC, giving RfaH time to contact its DNA recognition sequence and thereby activating an otherwise

autoinhibited RfaH¹²³. In contrast, the *dfn* hairpin results in no discernible TEC pause, implying that LoaP binds the *dfn* hairpin after it has left the TEC exit channel. This, in turn, suggests that the LoaP-hairpin complex associates with the TEC as it actively polymerizes away from the LoaP recruitment site. This might explain why a downstream transcriptional pause site assists LoaP antitermination both in vivo and in vitro; this study speculates that the downstream pausing event gives the LoaP-hairpin complex time to associate with the TEC. More specifically, it is posited that the stalled TEC may be engaged by LoaP bound to the flexible, nascent RNA, and that this interaction may be further driven by an increase in local LoaP concentration following recruitment to the RNA with sub-micromolar equilibrium binding affinity. That LoaP may be capable of binding the TEC and nascent RNA simultaneously, further suggests that LoaP may maintain contacts to the *dfn* hairpin in a way that is functionally meaningful for antitermination activity. Additionally, LoaP interferes with the ability of NusA to lengthen this pause event, further supporting the model that LoaP, by potentially disrupting the position of NusA at the exit tunnel, antagonizes NusA activity as part of the TEC. Together, these data reveal fundamental features of the LoaP mechanism (**Figure 4-14**) and provide key insights into the antitermination strategies used by Gram-positive bacteria. Furthermore, *loaP* genes are frequently located adjacent to secondary metabolite biosynthesis operons¹⁴⁸; therefore, it is likely that secondary metabolite pathways are frequent targets of LoaP regulons. Uncovering the molecular basis of LoaP regulation may aid in the discovery and production of natural products and other secondary metabolites. Therefore, knowledge of the LoaP mechanism may be used to develop novel genetic tools that improve heterologous expression of antibiotics and other natural products.

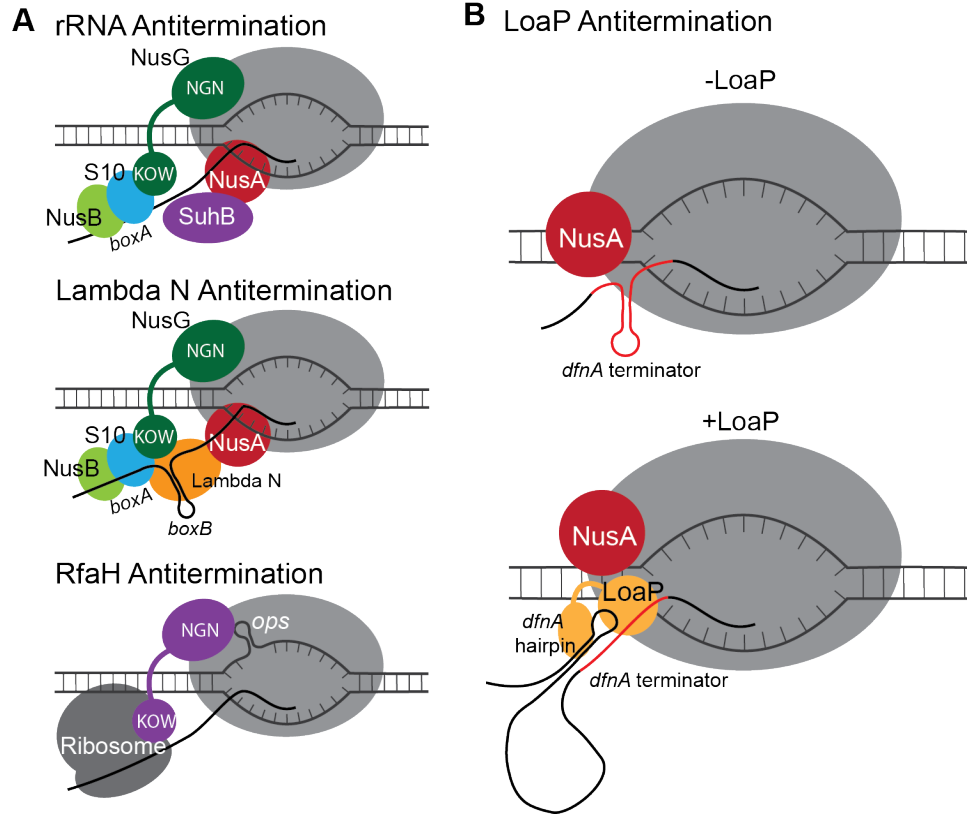


Figure 4-14. The molecular mechanism for LoaP antitermination is novel (A) Schematics of previously characterized *E. coli* antitermination mechanisms. Top: rRNA Middle: λ N Bottom: RfaH. (B) Schematic of LoaP antitermination. Top: NusA enhances intrinsic termination at *dfnA* in the absence of LoaP. Bottom: LoaP antagonizes NusA-enhanced intrinsic termination following the recognition of a UNCG RNA tetraloop in the 5' UTR of target gene clusters.

Materials and Methods

Strain Construction

All sequences and strains utilized in this study are summarized in the Appendices. To construct *E. coli* strains to overexpress and purify the proteins used in this study, the coding sequences for *loaP_{Bv}*, *nusA_{Bv}*, *nusG_{Bv}*, and *sigA_{Bs}* were PCR-amplified from genomic DNA (*B. velezensis* or *B. subtilis*) or purchased from Integrated DNA Technologies (IDT) and cloned via Gibson assembly into a plasmid (pAmr30) containing an IPTG-inducible N-terminal

decahistidine tag fused to a bdSENP-cleavable SUMO tag. Plasmids were transformed into *E. coli* XL10-Gold (Agilent) for plasmid replication and *E. coli* T7 Express (New England BioLabs) for overexpression. A *B. subtilis* strain lacking *nusG* and possessing a His-tagged RNAP was assembled by transforming strain 1A813 (*B. subtilis* 168, *rpoC*::6xHis tag, BGSC) with genomic DNA isolated from strain BKE01010 (*B. subtilis* 168 Δ *nusG*::*erm*, BGSC). In vivo reporter strains were constructed by PCR-amplifying the region corresponding to the *dfn* leader region and subcloning this sequence, followed by *yfp*, into the vector pDG1662, which integrates into the *amyE* non-essential gene locus. Mutations in this leader region were made via Site-Directed Mutagenesis. A plasmid that integrates a xylose-inducible *loaP_{Bv}* into the *thrC* non-essential gene locus (pJG108) was used to control LoaP production, whereas a multi-copy, conjugatable plasmid (pEP011) was used as the vector for a xylose-inducible *nusA_{Bv}*.

Protein Expression and Purification

Detailed in Chapter 3.

Synchronized In Vitro Transcription

Transcription elongation complexes were established by combining 100 nM linear DNA templates with 500 nM *B. subtilis* RNAP apoenzyme in transcription buffer (20 mM Tris-HCl pH 8.0, 40 mM KCl, 5 mM MgCl₂, and 1 mM β -mercaptoethanol). Transcription reactions were supplemented with 500 nM purified *B. subtilis* sigma factor A (σ A) and 100 μ M ApG to assist in initiation. These transcription mixtures included ATP and GTP at 5 μ M and UTP at 2.5 μ M, but lacked CTP. They were incubated for 15 min at 37 °C, resulting in elongation complexes being halted on the templates at C29 (**Figure 4-4C**). Elongation was stimulated by the addition of

nucleotides (ATP, CTP, GTP, and UTP at 200 μ M) and heparin (100 μ g/mL) to limit re-initiation of free polymerases. Transcription progress was tracked via 32 P derived from the incorporation of [α - 32 P]UTP (3,000 Ci/mmol). After 5 min, additional NTPs (Chase) were added (250 μ M each). When included, transcription factors NusA, NusG, and LoaP were added during the formation of the halted complex to final concentrations indicated in the figures. Samples were removed in the time frames specified in the figures and quenched in Stop Buffer (10 M Urea, 20 mM EDTA, 45 mM Tris-borate pH 8.3) prior to phenol extraction. The samples were heated for 2 min at 90 °C and resolved by 6% denaturing PAGE containing 1 $\times$ TBE and 4 M urea. RNA products were visualized and quantified using a Cytiva Amersham Typhoon PhosphorImaging System, ImageJ Software, and Microsoft Excel. Termination percentages were obtained by determining the signal intensity of bands corresponding to terminated products as a ratio of overall lane intensity, while normalizing for the contribution by background signal of individual lanes. Half-life calculations of the *dfn* leader pause site were performed as previously described¹⁹⁷. All synchronized transcription data were collected from at least three technical replicates.

Differential Radial Capillary Action of Ligand Assays (DRaCALAs)

All RNA and DNA species used in the in vitro binding assays are summarized in the Appendices and were purchased from IDT and 5'-radiolabeled with [γ - 32 P] ATP using T4 Polynucleotide Kinase (T4PNK; New England Biolabs). Longer nucleic acid species (>17 bp) were purified using a Zymo RNA Concentrator and Clean-up Kit. Binding reactions were set up in 96-well plates by incubating 2.5 nM radiolabeled RNA with increasing amounts of LoaP in Binding Buffer (50 mM NaH₂PO₄ pH 7.2, 100 mM NaCl, 0.1 mM EDTA, 1 mM MgCl₂,

100 $\mu\text{g/mL}$ BSA, 25 $\mu\text{g/mL}$ total yeast RNA) for 45 min at room temperature in the dark. 2 μL aliquots were spotted onto nitrocellulose membranes using a fixed replicator pin tool and allowed to air dry for 20 min. Nitrocellulose membranes were exposed to a phosphor screen for 20 min, visualized using an Amersham Typhoon phosphorimager, and quantified using ImageJ and GraphPad Prism software. All binding data were collected from at least three technical replicates.

Fluorescent Reporters

2xYT medium (containing the appropriate antibiotic) was inoculated from single colonies and incubated at 37 °C with shaking, overnight. These cultures were centrifuged at $5,000 \times g$ for 5 min and resuspended in fresh medium without antibiotics. Cultures were then diluted to an OD_{600} of 0.1 and incubated at 37 °C with shaking until reaching an OD_{600} of 0.5. D-xylose (1% w/v final) was added to the cultures, and the cells were incubated for an additional 3 hours. The cultures were then centrifuged at $5,000 \times g$ for 5 min and resuspended in 1x PBS. This step was repeated, and aliquots were placed on 1.5% low-melting-point agarose pads and allowed to dry for 10 min before being inverted on a glass coverslip. The cells were imaged at room temperature using a Zeiss Axio-Observer Z1 inverted fluorescence microscope equipped with a YFP filter cube and enclosed in a temperature-controlled chamber. Fluorescence intensity per cell was quantified using ImageJ software. Each mean fluorescence value was determined across a minimum of 30 cells.

Mass Spectrometry Analysis

Proteomics Sample Preparation and Normalization

To quantify the effect of LoaP on the proteome of *Bacillus velezensis* FZB42, WT, $\Delta loeP$, *loeP* complementation and *loeP* overexpression strains were prepared. At least four biological replicates of each strain were cultured in 5 mL LB, containing 0.5% D-xylose (w/v) for *loeP* expression, to an OD₆₀₀ of 0.8 at 37 °C. Dried protein pellets were resuspended in 500 μ L Lysis Buffer (5% sodium dodecyl sulfate, 50 mM triethylammonium bicarbonate) and sonicated with a Branson probe at 50% amplitude for 25 s, with 5 s on/off cycles. The protein concentration of each sample was assessed using a BCA Protein Assay kit (Pierce) following the manufacturer's protocol. A total of 300 μ g of each sample was used for proteomic digestion. Dithiothreitol (DTT) was added to each aliquot to a final concentration of 0.2 M, followed by incubation at 95 °C for 10 min. After cooling to room temperature, iodoacetamide was added to each sample to a final concentration of 0.1 M. Samples were then incubated in the dark at room temperature for 30 min. Formic acid (15%) was added to each sample at a final concentration of 1.5%, and S-Trap Buffer (90% methanol, 100 mM triethylammonium bicarbonate) was added to each sample at 6x the total reaction volume. The samples were loaded into S-Trap Mini columns (Protifi), and loading and wash steps were performed according to the manufacturer's protocol. Finally, 125 μ L of 50 mM triethylammonium bicarbonate containing Trypsin Lys-C (at a 1:20 enzyme-to-protein ratio; Promega) was added to each S-Trap Micro column filter. The columns were wrapped with parafilm to prevent evaporation and placed in a stationary incubator overnight at 37 °C. Elution was performed the following morning according to the manufacturer's instructions. Eluted peptides were dried and resuspended in water and 30% acetonitrile. Peptide concentrations of each sample were determined using a Colorimetric Peptide

Assay (Pierce) following the manufacturer's protocol. Peptides were dried once more, then resuspended in 5% acetonitrile and 0.1% formic acid in water to a final peptide concentration of 1 $\mu\text{g}/\mu\text{L}$ for analysis via LC/MS/MS. Proteomics sample preparation and normalization done in collaboration with Conor Jenkins.

Proteomics Data Acquisition

Each sample was analyzed using a Thermo Fisher Orbitrap Fusion mass spectrometer coupled to a Dionex Ultimate 3000 UHPLC (Thermo Fisher Scientific). Injections of each sample (2 μL) were first pre-concentrated on a reverse-phase trapping column (PepMap 300 $\mu\text{m} \times 5 \text{ mm}$ C18, 100 $\text{\AA}$, Thermo Fisher) and then resolved on a 75 $\mu\text{m} \times 500 \text{ mm}$ EASY-Spray column packed with PepMap RSLC C18, 2 μm , 100 $\text{\AA}$ particles (Thermo Fisher) using a 190 min multistep gradient [0–150 min: 6%–35% B, 150–158 min: 35%–60% B, 158–161 min: 60%–90% B, 161–171 min: 90% B hold, 171–172 min: 90–96% B, and 172–182 min: 6% B hold]. For the gradient, buffer A was 100% H_2O /0.1% formic acid, and buffer B was 98% acetonitrile/2% H_2O /0.1% formic acid. MS1 scans were acquired using the Orbitrap at a resolution of 120,000 with a scan range of m/z 350–1500 in profile mode. The top 15 precursors were selected for MS2 data-dependent fragmentation. MS2 spectra was acquired using the ion trap with the scan rate set to 'Turbo.' The minimum signal required to trigger a data-dependent scan was 5000.

Monoisotopic precursor selector was set to "On", and charge state filter was set to include charge states 2–7. High-energy C-trap dissociation (CID) was used to generate MS2 spectra with a collision energy of 65%. The AGC targets were set to 200,000 for MS1 and 10,000 for MS2. The maximum accumulation time for MS1 and MS2 was 50 ms. Dynamic exclusion was set for 70 s

with a 10 ppm mass window. Proteomics data acquisition done in collaboration with Conor Jenkins.

Proteomics Data Analysis

Including database discovery, relative abundance plots, cluster map, principal components analysis (PCA) plot, and volcano plot generation, untargeted proteomics data analysis was performed with Thermo Fisher Scientific Proteome Discoverer (ver. 2.5.0.400). A database search was performed against the Uniprot canonical and isoform FASTA (Taxon ID: 1390), downloaded from Uniprot (March 06, 2023), and searched using Sequest. The search parameters included an MS1 tolerance of 10 ppm and MS2 tolerance of 0.6 Da. A variable protein N-terminal acetylation was also included in the search, along with oxidation of methionines and static carbamidomethylation of cystines. FDR control was achieved utilizing the Percolator node. Post-translational modification (PTM) localization was performed with IMP-ptmRS. Label-free quantitation was performed using the Minora feature-detection node. Identified proteins from the analysis were then filtered by requiring at least one unique peptide, more than 1 PSM and Protein FDR at a confidence level (0.01 FDR). Proteomics data analysis done in collaboration with done in collaboration with Conor Jenkins.

Metabolomic Extraction

Complementation strains encoding *loaP* with amino acid mutations thought to inhibit the binding to the *dfn* hairpin (E40A, R41A, and I143A) were processed for metabolomic analysis. A strain with a *LoaP* K128A mutation was used as a negative control, as the mutation previously showed no observable change in RNA-binding affinity in vitro. Four biological replicates of each

strain were grown overnight at 37 °C with shaking, in 5 mL cultures of LB. Samples were then sub-cultured in 100 mL of LB and grown at 37 °C with shaking to an OD₆₀₀ of 0.5. D-xylose was added to the cultures at a final concentration of 0.5% (w/v) to induce *loaP* gene expression. The samples were then grown for 2 hours at 37 °C with shaking. Supernatants were collected by centrifugation, followed by the addition of 8 g of Amberlite XAD-16 resin. The samples were then shaken overnight at 4 °C. The resin was washed with 50 mL of LCMS grade water, followed by elution in 8 mL of LCMS grade methanol. The samples were dried to completion and resuspended in 200 µL of 90:10 H₂O:ACN with 0.1% FA. Metabolomic extraction done in collaboration with Conor Jenkins.

Metabolomic Data Acquisition and Analysis

The metabolomic samples were run on a Thermo Scientific Vanquish HPLC and Q Exactive Plus Hybrid Quadrupole-Orbitrap Mass Spectrometer. 10 µL of sample was injected onto a CORTECS T3 Column (120Å, 2.7 µm, 2.1 mm X 150 mm) and resolved on a 20 min gradient ranging from 0% B (ACN with 0.1% Formic Acid) and 100% A (H₂O with 0.1% Formic Acid) to 95% B and 5% A. Sample was injected twice for both positive and negative mode acquisitions. Metabolomic data analysis was initially performed using the FisH scoring metric in Compound Discoverer to obtain in silico fragmentation spectra for each analyte of interest. Data was then processed in Skyline with an MS1 ppm tolerance of 5 ppm and 0.02 fragment tolerance. The area under the curve was calculated using a built-in peak-finding tool. Metabolomic data acquisition and analysis done in collaboration with Conor Jenkins.

Chapter 5: Future Directions

Introduction

Universally, gene expression relies on the process of transcription, during which functional RNA molecules are synthesized based on the genetic elements of a DNA template. While the stages of transcription – initiation², elongation^{4,5} and termination⁶ – are well characterized, the longest stage, transcription elongation, is perhaps the least understood. This is likely because of the abundant diversity of regulatory mechanisms that control the transcription elongation complex.

The absence of a membrane-bound nucleus in bacterial species permits the transcription and translation processes to occur simultaneously and in close proximity⁶⁰. As a result, the TEC and a leading ribosome are capable of interacting on a nascent transcript. This tethering of the TEC to the leading ribosome results in the coupling of transcription and translation^{38,198}. Transcription-translation coupling is well-characterized in *E. coli*, which employs coupling, enhanced by the universal transcription factor NusG, as a means to prevent Rho termination⁶⁶. While these findings have set a precedent for antitermination mechanisms, particularly by NusG family proteins, current investigations in *Bacillus* species challenge this paradigm. A recent study supports that in *B. subtilis*, the Gram-positive model organism, transcription and translation are not coupled⁷². Instead, the in vivo rate of transcription is twice that of translation, resulting in a “runaway” transcription apparatus that outpaces the leading ribosome. These findings seem to be further evidenced by reduced or absent Rho activity across Gram-positive species⁴², diminished distance between stop codons and termination sites⁷², and the ubiquity of *cis*-acting regulatory elements in the leader regions of these organisms.

Prior work from the Winkler laboratory identified a novel subfamily of NusG homologs, LoaP, found across Bacillota, Actinomycetota and Spirochaetota¹⁴⁸. LoaP is often encoded adjacent to exceptionally long operons that are responsible for the biosynthesis of secondary metabolites and polysaccharides. In *B. velezensis*, the Gram-positive organism in which LoaP was discovered, LoaP was shown to promote readthrough at intrinsic termination sites within two polyketide synthesis operons controlling the production of the antibiotics diffcidin and macrolactin¹⁴⁸. Additionally, LoaP was found to associate with a UNCG RNA hairpin element (*dfn*) identified in the 5' UTR of these gene clusters with high affinity¹⁴⁹. Since its initial discovery, knowledge pertaining to the molecular mechanism employed by LoaP holds significant importance in demonstrating how the functions of NusG proteins may differ in organisms that do not couple transcription and translation. Given that *B. velezensis* is very closely related to *B. subtilis*, it is likely that the LoaP antitermination mechanism may broadly reflect the types of regulatory mechanisms employed by Gram-positive organisms.

All known antitermination mechanisms involve sequence-specific interactions with nucleic acids, often allowing multiple factors to associate with the TEC and regulate downstream termination events^{85,89}. Both the absence of transcription-translation coupling and the dispensability of Rho in Gram-positive organisms⁴² suggest that methods for regulating transcription elongation in these species may rely on direct contacts to the nascent RNA transcript.

The work detailed in this thesis was set out to determine the molecular mechanism by which LoaP enacts antitermination at target regulons. The co-crystal structure of the LoaP-*dfn* complex determined that LoaP interacts with its RNA ligand, the *dfn* hairpin, via a novel RNA-binding mechanism¹⁵². Although LoaP employs both terminal domains to contact the RNA

hairpin, structural analyses suggest that the LoaP NGN domain is capable of simultaneously binding the TEC, as expected for NusG proteins. Additionally, the proteomic and metabolomic data described in this dissertation further strengthened the hypothesis that LoaP selectively regulates operons, based on the presence of this RNA hairpin element in the corresponding 5' leader regions. The efforts described herein then turned to the reconstitution of LoaP antitermination in vitro as a means for identifying the required nucleic acid and protein determinants. The use of in vitro synchronized transcription methods determined that a NusA-dependent intrinsic terminator precedes the *dfn* operon and functions as an initial regulatory checkpoint, impeding the downstream expression of an otherwise costly operon. LoaP was found to antagonize this NusA control of the *dfn* operon without the need for additional protein factors, although it relies on the presence of the *dfn* hairpin element to suppress NusA activity. Additionally, a measurable transcription pause was found downstream of the *dfn* hairpin, but upstream of the *dfnA* terminator. This pause site enhances LoaP antitermination activity. Together, these data suggest a molecular mechanism wherein LoaP recognizes its cognate RNA hairpin ligand and, upon encountering the transcription pause site, is allotted sufficient time to interface concurrently with the TEC.

The molecular mechanism employed by LoaP is the first NusG specialized paralog to be characterized in a Gram-positive organism and represents a mode of antitermination so far unprecedented by NusG proteins. By antagonizing intrinsic termination events following recognition of a novel RNA determinant, and without the need for additional protein factors, LoaP expands the mechanistic diversity of NusG and NusG-like proteins. This diversity may offer insights into the regulatory mechanisms employed by other bacterial species during

transcription elongation. In addition, investigation of LoaP regulation will aid in the development of new methods for the discovery and production of natural products.

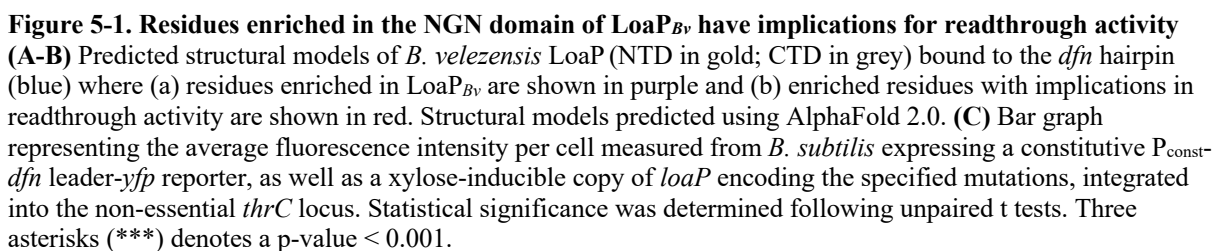
Conserved vs. enriched residues in LoaP proteins

The novelty of the LoaP antitermination mechanism, primarily the RNA-binding activity conferred by the LoaP NGN domain, suggests the presence of residues specific to LoaP in an N-terminal domain that is otherwise structurally comparable to those of NusG and RfaH^{18,37,152}. To determine conserved residues in the LoaP NTD which may confer unique abilities – likely in RNA recognition, but also potentially the recognition of non-template DNA, RNAP or additional cellular factors (e.g. NusA) – the Winkler laboratory performed a comparative sequence analysis to identify NTD residues enriched in LoaP proteins¹⁴⁹. To accomplish this, protein sequences annotated as NusG or LoaP were extracted from the NCBI genomic database and, after filtering for redundancy (> 60% sequence identity), were used to produce two distinct sequence clusters. These clusters were then used to generate a multiple sequence alignment in which NusG and LoaP sequences encoded by the same organism were compared.

By comparing the NGN domains of LoaP proteins with those of NusG proteins, residues that are heavily enriched in LoaP but infrequent in NusG were identified (**Figure 5-1A**). Given the evolutionary relation of LoaP as a specialized paralog of NusG, these enriched LoaP residues are likely to confer significant functions.

To test this hypothesis, residues enriched in the LoaP NTD were individually substituted for alanine and the resulting LoaP mutants were overexpressed from the non-essential *thrC* locus in a *B. subtilis* strain containing an *amyE*-integrated P_{const}-*dfn* leader-*yfp* reporter. The resulting single cell fluorescence intensity in the presence and absence of LoaP was calculated for each. It

was determined that only six of the tested LoaP mutants – those with alanine substitutions at Glu9, Phe24, Lys35, Lys42, Thr62 and Ile81 – exhibited a loss in readthrough activity suggesting significant function for LoaP antitermination (**Figure 5-1B-C**). Upon closer investigation of these residues, only one – Lys42 – is likely to engage with the *dfn* hairpin. In fact, this *B. velezensis* residue is likely homologous to Lys43 from *T. pseudethanolicus* LoaP, which was determined (Chapter 2) to interface with the *dfn* hairpin via a hydrogen bond to a NBO of U13. Notably, the rest of these functional, enriched residues show putative intramolecular interactions suggesting roles in maintaining LoaP protein structure. Somewhat surprisingly, none of these functional, enriched residues are located along the homologous, α -helical binding surfaces used by NusG and RfaH to contact RNAP or ntDNA sequences¹⁸. This is indicative of the possibility that LoaP does not employ novel means for either of these functions, but that LoaP may employ enriched residues in order to maintain a *dfn*-bound structure.



Heterologous express of LoaP with potential for drug discovery

The LoaP family of NusG paralogs is widespread in Bacillota, Actinomycetota and Spirochaetota, across which it is often associated with long, biosynthetic gene clusters responsible for the production of secondary metabolites¹⁴⁸. Although often encoded adjacent to these regulons, many LoaP proteins are encoded distally from the gene clusters that they regulate. With this, it can be expected that LoaP proteins in organisms aside from *B. velezensis* regulate the production of additional natural products, many of which may be undiscovered.

In fact, a study which set out to overexpress a secondary NusG protein encoded by *Clostridium cellulolyticum* discovered an antibiotic polythioamide compound, closthioamide¹⁹⁹. This NusG protein was later determined to be a LoaP, and the production of an antibiotic compound following its overexpression indicates the potential in natural product discovery following the characterization of LoaP proteins. The gene cluster responsible for the production of closthioamide (*cta*) was identified, however introduction of this gene cluster into *E. coli* only results in the production of intermediate compounds, and not closthioamide²⁰⁰. Instead, full-length expression of the *cta* gene cluster relies on LoaP^{201,202}. This finding indicates an application of LoaP proteins for the heterologous expression of natural products which are often encoded by long, energetically costly operons that require full length expression for functional metabolites.

In an effort to begin characterizing additional LoaP proteins, the Winkler laboratory has developed a list of candidate LoaP sequences from other organisms. Focusing initially on LoaP proteins present in Bacillota, sequences for LoaPs encoded by *Bacillus thuringiensis*, *Brevibacillus laterosporus* and *Paenibacillus polymyxa* were cloned into a xylose-inducible multi-copy plasmid. When these constructs were introduced into *B. subtilis* and cells were grown

in the presence of D-xylose, it was determined that the induction of heterologous LoaP proteins does not detrimentally affect cell growth (**Figure 5-2**), suggesting that overexpressing LoaP proteins, both in native and heterologous backgrounds, may be a viable path for synthetic biology tools and natural product production.

Polyketide synthesis (PKS) pathways have been identified across the genomes of *B. thuringiensis*^{203,204}, *B. laterosporus*²⁰⁵ and *P. polymyxa*²⁰⁶, yet the mechanisms which regulate the expression of these gene clusters are largely unknown. That these species each encode a LoaP suggests regulation of at least one or more of these operons following *loaP* expression. In fact, a prior study¹⁴⁹ identified RNA hairpin elements similar to *dfn* upstream of *loaP*-associated operons in *B. thuringiensis* and *B. laterosporus*; however, a *dfn*-like hairpin element has not yet been identified in *P. polymyxa*. Unsurprisingly, these proteins show incredible structural similarities (**Figure 5-3A-B**), with the global folds of their N- and C-terminal domains representing canonical NusG protein motifs¹⁵. With this, and the prevalence of PKS pathways in the genomes which encode them, it is likely that these proteins exhibit similar functions to *B. velezensis* LoaP.

To test this, the aforementioned overexpression constructs were used to express LoaP proteins in a *B. subtilis* background harboring a previously described (Chapter 3) *dfn* leader-*yfp* reporter construct. The resulting single cell fluorescence intensity in the presence and absence of each LoaP was determined (**Figure 5-3C**). The expression of *B. thuringiensis* and *B. laterosporus* LoaP proteins resulted in readthrough comparable to that by *B. velezensis* LoaP, while LoaP from *P. polymyxa* failed to enact antitermination activity on the *dfn* leader region. While unexpected, this finding has broader implications for the mechanistic diversity of LoaP proteins. Notably, both *B. thuringiensis* and *B. laterosporus* are more closely related to *B.*

velezensis than *P. polymyxa* and possess known *dfn* hairpin elements. Perhaps, LoaP proteins in more divergent organisms have evolved mechanisms for recognizing RNA elements which differ significantly from the *dfn* hairpin. That RNA-binding proteins tend to be more diverse in their modes of recognition than proteins that recognize DNA¹⁶³, suggests the potential for broad mechanistic diversity amongst LoaP proteins. Finally, these data support the heterologous expression of LoaP proteins for successful readthrough at target regulons, illustrating the potential in LoaP proteins as a tool for the discovery and production of novel natural products.

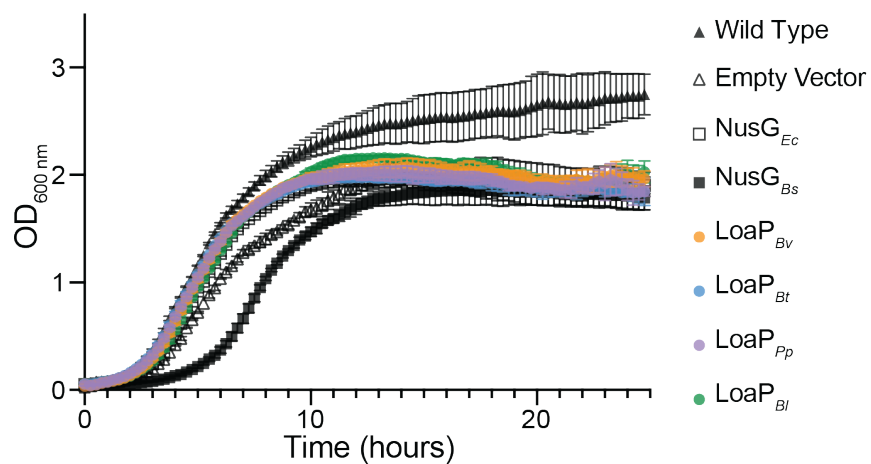


Figure 5-2. Heterologous overexpression of LoaP proteins does not impact cellular fitness Growth of *B. subtilis* strains overexpressing *nusG* and *loaP* sequences from other species (*Ec*, *E. coli*; *Bs*, *B. subtilis*; *Bv*, *B. velezensis*; *Bt*, *B. thuringiensis*; *Pp*, *P. polymyxa*; *Bl*, *B. laterosporus*). Cellular growth was carried out in rich media at 37 °C in the presence of 1% (w/v) D-xylose.

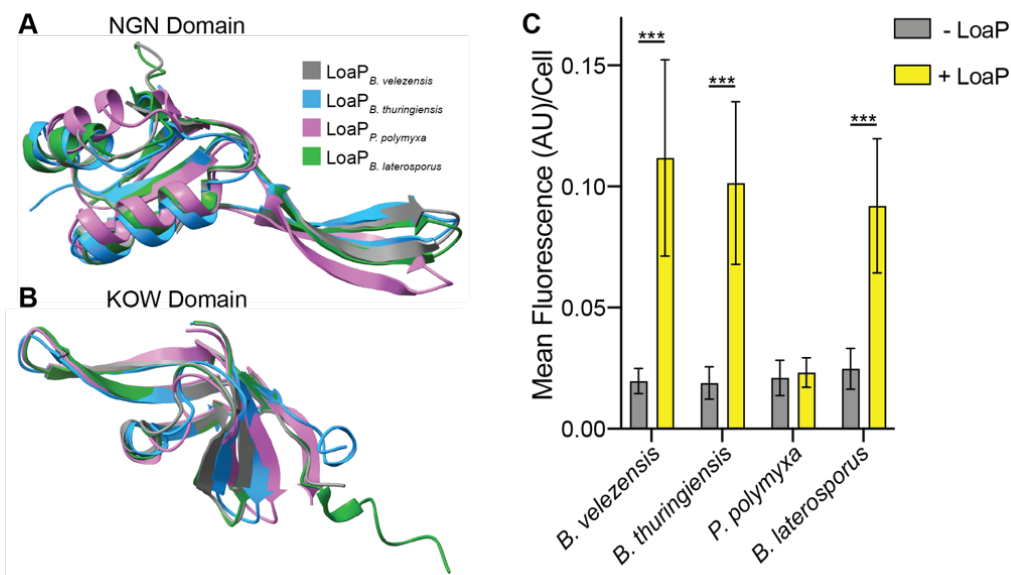


Figure 5-3. LoaP proteins from other species are structurally similar but may differ mechanistically (A-B) Superimposition of predicted structural models of the (a) NGN and (b) KOW domains of LoaP proteins from *B. velezensis* (grey), *B. thuringiensis* (blue), *P. polymyxa* (pink) and *B. laterosporus* (green). Structural models predicted using AlphaFold 2.0. (C) Bar graph representing the average fluorescence intensity per cell measured from *B. subtilis* expressing a constitutive $P_{const}-dfn$ leader-*yfp* reporter, as well as a xylose-inducible copy of *loaP* encoded by the species indicated, integrated into the non-essential *thrC* locus. Statistical significance was determined following unpaired t tests. Three asterisks (***) denotes a p-value < 0.001. Data collected in collaboration with Mansi Srivastava.

rRNA Antitermination in *B. subtilis*

During Rho-dependent termination in bacteria, Rho termination factor associates with G-rich, C-poor regions of nascent transcripts and employs ATP-dependent helicase activity in order to destabilize the TEC^{6,40–42}. This results in the dissociation of RNAP from the nascent RNA and thus the ceasing of transcript production. In this way, Rho also works to silence the global expression of antisense and non-coding RNAs^{42,57,58}. This function is often aided by NusG, but can be overcome at specific sequences by the ability of NusG to couple a leading ribosome to the TEC^{34,59,61}.

In *E. coli* the operons that encode rRNA transcripts (*rrn*) are not translated and therefore lack ribosomes, leaving them as targets for Rho-mediated suppression⁴². The successful expression of rRNA transcripts then relies on a multipartite nucleoprotein complex resulting from the association of a NusB:S10 heterodimer at the *boxA* RNA element present in *rrn* leader regions^{98,109}. Ribosomal protein S10 interacts directly with the KOW domain of NusG, and in concert with additional required factors NusA and SuhB. The resulting complex works to manipulate the exit tunnel of RNAP and tethers ribosomal proteins as a means of blocking Rho^{111,113}.

This mechanism however cannot be meaningfully extrapolated to *B. subtilis*, which does not couple transcription and translation machinery⁷². Moreover, both *nusG* and *rho* are dispensable in *B. subtilis*, suggesting a diminished role of Rho termination for the suppression of long, non-coding transcripts⁷⁰. Collectively, these findings suggest that antitermination for the sake of anti-Rho activity would be largely absent from *B. subtilis*, particular at *rrn* operons. However, that *B. subtilis* encodes for NusB is peculiar, given that NusB has seemingly no characterized function outside of facilitating S10 contacts within the antitermination complex. Additionally, all ten *rrn* operons encoded by *B. subtilis* display conserved *boxA* and *boxB* sequences, suggesting some regulatory function of the NusB:S10 complex in rRNA expression.

To investigate this further, in vivo reporter constructs were generated wherein a sequence (*cssS* antisense) previously shown to confer strong Rho activity in *B. subtilis*⁷² was translationally fused to *yfp*. The promoter region controlling the expression of one of the *B. subtilis* rRNA operons (*rrnO*) was placed upstream of *cssS* (**Figure 5-4A**). These reporters were then integrated single copy into a non-essential locus in the genomes of wildtype *B. subtilis* and $\Delta nusG$, $\Delta nusB$, Δrho and $\Delta suhB$ mutants. The averaged single cell fluorescence determined for

each of the resulting strains demonstrate readthrough past *cssS* only in the absence of Rho (Figure 5-4B). While this was expected, the absence of readthrough in a wildtype background, and that a $\Delta nusG$ background did not phenocopy Δrho , not only suggest a minimal role for NusG in aiding Rho-dependent termination, but that an anti-Rho termination complex is not assembled on the leaders of *B. subtilis rrn* transcripts.

This finding, although striking, is not entirely novel. It has been previously proposed that the rRNA antitermination complex in *E. coli* does not function primarily to prevent premature termination events, but rather enables successful rRNA maturation and functional ribosome biogenesis^{111,207}. While it is known that the assembly of Nus factors at the TEC regulates transcription elongation with impacts on RNA folding, a recent study suggests that these factors, in coordination with RNase III, work to promote proper cotranscriptional folding of 16S rRNA in a way that promotes the formation of fully functional ribosomes. In this process, it is thought that the additional cellular factor, SuhB, maintains crucial contacts between RNAP and *boxA*-bound Nus factors. Additionally, NusB:S10 binding sites have been found upstream of genes unrelated to rRNA production in other organisms²⁰⁷. These data then, supplement a role for the rRNA antitermination complex outside of anti-Rho activity. More importantly, this offers insights into the functions of these factors in a cellular background which is thought to seldom employ Rho and does not couple transcription and translation. In emphasizing key differences in transcription elongation across model organisms, this highlights functions of a TEC complex which may have broader implications across Gram-positive organisms.

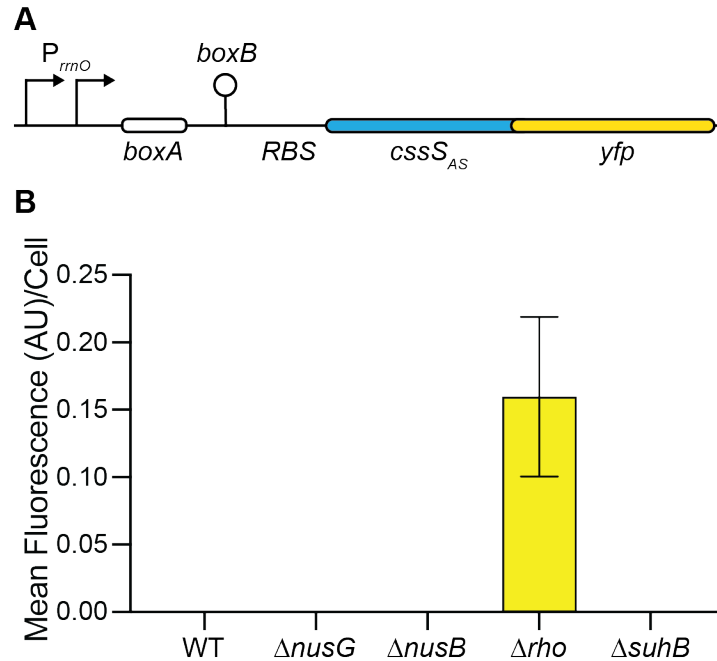


Figure 5-4. *B. subtilis* does not show evidence of antitermination for the expression of *rrn* operons in a Rho-dependent reporter system (A) Cartoon schematic of a *P_{rrnO}-cssS* antisense-*yfp* translational fusion construct (B) Bar graph representing the average fluorescence intensity per cell measured from *B. subtilis* strains expressing a *P_{rrnO}-cssS* antisense -*yfp* reporter integrated at the non-essential *amyE* locus.

Final Thoughts

NusG is a universal transcription factor present in the TECs of all cell types. Bacterial NusG proteins regulate transcription elongation by facilitating interactions with other elongation factors, controlling transcription processivity and termination^{6,15,18}. As discussed throughout this dissertation, all organisms rely on a core, housekeeping NusG, yet many also encode for one or more specialized paralogs controlling the expression of select gene clusters. At the onset of the work described in this dissertation, the only well-characterized NusG paralog was RfaH in *E. coli*. RfaH associates with the TEC following the recognition of a hairpin element in the non-

template DNA strand¹²³. Similar to NusG during antitermination, RfaH bridges interactions between RNAP and the leading ribosome as a means to prevent Rho-dependent termination¹⁵⁷. Knowledge pertaining to NusG and RfaH in *E.coli* set a precedent for antitermination by NusG proteins.

LoaP, the first NusG specialized paralog characterized in a Gram-positive organism, challenges this paradigm. Discovered in *B. velezensis* but widespread amongst other Gram-positive phyla, LoaP is often associated with the expression of operons encoding the biosynthesis of secondary metabolites and polysaccharides¹⁴⁸. However, in stark contrast to RfaH, LoaP was shown to promote readthrough at intrinsic termination sites following the novel recognition of an RNA hairpin element^{148,149,152}. Additionally, the ability of LoaP to binds this RNA hairpin with high affinity and high specificity appears to be the only molecular determinant for LoaP antitermination activity.

The LoaP antitermination mechanism expands the mechanistic diversity of NusG family proteins and likely represents control over termination in an organism that does not couple transcription and translation. Moreover, further investigation into the molecular mechanisms employed by LoaP proteins will aid in the discovery and production of novel natural products.

Appendices

A. Strains used in this dissertation

Strain Name	Strain Background	Genotype	Plasmids Used	Antibiotic Marker	Source
Amr101	<i>E. coli</i> T7 Express	10xHis-SUMO-LoaP _{BV}	pAmr49	Amp	Elghondakly et al., 2021
Amr039	<i>E. coli</i> T7 Express	10xHis-SUMO-NusA _{BV}	pAmr017	Gent	
MDJ151	<i>E. coli</i> T7 Express	10xHis-SUMO-NusA _{BS}	pMJ064	Amp	<i>This study</i>
Amr042	<i>E. coli</i> T7 Express	10xHis-SUMO-NusG _{BV}	pAmr028	Gent	
MDJ141	<i>E. coli</i> T7 Express	10xHis-SUMO-NusG _{BS}	pMJ058	Amp	<i>This study</i>
MDJ143	<i>E. coli</i> T7 Express	10xHis-SUMO-NusG _{EC}	pMJ059	Amp	<i>This study</i>
Amr85	<i>E. coli</i> T7 Express	10xHis-SUMO-S10 _{BV}	pAmr047	Amp	
MDJ145	<i>E. coli</i> T7 Express	10xHis-SUMO-S10 _{BS}	pMJ060	Amp	<i>This study</i>
MDJ147	<i>E. coli</i> T7 Express	10xHis-SUMO-S10 _{EC}	pMJ061	Amp	<i>This study</i>
Amr87	<i>E. coli</i> T7 Express	10xHis-SUMO-NusB _{BS}	pAmr48	Amp	
Amr034	<i>E. coli</i> T7 Express	10xHis-SUMO-Rho _{BV}	pAmr022	Gent	
TMM286	<i>B. subtilis</i> 168	Twin-strep-tagged <i>rpoC</i> (<i>rpoC</i> -2xstrep)	pTM111	Cm	<i>This study</i>
MDJ354	<i>B. subtilis</i> 168	Δ <i>nusG::erm rpoC</i> -2xstrep	pTM111	Erm, Cm	<i>This study</i>
MDJ355	<i>B. subtilis</i> 168	Δ <i>nusB::erm rpoC</i> -2xstrep	pTM111	Erm, Cm	<i>This study</i>
MDJ356	<i>B. subtilis</i> 168	Δ <i>rho::erm rpoC</i> -2xstrep	pTM111	Erm, Cm	<i>This study</i>
FZB42	<i>B. velezensis</i> FZB42	Wild-type <i>B. velezensis</i>			
JG091	<i>B. velezensis</i> FZB42	<i>loaP::erm</i>	pJG030	Erm	Goodson et al., 2017

JG098	<i>B. velezensis</i> FZB42	<i>amyE::Pxyl-loaP</i>	pJG032	Cm	Goodson et al., 2017
JG099	<i>B. velezensis</i> FZB42	<i>loaP::erm</i> <i>amyE::Pxyl-loaP</i>	pJG030 pJG032	Erm, Cm	Goodson et al., 2017
MDJ174	<i>B. velezensis</i> FZB42	<i>loaP::erm</i> <i>amyE::Pxyl-loaP</i> E40A	pJG030 pMJ080	Erm, Cm	<i>This study</i>
MDJ175	<i>B. velezensis</i> FZB42	<i>loaP::erm</i> <i>amyE::Pxyl-loaP</i> R41A	pJG030 pMJ081	Erm, Cm	<i>This study</i>
MDJ189	<i>B. velezensis</i> FZB42	<i>loaP::erm</i> <i>amyE::Pxyl-loaP</i> K128A	pJG030 pMJ083	Erm, Cm	<i>This study</i>
MDJ190	<i>B. velezensis</i> FZB42	<i>loaP::erm</i> <i>amyE::Pxyl-loaP</i> I143A	pJG030 pMJ084	Erm, Cm	<i>This study</i>
JG257	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> <i>amyE::Pconst-dfn-yfp</i>	pJG101 pJG108	Erm, Cm	Elghondakly et al., 2021
MDJ330	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> <i>amyE::Pconst-dfnΔhp-yfp</i>	pJG108 pMJ136	Erm, Cm	<i>This study</i>
MDJ331	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> <i>amyE::Pconst-dfnΔpause-yfp</i>	pJG108 pMJ137	Erm, Cm	<i>This study</i>
MDJ332	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> <i>amyE::Pconst-dfnΔterm-yfp</i>	pJG108 pMJ138	Erm, Cm	<i>This study</i>
MDJ334	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> , <i>amyE::Pconst-dfnΔ+3-12-yfp</i>	pJG108 pMJ139	Erm, Cm	<i>This study</i>
MDJ335	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> , <i>amyE::Pconst-dfnΔ+13-22-yfp</i>	pJG108 pMJ140	Erm, Cm	<i>This study</i>
MDJ336	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> , <i>amyE::Pconst-dfnΔ+23-32-yfp</i>	pJG108 pMJ141	Erm, Cm	<i>This study</i>
MDJ337	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> , <i>amyE::Pconst-dfnΔ+33-42-yfp</i>	pJG108 pMJ142	Erm, Cm	<i>This study</i>
MDJ338	<i>B. subtilis</i> 168	<i>thrC::Pxyl-loaP</i> <i>amyE::Pconst-dfnΔ+43-52-yfp</i>	pJG108 pMJ142	Erm, Cm	<i>This study</i>
MDJ345	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> pMJ106(Pxyl-nusA)	pJG101 pMJ106	Erm, Cm	<i>This study</i>
MDJ347	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfnΔterm-yfp</i> pMJ106(Pxyl-nusA)	pMJ138 pMJ106	Erm, Cm	<i>This study</i>
MDJ361	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (K42A)	pJG101 pMJ016	Erm, Cm	<i>This study</i>
MDJ362	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (K46A)	pJG101 pMJ017	Erm, Cm	<i>This study</i>
MDJ363	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (K52A)	pJG101 pMJ019	Erm, Cm	<i>This study</i>
MDJ364	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (T62A)	pJG101 pMJ038	Erm, Cm	<i>This study</i>
MDJ365	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i>	pJG101	Erm, Cm	<i>This study</i>

		<i>thrC::Pxyl-loaP</i> (K71A)	pMJ039		
MDJ366	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (S78A)	pJG101 pMJ040	Erm, Cm	<i>This study</i>
MDJ367	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (C79A)	pJG101 pMJ041	Erm, Cm	<i>This study</i>
MDJ368	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (R80A)	pJG101 pMJ042	Erm, Cm	<i>This study</i>
MDJ369	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (I81A)	pJG101 pMJ043	Erm, Cm	<i>This study</i>
MDJ370	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (N83A)	pJG101 pMJ044	Erm, Cm	<i>This study</i>
MDJ371	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (E9A)	pJG101 pMJ033	Erm, Cm	<i>This study</i>
MDJ372	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (T15A)	pJG101 pMJ011	Erm, Cm	<i>This study</i>
MDJ373	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (F24A)	pJG101 pMJ034	Erm, Cm	<i>This study</i>
MDJ374	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (E26A)	pJG101 pMJ035	Erm, Cm	<i>This study</i>
MDJ375	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (S31A)	pJG101 pMJ036	Erm, Cm	<i>This study</i>
MDJ376	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (K35A)	pJG101 pMJ014	Erm, Cm	<i>This study</i>
MDJ377	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>thrC::Pxyl-loaP</i> (K36A)	pJG101 pMJ037	Erm, Cm	<i>This study</i>
MS085	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>Pxyl-loaP</i> (<i>B. velezensis</i>)	pJG101 pMJ085	Erm, Cm	<i>This study</i>
MS082	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>Pxyl-loaP</i> (<i>B. thuringiensis</i>)	pJG101 pMJ095	Erm, Cm	<i>This study</i>
MS083	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>Pxyl-loaP</i> (<i>P. polymyxa</i>)	pJG101 pMJ104	Erm, Cm	<i>This study</i>
MS086	<i>B. subtilis</i> 168	<i>amyE::Pconst-dfn-yfp</i> <i>Pxyl-loaP</i> (<i>B. laterosporus</i>)	pJG101 pMJ111	Erm, Cm	<i>This study</i>
MDJ380	<i>B. subtilis</i> 168	<i>amyE::PrnO-rrn</i> UTR- <i>cssS</i> antisense- <i>yfp</i>	pMJ153	Cm	<i>This study</i>
MDJ392	<i>B. subtilis</i> 168	Δ <i>nusG::erm</i> <i>amyE::PrnO-rrn</i> UTR- <i>cssS</i> antisense- <i>yfp</i>	pMJ153	Erm, Cm	<i>This study</i>
MDJ381	<i>B. subtilis</i> 168	Δ <i>nusB::erm</i> <i>amyE::PrnO-rrn</i> UTR- <i>cssS</i> antisense- <i>yfp</i>	pMJ153	Erm, Cm	<i>This study</i>
MDJ382	<i>B. subtilis</i> 168	Δ <i>rho::erm</i> <i>amyE::PrnO-rrn</i> UTR- <i>cssS</i> antisense- <i>yfp</i>	pMJ153	Erm, Cm	<i>This study</i>
MDJ393	<i>B. subtilis</i> 168	Δ <i>suhB::erm</i> <i>amyE::PrnO-rrn</i> UTR- <i>cssS</i> antisense- <i>yfp</i>	pMJ153	Erm, Cm	<i>This study</i>

- (full list of M Jermain strains located in Winkler Lab drive in ‘M Jermain Strains.xlsx’)

B. DNA/RNA oligos used in this dissertation

- (full list of M Jermain oligos located in Winkler Lab drive in 'M Jermain Oligos.xlsx')

C. Plasmids used in this dissertation

- (full collection of M Jermain plasmids located in Winkler Lab drive in folder 'M Jermain Plasmids')

D. RNA sequences used in equilibrium binding assays (with 5' GGAAA extensions)

1. Wild type *dfn* hairpin

5' – (GGAAA)GGCAAUCGCUGCUUCGGCACGUUGCC – 3'

2. *boxB* hairpin

5' – (GGAAA)GGGCCCUGAAGAAGGGCCC – 3'

3. dA5, *ps-C7 dfn* hairpin

5' – (GGAAA)GGCAAdAUC*GCUGCUUCGGCACGUUGCC – 3'

4. *ps-dU6 dfn* hairpin

5' – (GGAAA)GGCAAdU*CGCUGCUUCGGCACGUUGCC – 3'

5. *ps-3-methyl U6 dfn* hairpin

5' – (GGAAA)GGCAAU3m*CGCUGCUUCGGCACGUUGCC – 3'

6. *ps-G8 dfn* hairpin

5' – (GGAAA)GGCAAUCG*CGCUGCUUCGGCACGUUGCC – 3'

7. *ps-G11 dfn* hairpin

5' – (GGAAA)GGCAAUCGCUG*CUUCGGCACGUUGCC – 3'

8. *ps-C12 dfn* hairpin

5' – (GGAAA)GGCAAUCGCUGC*UUCGGCACGUUGCC – 3'

9. 4-thio-U15 *dfn* hairpin

5' – (GGAAA)GGCAAUCGCUGCUU-4-thio-UGGCACGUUGCC – 3'

10. *ps*-dC15 *dfn* hairpin

5' – (GGAAA)GGCAAUCGCUGCUUdC*GGCACGUUGCC – 3'

11. *ps*-G16 *dfn* hairpin

5' – (GGAAA)GGCAAUCGCUGCUUCG*GCACGUUGCC – 3'

12. *dfnA* terminator hairpin

5' – (GGAAA)CACAAUGCGUUUGCGAUUUUAAACCUGUGCUGAAUCGCAAA
CGCGUUUGUG – 3'

13. mutant *dfnA* terminator hairpin

5' – (GGAAA)CACAGUUUGCGAUUUUAAACCUGUGCUGAAUCGCAAACUGU
G – 3'

E. DNA sequences used in equilibrium binding assays

1. *dfn* UTR segment 1 (non-template)

5' – AATAGCAACT – 3'

2. *dfn* UTR segment 2 (non-template)

5' – ATTTGTTAAA – 3'

3. *dfn* UTR segment 3 (non-template)

5' – TTTATTAAAT – 3'

4. *dfn* UTR segment 4 (non-template)

5' – TGATGTAAAG – 3'

5. *dfn* UTR segment 5 (non-template)

5' – CACCTATTAA – 3'

6. *dfn* UTR segment 1 (template)

5' – AGTTGCTATT – 3'

7. *dfn* UTR segment 2 (template)

5' – TTAAACAAAT – 3'

8. *dfn* UTR segment 3 (template)

5' – ATTTAATAAAA – 3'

9. *dfn* UTR segment 4 (template)

5' – CTTTACATCA – 3'

10. *dfn* UTR segment 5 (template)

5' – TTAATAGGTG – 3'

F. *dfn* leader transcription template

Promoter

[Transcription start site]

C-less cassette

dfn hairpin

pause site

dfnA terminator

AATACTGAATTGTAAGGATCCAAGAACAGCTTGACAAATACACAAGAGTGTGTTATAATGCAATT[A
|GAATGAGTTGAGTTAGAGAATAGGGTAGCGAAAGCTACCCTCGAGTTAAATAGCAACTATTTGTTAAATT
TATTAAATTGATGTAAAGCACCTATTAAA|TTCAATCGCTGTTTCGGCACGTTGAAGAGGAAATAG
GGATTATAAAAAATGATTGTAGTGTATATCTACATATACATGATAAGGGAGAAAGTGTGCCGTCATG
GCTGTTTTCGGGCATATTGGCACTGCAACTTGTTCATGCGAAACATTTCAGTAAACAAGCAAGATGTCT
TTATATAAATTATGTTAGAAAAGGAAAA|CACAATGCGTTTGCGATTTTAAACCTGTGCTGAATCG
CAAACGCGTTTGTTTATTTTATTTTATGCCGGTGTTTACACCAGTCGAATTCAAGAACGAATTCAAG
CTTGGCACTGGCCGTCGTTTTACAACGTCGTGAC

G. Protein sequences used in this dissertation

B. velezensis LoaP

MKWYALFVESGKEETVQKFLRLQFDEQALYSIIPKKKVTERKAGIKYEALKKMFPGYVLFKTKMTERTFH
KIKELPISCRIVNNGAYYSKERKTYFTTIKDEEILPIIRLIGEGDVTVDYSKVYIENSKVTVASGPLKGMEGIIKK
IDKRKRRAKICLSFMGLDKMVNVGIEVLSKP

T. pseudethanolicus LoaP

MKKWYVIFTRSGYENKVRDIENCfKEEVKLLIPKRKIIERVKGQPVKIKLLFPGYVVFVNAEMSDDLYYKIS
EVLKRGIFLKEGKRPAFVKEEEMKIILSLTKNSDLIDLKSGIMEGERVKIIEGPLKGYEGLIKKIDKRKKRAK
VIFSIAGELKSVDLAIEVMENVSEQQRSLVYAC

B. thuringiensis LoaP

MKILSWYALCVRSGYEDLVESLILRDLNETLVSVLPKKTVPKRGHFKWVTRKLLPGYILINTSMDAKM
YYKLKEIPNCYNFLNKFQDDDDIYFLEIPDQEITPIIQLLNDRGITIECSEAYFEGSKVIVKSGPLKGREGIIRRID
KRKKRAKLLNFLGDTKMIDVGVEFLSSQKLE

***B. laterosporus* LoaP**

MKWYALYVRSGGEEIVKDHLNYYFDEQTLIALIPKRKVPEKKNGAIFHNIKILFPGYVLIKTDMTSSIFYKI
KEIPNLYKFVNIGPIYSKLTSSNHAKKDISITNFSASIDDEEMSPLLRLLGNTDIITYSKIHLENSEVYVNSGPLK
GMEGQIKKVDRHKKRAKVLIFMGSEKLVDLGVEILSSLKLNLSN

***P. polymyxa* LoaP**

MNWYVLFVRTGREERVKQLFNKWFDESEVYKPIPLQERLFKVAGTVKKELAPLFPSYVFIESNLPDLQFVT
STNSMIYTSSDIIRLLRYSKLEASMRDCERQVLESCLNDRYCIECSTGIIEGDNIRIIEGPLKGRSSIVKKIDRHK
RQAVIQLEFMGDIRLVNVALEIISKV

***B. velezensis* NusA**

MSELDDALTILEKEKGISKDIIIEAIEAALISAYKRNFNQAQNVVRVDLNRETGTIRVFARKDVVDEVYDSRL
EISLDDAQHIHPNYLVGDVVEIEVTPKDFGRIAAQTAKQVVTQVRVREAERGVIYSEFIDREEDIMTGIVQRLD
SKFIYVSLGKIEALLPVNEQMPNESYKPHDRIKVFITKVEKTTKGPQIYVSRTHPGLLKRLFIEVPEIYDGTV
ELKSVAREAGDRSKISVRTDDPDVDPVGSVCVGPKGQQRVQAIVNELKGEKIDIVNWSDDPVEFVANALSPSK
VLDVIVSEEEKATTVIVPDYQLSLAIGKRGQNARLAAKLTGWKIDIKSETDARDLGIYPREHEEDDDLEPLF
MEPETAESDE

***B. subtilis* NusB**

MKRRTAREKALQALFQIDVSDIAVNEAIEHALDEEKTDPFFEQLVHGVLEHQDQLDEMISKHLVNWKLDR
ANVDRAILRLAAYEMAYAEDIPVNVSMNEAIELAKRFGDDKATKVFVNGVLSNIKSIEQS

***B. velezensis* NusB**

MKRRTAREKALQALFQIDVSDIAPNDAIEHALDGEKTDAPFEQLVYGVIEHQVQLDEMISGHLVNWKLDR
ANVDRAILRLAVYEMVYSDDIPANVSLNEAIELAKRFGDDKAAKFVNGVLSNIKTIES

***E. coli* NusG**

MSEAPKKRWYVVQAFSGFEGRVATSLREHIKLNHNMEDLFGEVMVPTEEVVEIRGGQRRKSERKFFPGYVL
VQMVMNDASWHLVRSVPRVMGFIGGTSRDPAPISDKEVDAIMNRLQQVGDKPRPKTLFEPGEMVRVNDG
PFADFNGVVEEVDYKESRLKVSVSIFGRATPVELDFSQVEKA

***B. subtilis* NusG**

MEKNWYVVHTYSGYENKVKANLEKRVESMGMQDKIFRVVVPPEEEETDIKNGKKKVKKKVFPGYVLVE
IVMTDDSWYVVRNTPGVTGFVGSAGSGSKPTPLPGEAETILKRMGMDERKTDIDFELKETVKVIDGPAN
FTGSIEEIDYDKSKVKVFNMFGRETPVELEFTQIDKL

***B. velezensis* NusG**

MEKNWYVVHTYSGYENKVKANLEKRVESMGMQDKIFRVVVPPEEEETDIKNGKKKVKKKVFPGYVLVE
IVMTDDSWYVVRNTPGVTGFVGSAGSGSKPTPLPGEAETILKRMGMDERKTDIDFELNETVKVIDGPAN
FTGSIEEIDYDKSKVKVFNMFGRETPVELEFTQIDKL

***E. coli* S10 (NusE)**

MQNQIRIRLKAQFDHRLIDQATAEIVETAKRTGAQVRGPIPLPTRKERFTVLISPHVNKD
ARDQYEIRTHRLRLVDIVEPTEKTVDALMRLDLAAGVDVQISLG

***B. subtilis* S10 (NusE)**

MAKQKIRIRLKAYDHRILDQSAEKIVETAKRSGASVSGPIPLPTEKSVYTILRAVHKYKD
SREQFEMRTHKRLIDIVNPTPQTVDALMRLDLPSGVDIEIKL

***B. velezensis* S10 (NusE)**

MAKQKIRIRLKAYDHRILDQSAEKIVETAKRSGASVSGPIPLPTEKSVYTILRAVHKYKD
SREQFEMRTHKRLIDIVNPTPQTVDALMRLDLPSGVDIEIKL

***B. velezensis* Rho**

MKDVSISSENMKLKELYELARHYKISYYSKLTKKELIFAILKANAEQEDLLFMEGVLEIIQSEGFGFLRPIN
YSPSEDIYISASQIRRFDLRNGDKVSGKVRPPKENERYYYGLLHVEAVNGDDPESAKERVHFPALTPLYPDR
QMVLETKPNYLSTRILDMMAPVGFQGRGLIVAPPKAGKTMLLKEIANSITTNPDAELIVLLIDERPEEVTDI
ERSVAGDVVSSTFDEVPHIKVAELVLERAMRLVEHKKDVIILMDSITRLARAYNLVIPPSGRTLSGGIDPA
AFHRPKRFFGAARNIEEGSLTILATALVDTGSRMDDVIYEEFKGTGNMELHLDRSLAERRIFPAIDIRRS
RKEELLVPKEHLDRWLWAIKSMSPDFAEKFMRKMKKTKTNQEFFDLINQEWKQANLSSSARR

***B. subtilis* RNAP β' subunit**

MLDVNNFEYMNIGLASPDKIRSWSFGEVKKPETINYRTLKPEKDGLFCERIFGPTKDWECHCGKYKRVRYK
GVVCDRCGVEVTRAKVRRERMGHIELAAPVSHIWFYFKGIPSRMGLVLDMSPRALEEVIYFASYVVTDPAN
TPLEKKQLLSEKEYRAYLDKYGNKFQASMGAEAIHKLQDIDLVEVDMLKEELKTSQGQRRTRAIKRL
VLEAFRNSGNKPSWMILDVLPVIPPELRPMVQLDGGRFATSDLNLYRRVINRNNRLKRLLDLGAPSIIVQN
EKRMQLQEAVDALIDNGRRGRPVTGPGNRPLKSLSHMLKGKQGRFRQNLGKRVDSGRSVIVVGPHLKM
YQCGLPKEMALELFPFVMKELVEKGLAHNIKSAKRKIERVQPEVWDVLESVIKEHPVLLNRAPTLHRLGI
QAFEPTLVEGRAIRLHPLVCTAYNADFDGDQMAVHVPLSAEAQAEARILMLAAQNILNPKDGKPVVTPSQ
DMVLGNYYLTLEAGAVGEGMVFKNTDEALLAYQNGYVHLHTRVAVAAANSLKNVTFTEEQRSKLLITTV
GKLVFNEILPESFPYMNEPTKSNIEEKTDRFFLEKGADVKA VIAQQPINAPFKKGILGKIIAEIFKRFRHITETS
KMLDRMKNLGFKYSTKAGITVGVSDIVVLDDKQEILEEAQSKVDNVMKQFRRGLITEEERYERVISIWSAA
KDVIQGKLMKSLDELNPIYMMSDSGARGNASNFTQLAGMRGLMANPAGRIIHELPIKSSFREGLTVLEYFIST
HGARKGLADTALKTADSGYLTRRLVDVAQDVIIRETDCGTDRGILAKPLKEGTETIERLEERLIGRFARKQV
KHPETGEVLVNENELIDEDKALEIVEAGIEEVWIRSAFTCNTPHGVCKRCYGRNLATGSDVEVGEAVGIIAA
QSIGEPGTQLTMRTFHTGGVAGDDITQGLPRIQELFEARNPKGQATITEIDGTVEINEVRDKQQEIVVQGA
VETRSTAPYNSRLKVAEGDKITRGQVLTEGSIDPKELLKVTDLTTVQEYLLHEVQKVYRMQGVIGDKH
VEVMVRQMLRKVRVIDAGDTDVLPGTLLDIHQFTEANKKVLLEGNRPATGRPVLGITKASLETDSFLSAA
SFQETTRVLTDAAIKGKRDELLGLKENVIIGKLVPAGTGMMKYRKVKPVSNVQPTDDMVPVE

References

1. Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the whole story. *Curr. Opin. Struct. Biol.* 13, 31–39. [https://doi.org/10.1016/s0959-440x\(02\)00005-2](https://doi.org/10.1016/s0959-440x(02)00005-2).
2. Browning, D.F., and Busby, S.J.W. (2016). Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* 14, 638–650. <https://doi.org/10.1038/nrmicro.2016.103>.
3. Ruff, E.F., Record, M.T., and Artsimovitch, I. (2015). Initial Events in Bacterial Transcription Initiation. *Biomolecules* 5, 1035–1062. <https://doi.org/10.3390/biom5021035>.
4. Belogurov, G.A., and Artsimovitch, I. (2014). Regulation of Transcript Elongation. *Annu. Rev. Microbiol.* 69, 1–21. <https://doi.org/10.1146/annurev-micro-091014-104047>.
5. Mustaev, A., Roberts, J., and Gottesman, M. (2017). Transcription elongation. *Transcription* 8, 150–161. <https://doi.org/10.1080/21541264.2017.1289294>.
6. Ray-Soni, A., Bellecourt, M.J., and Landick, R. (2015). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.* 85, 1–29. <https://doi.org/10.1146/annurev-biochem-060815-014844>.
7. Helmann, J.D., and Chamberlin, M.J. (1988). Structure and Function of Bacterial Sigma Factors. *Annu. Rev. Biochem.* 57, 839–872. <https://doi.org/10.1146/annurev.bi.57.070188.004203>.
8. Feklistov, A., Sharon, B.D., Darst, S.A., and Gross, C.A. (2014). Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. *Annu. Rev. Microbiol.* 68, 1–20. <https://doi.org/10.1146/annurev-micro-092412-155737>.
9. Murakami, K.S. (2013). X-ray Crystal Structure of Escherichia coli RNA Polymerase σ 70 Holoenzyme*. *J. Biol. Chem.* 288, 9126–9134. <https://doi.org/10.1074/jbc.m112.430900>.
10. Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R., and Ebright, R.H. (2002). Structural Organization of Bacterial RNA Polymerase Holoenzyme and the RNA Polymerase-Promoter Open Complex. *Cell* 108, 599–614. [https://doi.org/10.1016/s0092-8674\(02\)00667-0](https://doi.org/10.1016/s0092-8674(02)00667-0).
11. Lloyd, G., Landini, P., and Busby, S. (2001). Activation and repression of transcription initiation in bacteria. *Essays Biochem.* 37, 17–31. <https://doi.org/10.1042/bse0370017>.

12. Busby, S.J.W. (2019). Transcription activation in bacteria: ancient and modern. *Microbiology* 165, 386–395. <https://doi.org/10.1099/mic.0.000783>.
13. Belogurov, G.A., and Artsimovitch, I. (2019). The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *J. Mol. Biol.* 431, 3975–4006. <https://doi.org/10.1016/j.jmb.2019.05.042>.
14. Blombach, F., Daviter, T., Fielden, D., Grohmann, D., Smollett, K., and Werner, F. (2013). Archaeology of RNA polymerase: factor swapping during the transcription cycle. *Biochem. Soc. Trans.* 41, 362–367. <https://doi.org/10.1042/bst20120274>.
15. Wang, B., and Artsimovitch, I. (2021). NusG, an Ancient Yet Rapidly Evolving Transcription Factor. *Front Microbiol* 11, 619618. <https://doi.org/10.3389/fmicb.2020.619618>.
16. Tomar, S.K., and Artsimovitch, I. (2013). NusG-Spt5 Proteins—Universal Tools for Transcription Modification and Communication. *Chem. Rev.* 113, 8604–8619. <https://doi.org/10.1021/cr400064k>.
17. Ponting, C.P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* 30, 3643–3652. <https://doi.org/10.1093/nar/gkf498>.
18. Kang, J.Y., Mooney, R.A., Nedialkov, Y., Saba, J., Mishanina, T.V., Artsimovitch, I., Landick, R., and Darst, S.A. (2018). Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell* 173, 1650–1662.e14. <https://doi.org/10.1016/j.cell.2018.05.017>.
19. Grachev, M.A., Zaychikov, E.F., Ivanova, E.M., Komarova, N.I., Kutayvin, I.V., Sidelnikova, N.P., and Frolova, I.P. (1984). Oligonucleotides complementary to a promoter over the region -8...+2 as transcription primers for *E. coli* RNA polymerase. *Nucleic acids Res.* 12, 8509–8524. <https://doi.org/10.1093/nar/12.22.8509>.
20. Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural Basis of Transcription Initiation. *Science* 338, 1076–1080. <https://doi.org/10.1126/science.1227786>.
21. Weixlbaumer, A., Leon, K., Landick, R., and Darst, S.A. (2013). Structural Basis of Transcriptional Pausing in Bacteria. *Cell* 152, 431–441. <https://doi.org/10.1016/j.cell.2012.12.020>.
22. Hein, P.P., Kolb, K.E., Windgassen, T., Bellecourt, M.J., Darst, S.A., Mooney, R.A., and Landick, R. (2014). RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* 21, 794–802. <https://doi.org/10.1038/nsmb.2867>.
23. Gusarov, I., and Nudler, E. (1999). The Mechanism of Intrinsic Transcription Termination. *Mol. Cell* 3, 495–504. [https://doi.org/10.1016/s1097-2765\(00\)80477-3](https://doi.org/10.1016/s1097-2765(00)80477-3).

24. Yarnell, W.S., and Roberts, J.W. (1999). Mechanism of Intrinsic Transcription Termination and Antitermination. *Science* 284, 611–615. <https://doi.org/10.1126/science.284.5414.611>.
25. Ha, K.S., Touloukhonov, I., Vassilyev, D.G., and Landick, R. (2010). The NusA N-Terminal Domain Is Necessary and Sufficient for Enhancement of Transcriptional Pausing via Interaction with the RNA Exit Channel of RNA Polymerase. *J. Mol. Biol.* 401, 708–725. <https://doi.org/10.1016/j.jmb.2010.06.036>.
26. Lee, D.N., Phung, L., Stewart, J., and Landick, R. (1990). Transcription pausing by *Escherichia coli* RNA polymerase is modulated by downstream DNA sequences. *J. Biol. Chem.* 265, 15145–15153. [https://doi.org/10.1016/s0021-9258\(18\)77234-7](https://doi.org/10.1016/s0021-9258(18)77234-7).
27. Nudler, E., Mustaev, A., Goldfarb, A., and Lukhtanov, E. (1997). The RNA–DNA Hybrid Maintains the Register of Transcription by Preventing Backtracking of RNA Polymerase. *Cell* 89, 33–41. [https://doi.org/10.1016/s0092-8674\(00\)80180-4](https://doi.org/10.1016/s0092-8674(00)80180-4).
28. Komissarova, N., and Kashlev, M. (1997). Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc. Natl. Acad. Sci.* 94, 1755–1760. <https://doi.org/10.1073/pnas.94.5.1755>.
29. Komissarova, N., and Kashlev, M. (1997). RNA Polymerase Switches between Inactivated and Activated States By Translocating Back and Forth along the DNA and the RNA*. *J. Biol. Chem.* 272, 15329–15338. <https://doi.org/10.1074/jbc.272.24.15329>.
30. Guajardo, R., and Sousa, R. (1997). A model for the mechanism of polymerase translocation. *J. Mol. Biol.* 265, 8–19. <https://doi.org/10.1006/jmbi.1996.0707>.
31. 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. <https://doi.org/10.1074/jbc.m115.704189>.
32. Herbert, K.M., Zhou, J., Mooney, R.A., Porta, A.L., Landick, R., and Block, S.M. (2010). *E. coli* NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward Translocation of RNA Polymerase. *J. Mol. Biol.* 399, 17–30. <https://doi.org/10.1016/j.jmb.2010.03.051>.
33. 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. [https://doi.org/10.1016/s0968-0004\(96\)30036-4](https://doi.org/10.1016/s0968-0004(96)30036-4).
34. Lawson, M.R., Ma, W., Bellecourt, M.J., Artsimovitch, I., Martin, A., Landick, R., Schulten, K., and Berger, J.M. (2018). Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. *Mol Cell* 71, 911–922.e4. <https://doi.org/10.1016/j.molcel.2018.07.014>.

35. Saxena, S., Myka, K.K., Washburn, R., Costantino, N., Court, D.L., and Gottesman, M.E. (2018). Escherichia coli transcription factor NusG binds to 70S ribosomes. *Mol Microbiol* 108, 495–504. <https://doi.org/10.1111/mmi.13953>.
36. Mooney, R.A., Schweimer, K., Rösch, P., Gottesman, M., and Landick, R. (2009). Two Structurally Independent Domains of E. coli NusG Create Regulatory Plasticity via Distinct Interactions with RNA Polymerase and Regulators. *J. Mol. Biol.* 391, 341–358. <https://doi.org/10.1016/j.jmb.2009.05.078>.
37. Wang, C., Molodtsov, V., Firlar, E., Kaelber, J.T., Blaha, G., Su, M., and Ebright, R.H. (2020). Structural basis of transcription-translation coupling. *Science* 369, 1359–1365. <https://doi.org/10.1126/science.abb5317>.
38. Webster, M.W., Takacs, M., Zhu, C., Vidmar, V., Eduljee, A., Abdelkareem, M., and Weixlbaumer, A. (2020). Structural basis of transcription-translation coupling and collision in bacteria. *Science* 369, 1355–1359. <https://doi.org/10.1126/science.abb5036>.
39. Yakhnin, A.V., FitzGerald, P.C., McIntosh, C., Yakhnin, H., Kireeva, M., Turek-Herman, J., Mandell, Z.F., Kashlev, M., and Babitzke, P. (2020). NusG controls transcription pausing and RNA polymerase translocation throughout the Bacillus subtilis genome. *Proc National Acad Sci* 117, 21628–21636. <https://doi.org/10.1073/pnas.2006873117>.
40. Epshtein, V., Dutta, D., Wade, J., and Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463, 245–249. <https://doi.org/10.1038/nature08669>.
41. Holmes, W.M., Platt, T., and Rosenberg, M. (1983). Termination of transcription in E. coli. *Cell* 32, 1029–1032. [https://doi.org/10.1016/0092-8674\(83\)90287-8](https://doi.org/10.1016/0092-8674(83)90287-8).
42. Mitra, P., Ghosh, G., Hafeezunnisa, Md., and Sen, R. (2017). Rho Protein: Roles and Mechanisms. *Annu. Rev. Microbiol.* 71, 1–23. <https://doi.org/10.1146/annurev-micro-030117-020432>.
43. Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. (2002). Shortening of RNA:DNA Hybrid in the Elongation Complex of RNA Polymerase Is a Prerequisite for Transcription Termination. *Mol. Cell* 10, 1151–1162. [https://doi.org/10.1016/s1097-2765\(02\)00738-4](https://doi.org/10.1016/s1097-2765(02)00738-4).
44. Mondal, S., Yakhnin, A.V., Sebastian, A., Albert, I., and Babitzke, P. (2016). NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat. Microbiol.* 1, 15007. <https://doi.org/10.1038/nmicrobiol.2015.7>.
45. McSwiggen, J.A., Bear, D.G., and Hippel, P.H. von (1988). Interactions of Escherichia coli transcription termination factor rho with RNA. *J. Mol. Biol.* 199, 609–622. [https://doi.org/10.1016/0022-2836\(88\)90305-1](https://doi.org/10.1016/0022-2836(88)90305-1).

46. Hart, C.M., and Roberts, J.W. (1991). Rho-dependent transcription termination. Characterization of the requirement for cytidine in the nascent transcript. *J. Biol. Chem.* 266, 24140–24148. [https://doi.org/10.1016/s0021-9258\(18\)54405-7](https://doi.org/10.1016/s0021-9258(18)54405-7).
47. Gogol, E.P., Seifried, S.E., and Hippel, P.H. von (1991). Structure and assembly of the *Escherichia coli* transcription termination factor rho and its interactions with RNA I. Cryoelectron microscopic studies. *J. Mol. Biol.* 221, 1127–1138. [https://doi.org/10.1016/0022-2836\(91\)90923-t](https://doi.org/10.1016/0022-2836(91)90923-t).
48. Briercheck, D.M., Wood, T.C., Allison, T.J., Richardson, J.P., and Rule, G.S. (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. <https://doi.org/10.1038/nsb0598-393>.
49. Allfano, P., Rivellini, F., Limauro, D., Bruni, C.B., and Carlomagno, M.S. (1991). A consensus motif common to all rho-dependent prokaryotic transcription terminators. *Cell* 64, 553–563. [https://doi.org/10.1016/0092-8674\(91\)90239-u](https://doi.org/10.1016/0092-8674(91)90239-u).
50. Skordalakes, E., and Berger, J.M. (2003). Structure of the Rho Transcription Terminator Mechanism of mRNA Recognition and Helicase Loading. *Cell* 114, 135–146. [https://doi.org/10.1016/s0092-8674\(03\)00512-9](https://doi.org/10.1016/s0092-8674(03)00512-9).
51. Koslover, D.J., Fazal, F.M., Mooney, R.A., Landick, R., and Block, S.M. (2012). Binding and Translocation of Termination Factor Rho Studied at the Single-Molecule Level. *J. Mol. Biol.* 423, 664–676. <https://doi.org/10.1016/j.jmb.2012.07.027>.
52. Thomsen, N.D., and Berger, J.M. (2009). Running in Reverse: The Structural Basis for Translocation Polarity in Hexameric Helicases. *Cell* 139, 523–534. <https://doi.org/10.1016/j.cell.2009.08.043>.
53. Browne, R.J., Barr, E.W., and Stitt, B.L. (2005). Catalytic Cooperativity among Subunits of *Escherichia coli* Transcription Termination Factor Rho KINETICS AND SUBSTRATE STRUCTURAL REQUIREMENTS*. *J. Biol. Chem.* 280, 13292–13299. <https://doi.org/10.1074/jbc.m500221200>.
54. Soares, E., Schwartz, A., Nollmann, M., Margeat, E., and Boudvillain, M. (2014). The RNA-mediated, asymmetric ring regulatory mechanism of the transcription termination Rho helicase decrypted by time-resolved Nucleotide Analog Interference Probing (trNAIP). *Nucleic Acids Res.* 42, 9270–9284. <https://doi.org/10.1093/nar/gku595>.
55. Brennan, C.A., Dombroski, A.J., and Platt, T. (1987). Transcription termination factor rho is an RNA-DNA helicase. *Cell* 48, 945–952. [https://doi.org/10.1016/0092-8674\(87\)90703-3](https://doi.org/10.1016/0092-8674(87)90703-3).
56. Lowery-Goldhammer, C., and Richardson, J.P. (1974). An RNA-Dependent Nucleoside Triphosphate Phosphohydrolase (ATPase) Associated with Rho Termination Factor. *Proc. Natl. Acad. Sci.* 71, 2003–2007. <https://doi.org/10.1073/pnas.71.5.2003>.

57. Peters, J.M., Mooney, R.A., Grass, J.A., Jessen, E.D., Tran, F., and Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev.* 26, 2621–2633. <https://doi.org/10.1101/gad.196741.112>.
58. Bossi, L., Ratel, M., Laurent, C., Kerboriou, P., Camilli, A., Eveno, E., Boudvillain, M., and Figueroa-Bossi, N. (2019). NusG prevents transcriptional invasion of H-NS-silenced genes. *Plos Genet* 15, e1008425. <https://doi.org/10.1371/journal.pgen.1008425>.
59. Lawson, M.R., and Berger, J.M. (2019). Tuning the sequence specificity of a transcription terminator. *Curr. Genet.* 65, 729–733. <https://doi.org/10.1007/s00294-019-00939-1>.
60. Jr., O.L.M., Hamkalo, B.A., and Jr., C.A.T. (1970). Visualization of Bacterial Genes in Action. *Science* 169, 392–395. <https://doi.org/10.1126/science.169.3943.392>.
61. 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. *Sci. (N. York, NY)* 328, 501–504. <https://doi.org/10.1126/science.1184953>.
62. Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010). Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science* 328, 504–508. <https://doi.org/10.1126/science.1184939>.
63. Stevenson-Jones, F., Woodgate, J., Castro-Roa, D., and Zenkin, N. (2020). Ribosome reactivates transcription by physically pushing RNA polymerase out of transcription arrest. *Proc. Natl. Acad. Sci. United States Am.* 117, 8462–8467. <https://doi.org/10.1073/pnas.1919985117>.
64. Gowrishankar, J., Leela, J.K., and Anupama, K. (2013). R-loops in bacterial transcription: Their causes and consequences. *Transcription* 4, 153–157. <https://doi.org/10.4161/trns.25101>.
65. Iost, I., and Dreyfus, M. (1995). The stability of *Escherichia coli* lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J.* 14, 3252–3261. <https://doi.org/10.1002/j.1460-2075.1995.tb07328.x>.
66. Washburn, R.S., Zuber, P.K., Sun, M., Hashem, Y., Shen, B., Li, W., Harvey, S., Reyes, F.J.A., Gottesman, M.E., Knauer, S.H., et al. (2020). *Escherichia coli* NusG Links the Lead Ribosome with the Transcription Elongation Complex. *Iscience* 23, 101352. <https://doi.org/10.1016/j.isci.2020.101352>.
67. Richardson, J.P. (1991). Preventing the synthesis of unused transcripts by rho factor. *Cell* 64, 1047–1049. [https://doi.org/10.1016/0092-8674\(91\)90257-y](https://doi.org/10.1016/0092-8674(91)90257-y).
68. Lane, W.J., and Darst, S.A. (2010). Molecular Evolution of Multisubunit RNA Polymerases: Sequence Analysis. *J. Mol. Biol.* 395, 671–685. <https://doi.org/10.1016/j.jmb.2009.10.062>.

69. Downing, W.L., Sullivan, S.L., Gottesman, M.E., and Dennis, P.P. (1990). Sequence and transcriptional pattern of the essential *Escherichia coli* secE-nusG operon. *J. Bacteriol.* *172*, 1621–1627. <https://doi.org/10.1128/jb.172.3.1621-1627.1990>.
70. Ingham, C.J., Dennis, J., and Furneaux, P.A. (1999). Autogenous regulation of transcription termination factor Rho and the requirement for Nus factors in *Bacillus subtilis*. *Mol. Microbiol.* *31*, 651–663. <https://doi.org/10.1046/j.1365-2958.1999.01205.x>.
71. D’Heygère, F., Rabhi, M., and Boudvillain, M. (2013). Phyletic distribution and conservation of the bacterial transcription termination factor Rho. *Microbiology* *159*, 1423–1436. <https://doi.org/10.1099/mic.0.067462-0>.
72. Johnson, G.E., Lalanne, J.-B., Peters, M.L., and Li, G.-W. (2020). Functionally uncoupled transcription–translation in *Bacillus subtilis*. *Nature* *585*, 124–128. <https://doi.org/10.1038/s41586-020-2638-5>.
73. Saxena, S., and Gowrishankar, J. (2011). Compromised Factor-Dependent Transcription Termination in a nusA Mutant of *Escherichia coli*: Spectrum of Termination Efficiencies Generated by Perturbations of Rho, NusG, NusA, and H-NS Family Proteins. *J. Bacteriol.* *193*, 3842–3850. <https://doi.org/10.1128/jb.00221-11>.
74. Artsimovitch, I., Svetlov, V., Anthony, L., Burgess, R.R., and Landick, R. (2000). RNA polymerases from *Bacillus subtilis* and *Escherichia coli* differ in recognition of regulatory signals in vitro. *J. Bacteriol.* *182*, 6027–6035. <https://doi.org/10.1128/jb.182.21.6027-6035.2000>.
75. Yakhnin, A.V., Kashlev, M., and Babitzke, P. (2020). NusG-dependent RNA polymerase pausing is a frequent function of this universally conserved transcription elongation factor. *Crit. Rev. Biochem. Mol. Biol.* *55*, 716–728. <https://doi.org/10.1080/10409238.2020.1828261>.
76. Belogurov, G.A., Mooney, R.A., Svetlov, V., Landick, R., and Artsimovitch, I. (2009). Functional specialization of transcription elongation factors. *Embo J* *28*, 112–122. <https://doi.org/10.1038/emboj.2008.268>.
77. Sevostyanova, A., Svetlov, V., Vassilyev, D.G., and Artsimovitch, I. (2008). The elongation factor RfaH and the initiation factor σ bind to the same site on the transcription elongation complex. *Proc. Natl. Acad. Sci.* *105*, 865–870. <https://doi.org/10.1073/pnas.0708432105>.
78. Turnbough, C.L. (2019). Regulation of Bacterial Gene Expression by Transcription Attenuation. *Microbiol. Mol. Biol. Rev.* *83*, 10.1128/mmbr.00019-19. <https://doi.org/10.1128/mmbr.00019-19>.
79. Merino, E., and Yanofsky, C. (2005). Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.* *21*, 260–264. <https://doi.org/10.1016/j.tig.2005.03.002>.

80. Montange, R.K., and Batey, R.T. (2008). Riboswitches: Emerging Themes in RNA Structure and Function. *Annu. Rev. Biophys.* 37, 117–133.
<https://doi.org/10.1146/annurev.biophys.37.032807.130000>.
81. Winkler, W.C. (2005). Metabolic monitoring by bacterial mRNAs. *Arch. Microbiol.* 183, 151–159. <https://doi.org/10.1007/s00203-005-0758-9>.
82. Winkler, W.C., and Breaker, R.R. (2005). REGULATION OF BACTERIAL GENE EXPRESSION BY RIBOSWITCHES. *Annu. Rev. Microbiol.* 59, 487–517.
<https://doi.org/10.1146/annurev.micro.59.030804.121336>.
83. Winkler, W.C. (2005). Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr. Opin. Chem. Biol.* 9, 594–602. <https://doi.org/10.1016/j.cbpa.2005.09.016>.
84. Landick, R., Carey, J., and Yanofsky, C. (1985). Translation activates the paused transcription complex and restores transcription of the trp operon leader region. *Proc. Natl. Acad. Sci.* 82, 4663–4667. <https://doi.org/10.1073/pnas.82.14.4663>.
85. Santangelo, T.J., and Artsimovitch, I. (2011). Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.* 9, 319–329.
<https://doi.org/10.1038/nrmicro2560>.
86. Weisberg, R.A., and Gottesman, M.E. (1999). Processive Antitermination. *J. Bacteriol.* 181, 359–367. <https://doi.org/10.1128/jb.181.2.359-367.1999>.
87. Greenblatt, J., Nodwell, J.R., and Mason, S.W. (1993). Transcriptional antitermination. *Nature* 364, 401–406. <https://doi.org/10.1038/364401a0>.
88. Mandell, Z.F., Vishwakarma, R.K., Yakhnin, H., Murakami, K.S., Kashlev, M., and Babitzke, P. (2022). Comprehensive transcription terminator atlas for *Bacillus subtilis*. *Nat Microbiol* 7, 1918–1931. <https://doi.org/10.1038/s41564-022-01240-7>.
89. Goodson, J.R., and Winkler, W.C. (2018). Processive Antitermination. *Microbiol. Spectr.* 6, 10.1128/microbiolspec.rwr-0031–2018. <https://doi.org/10.1128/microbiolspec.rwr-0031-2018>.
90. Patterson, T.A., Zhang, Z., Baker, T., Johnson, L.L., Friedman, D.I., and Court, D.L. (1994). Bacteriophage Lambda N-Dependent Transcription Antitermination. *J. Mol. Biol.* 236, 217–228.
<https://doi.org/10.1006/jmbi.1994.1131>.
91. Krupp, F., Said, N., Huang, Y.-H., Loll, B., Bürger, J., Mielke, T., Spahn, C.M.T., and Wahl, M.C. (2019). Structural Basis for the Action of an All-Purpose Transcription Anti-termination Factor. *Mol. Cell* 74, 143–157.e5. <https://doi.org/10.1016/j.molcel.2019.01.016>.
92. Said, N., Krupp, F., Anedchenko, E., Santos, K.F., Dybkov, O., Huang, Y.-H., Lee, C.-T., Loll, B., Behrmann, E., Bürger, J., et al. (2016). Structural basis for λ N-dependent processive

transcription antitermination. *Nat. Microbiol.* 2, 17062.

<https://doi.org/10.1038/nmicrobiol.2017.62>.

93. Mogridge, J., Mah, T.-F., and Greenblatt, J. (1998). Involvement of boxA Nucleotides in the Formation of a Stable Ribonucleoprotein Complex Containing the Bacteriophage λ N Protein*. *J. Biol. Chem.* 273, 4143–4148. <https://doi.org/10.1074/jbc.273.7.4143>.

94. Gusarov, I., and Nudler, E. (2001). Control of Intrinsic Transcription Termination by N and NusA. *Cell* 107, 437–449. [https://doi.org/10.1016/s0092-8674\(01\)00582-7](https://doi.org/10.1016/s0092-8674(01)00582-7).

95. Thapar, R., Denmon, A.P., and Nikonowicz, E.P. (2014). Recognition modes of RNA tetraloops and tetraloop-like motifs by RNA-binding proteins. *Wiley Interdiscip. Rev.: RNA* 5, 49–67. <https://doi.org/10.1002/wrna.1196>.

96. Legault, P., Li, J., Mogridge, J., Kay, L.E., and Greenblatt, J. (1998). NMR Structure of the Bacteriophage λ N Peptide/boxB RNA Complex: Recognition of a GNRA Fold by an Arginine-Rich Motif. *Cell* 93, 289–299. [https://doi.org/10.1016/s0092-8674\(00\)81579-2](https://doi.org/10.1016/s0092-8674(00)81579-2).

97. Conant, C.R., Goodarzi, J.P., Weitzel, S.E., and Hippel, P.H. von (2008). The Antitermination Activity of Bacteriophage λ N Protein is Controlled by the Kinetics of an RNA-Looping-Facilitated Interaction with the Transcription Complex. *J. Mol. Biol.* 384, 87–108. <https://doi.org/10.1016/j.jmb.2008.05.014>.

98. 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. [https://doi.org/10.1016/0092-8674\(93\)90665-d](https://doi.org/10.1016/0092-8674(93)90665-d).

99. Das, A. (1992). How the phage lambda N gene product suppresses transcription termination: communication of RNA polymerase with regulatory proteins mediated by signals in nascent RNA. *J. Bacteriol.* 174, 6711–6716. <https://doi.org/10.1128/jb.174.21.6711-6716.1992>.

100. Das, A. (1993). Control of Transcription Termination by RNA-Binding Proteins. *Annu. Rev. Biochem.* 62, 893–930. <https://doi.org/10.1146/annurev.bi.62.070193.004333>.

101. Mason, S.W., Li, J., and Greenblatt, J. (1992). 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. [https://doi.org/10.1016/s0021-9258\(18\)41792-9](https://doi.org/10.1016/s0021-9258(18)41792-9).

102. Rees, W.A., Weitzel, S.E., Das, A., and Hippel, P.H. von (1997). Regulation of the elongation-termination decision at intrinsic terminators by antitermination protein N of phage lambda. *J. Mol. Biol.* 273, 797–813. <https://doi.org/10.1006/jmbi.1997.1327>.

103. Nudler, E., and Gottesman, M.E. (2002). Transcription termination and anti-termination in *E. coli*. *Genes Cells* 7, 755–768. <https://doi.org/10.1046/j.1365-2443.2002.00563.x>.

104. Liu, K., Zhang, Y., Severinov, K., Das, A., and Hanna, M.M. (1996). Role of *Escherichia coli* RNA polymerase alpha subunit in modulation of pausing, termination and anti-termination by the transcription elongation factor NusA. *EMBO J.* 15, 150–161. <https://doi.org/10.1002/j.1460-2075.1996.tb00343.x>.
105. Valabhoju, V., Agrawal, S., and Sen, R. (2016). Molecular Basis of NusG-mediated Regulation of Rho-dependent Transcription Termination in Bacteria*. *J Biol Chem* 291, 22386–22403. <https://doi.org/10.1074/jbc.m116.745364>.
106. Herskowitz, I., and Signer, E.E. (1970). A site essential for expression of all late genes in bacteriophage λ . *J. Mol. Biol.* 47, 545–556. [https://doi.org/10.1016/0022-2836\(70\)90321-9](https://doi.org/10.1016/0022-2836(70)90321-9).
107. Yarnell, W.S., and Roberts, J.W. (1992). The phage λ gene Q transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. *Cell* 69, 1181–1189. [https://doi.org/10.1016/0092-8674\(92\)90639-t](https://doi.org/10.1016/0092-8674(92)90639-t).
108. Li, S.C., Squires, C.L., and 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. [https://doi.org/10.1016/0092-8674\(84\)90280-0](https://doi.org/10.1016/0092-8674(84)90280-0).
109. Berg, K.L., Squires, C., and Squires, C.L. (1989). Ribosomal RNA operon anti-termination Function of leader and spacer region box B-box A sequences and their conservation in diverse micro-organisms. *J. Mol. Biol.* 209, 345–358. [https://doi.org/10.1016/0022-2836\(89\)90002-8](https://doi.org/10.1016/0022-2836(89)90002-8).
110. 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.* 90, 970–974. <https://doi.org/10.1073/pnas.90.3.970>.
111. Singh, N., Bubunenko, M., Smith, C., Abbott, D.M., Stringer, A.M., Shi, R., Court, D.L., and Wade, J.T. (2016). SuhB Associates with Nus Factors To Facilitate 30S Ribosome Biogenesis in *Escherichia coli*. *mBio* 7, 10.1128/mbio.00114-16. <https://doi.org/10.1128/mbio.00114-16>.
112. Dudenhoefter, B.R., Schneider, H., Schweimer, K., and Knauer, S.H. (2019). SuhB is an integral part of the ribosomal antitermination complex and interacts with NusA. *Nucleic Acids Res.* 47, 6504–6518. <https://doi.org/10.1093/nar/gkz442>.
113. Huang, Y.-H., Hilal, T., Loll, B., Bürger, J., Mielke, T., Böttcher, C., Said, N., and Wahl, M.C. (2020). Structure-Based Mechanisms of a Molecular RNA Polymerase/Chaperone Machine Required for Ribosome Biosynthesis. *Mol Cell* 79, 1024-1036.e5. <https://doi.org/10.1016/j.molcel.2020.08.010>.
114. King, R.A., Banik-Maiti, S., Jin, D.J., and Weisberg, R.A. (1996). Transcripts That Increase the Processivity and Elongation Rate of RNA Polymerase. *Cell* 87, 893–903. [https://doi.org/10.1016/s0092-8674\(00\)81996-0](https://doi.org/10.1016/s0092-8674(00)81996-0).

115. Clerget, M., Jin, D.J., and Weisberg, R.A. (1995). A Zinc-binding Region in the β' Subunit of RNA Polymerase is Involved in Antitermination of Early Transcription of Phage HK022. *J. Mol. Biol.* 248, 768–780. <https://doi.org/10.1006/jmbi.1995.0259>.
116. Banik-Maiti, S., King, R.A., and Weisberg, R.A. (1997). The antiterminator RNA of phage HK022. *J. Mol. Biol.* 272, 677–687. <https://doi.org/10.1006/jmbi.1997.1251>.
117. King, R.A., and Weisberg, R.A. (2003). Suppression of factor-dependent transcription termination by antiterminator RNA. *J. Bacteriol.* 185, 7085–7091. <https://doi.org/10.1128/jb.185.24.7085-7091.2003>.
118. Robert, J., Sloan, S.B., Weisberg, R.A., Gottesman, M.E., Robledo, R., and Harbrecht, D. (1987). The remarkable specificity of a new transcription termination factor suggests that the mechanisms of termination and antitermination are similar. *Cell* 51, 483–492. [https://doi.org/10.1016/0092-8674\(87\)90644-1](https://doi.org/10.1016/0092-8674(87)90644-1).
119. Hung, S.C., and Gottesman, M.E. (1997). The Nun protein of bacteriophage HK022 inhibits translocation of Escherichia coli RNA polymerase without abolishing its catalytic activities. *Genes Dev.* 11, 2670–2678. <https://doi.org/10.1101/gad.11.20.2670>.
120. Vitiello, C.L., Kireeva, M.L., Lubkowska, L., Kashlev, M., and Gottesman, M. (2014). Coliphage HK022 Nun protein inhibits RNA polymerase translocation. *Proc. Natl. Acad. Sci.* 111, E2368–E2375. <https://doi.org/10.1073/pnas.1319740111>.
121. Oberto, J., Clerget, M., Ditto, M., Cam, K., and Weisberg, R.A. (1993). Antitermination of Early Transcription in Phage HK022 Absence of a Phage-encoded Antitermination Factor. *J. Mol. Biol.* 229, 368–381. <https://doi.org/10.1006/jmbi.1993.1040>.
122. Irnov, I., and Winkler, W.C. (2010). A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. *Mol Microbiol* 76, 559–575. <https://doi.org/10.1111/j.1365-2958.2010.07131.x>.
123. 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. <https://doi.org/10.1016/j.cell.2012.05.042>.
124. Leeds, J.A., and Welch, R.A. (1996). RfaH enhances elongation of Escherichia coli hlyCABD mRNA. *J. Bacteriol.* 178, 1850–1857. <https://doi.org/10.1128/jb.178.7.1850-1857.1996>.
125. Bailey, M.J.A., Hughes, C., and Koronakis, V. (1997). RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. *Mol. Microbiol.* 26, 845–851. <https://doi.org/10.1046/j.1365-2958.1997.6432014.x>.

126. Bailey, M.J.A., Koronakis, V., Schmoll, T., and Hughes, C. (1992). *Escherichia coli* HlyT protein, a transcriptional activator of haemolysin synthesis and secretion, is encoded by the *rfaH* (*sfrB*) locus required for expression of sex factor and lipopolysaccharide genes. *Mol. Microbiol.* 6, 1003–1012. <https://doi.org/10.1111/j.1365-2958.1992.tb02166.x>.
127. Bailey, M.J.A., Hughes, C., and Koronakis, V. (1996). Increased distal gene transcription by the elongation factor RfaH, a specialized homologue of NusG. *Mol. Microbiol.* 22, 729–737. <https://doi.org/10.1046/j.1365-2958.1996.d01-1726.x>.
128. Zuber, P.K., Schweimer, K., Rösch, P., Artsimovitch, I., and Knauer, S.H. (2019). Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. *Nat Commun* 10, 702. <https://doi.org/10.1038/s41467-019-08567-6>.
129. 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. [https://doi.org/10.1016/s0092-8674\(02\)00724-9](https://doi.org/10.1016/s0092-8674(02)00724-9).
130. 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: Struct., Funct., Bioinform.* 56, 40–51. <https://doi.org/10.1002/prot.20054>.
131. Liu, B., and Steitz, T.A. (2017). Structural insights into NusG regulating transcription elongation. *Nucleic Acids Res.* 45, 968–974. <https://doi.org/10.1093/nar/gkw1159>.
132. Svetlov, V., Belogurov, G.A., Shabrova, E., Vassylyev, D.G., and Artsimovitch, I. (2007). Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic Acids Res.* 35, 5694–5705. <https://doi.org/10.1093/nar/gkm600>.
133. Turtola, M., and Belogurov, G.A. (2016). NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *eLife* 5, e18096. <https://doi.org/10.7554/elife.18096>.
134. Nedialkov, Y., Svetlov, D., Belogurov, G.A., and Artsimovitch, I. (2018). Locking the nontemplate DNA to control transcription. *Mol. Microbiol.* 109, 445–457. <https://doi.org/10.1111/mmi.13983>.
135. Zuber, P.K., Artsimovitch, I., NandyMazumdar, M., Liu, Z., Nedialkov, Y., Schweimer, K., Rösch, P., and Knauer, S.H. (2018). The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. *Elife* 7, e36349. <https://doi.org/10.7554/elife.36349>.
136. Shi, D., Svetlov, D., Abagyan, R., and Artsimovitch, I. (2017). Flipping states: a few key residues decide the winning conformation of the only universally conserved transcription factor. *Nucleic Acids Res.* 45, 8835–8843. <https://doi.org/10.1093/nar/gkx523>.

137. Hu, K., and Artsimovitch, I. (2017). A Screen for *rfaH* Suppressors Reveals a Key Role for a Connector Region of Termination Factor Rho. *mBio* 8, e00753-17. <https://doi.org/10.1128/mbio.00753-17>.
138. Wang, B., Gumerov, V.M., Andrianova, E.P., Zhulin, I.B., and Artsimovitch, I. (2020). Origins and Molecular Evolution of the NusG Paralog RfaH. *Mbio* 11. <https://doi.org/10.1128/mbio.02717-20>.
139. Chatzidaki-Livanis, M., Coyne, M.J., and Comstock, L.E. (2009). A Family of Transcriptional Antitermination Factors Necessary for Synthesis of the Capsular Polysaccharides of *Bacteroides fragilis*. *J Bacteriol* 191, 7288–7295. <https://doi.org/10.1128/jb.00500-09>.
140. Chatzidaki-Livanis, M., Weinacht, K.G., and Comstock, L.E. (2010). Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont *Bacteroides fragilis*. *Proc. Natl. Acad. Sci.* 107, 11976–11980. <https://doi.org/10.1073/pnas.1005039107>.
141. Saba, J., Flores, K., Marshall, B., Engstrom, M.D., Peng, Y., Garje, A.S., Comstock, L.E., and Landick, R. (2024). *Bacteroides* expand the functional versatility of a conserved transcription factor and transcribed DNA to program capsule diversity. *Nat. Commun.* 15, 10862. <https://doi.org/10.1038/s41467-024-55215-9>.
142. Varon, M., Fuchs, N., Monosov, M., Tolchinsky, S., and Rosenberg, E. (1992). Mutation and mapping of genes involved in production of the antibiotic TA in *Myxococcus xanthus*. *Antimicrob. Agents Chemother.* 36, 2316–2321. <https://doi.org/10.1128/aac.36.10.2316>.
143. Paitan, Y., Orr, E., Ron, E.Z., and Rosenberg, E. (1999). A NusG-like transcription anti-terminator is involved in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus*. *FEMS Microbiol. Lett.* 170, 221–227. <https://doi.org/10.1111/j.1574-6968.1999.tb13377.x>.
144. Núñez, B., Avila, P., and Cruz, F.D.L. (1997). Genes involved in conjugative DNA processing of plasmid R6K. *Mol. Microbiol.* 24, 1157–1168. <https://doi.org/10.1046/j.1365-2958.1997.4111778.x>.
145. Arutyunov, D., Arenson, B., Manchak, J., and Frost, L.S. (2010). F Plasmid TraF and TraH Are Components of an Outer Membrane Complex Involved in Conjugation. *J. Bacteriol.* 192, 1730–1734. <https://doi.org/10.1128/jb.00726-09>.
146. NandyMazumdar, M., and Artsimovitch, I. (2015). Ubiquitous transcription factors display structural plasticity and diverse functions. *BioEssays* 37, 324–334. <https://doi.org/10.1002/bies.201400177>.
147. Jones, C.S., Osborne, D.J., and Stanley, J. (1993). Molecular comparison of the IncX plasmids allows division into IncX1 and IncX2 subgroups. *Microbiology* 139, 735–741. <https://doi.org/10.1099/00221287-139-4-735>.

148. Goodson, J.R., Klupt, S., Zhang, C., Straight, P., and Winkler, W.C. (2017). LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in *Bacillus amyloliquefaciens*. *Nat. Microbiol.* 2, 17003. <https://doi.org/10.1038/nmicrobiol.2017.3>.
149. Elghondakly, A., Wu, C.H., Klupt, S., Goodson, J., and Winkler, W.C. (2021). A NusG Specialized Paralog That Exhibits Specific, High-Affinity RNA-Binding Activity. *J. Mol. Biol.* 433, 167100. <https://doi.org/10.1016/j.jmb.2021.167100>.
150. Yakhnin, A.V., and Babitzke, P. (2014). NusG/Spt5: are there common functions of this ubiquitous transcription elongation factor? *Curr. Opin. Microbiol.* 18, 68–71. <https://doi.org/10.1016/j.mib.2014.02.005>.
151. Bailey, E.J., Gottesman, M.E., and Gonzalez, R.L. (2022). NusG-mediated coupling of transcription and translation enhances gene expression by suppressing RNA polymerase backtracking. *J Mol Biol* 434, 167330. <https://doi.org/10.1016/j.jmb.2021.167330>.
152. Elghondakly, A., Jermain, M.D., Winkler, W.C., and Ferré-D'Amaré, A.R. (2024). Major-groove sequence-specific RNA recognition by LoaP, a paralog of transcription elongation factor NusG. *Structure* 32, 1488-1497.e5. <https://doi.org/10.1016/j.str.2024.06.001>.
153. Woese, C.R., Winker, S., and Gutell, R.R. (1990). Architecture of ribosomal RNA: constraints on the sequence of “tetra-loops”. *Proc. Natl. Acad. Sci.* 87, 8467–8471. <https://doi.org/10.1073/pnas.87.21.8467>.
154. Donaldson, G.P., Roelofs, K.G., Luo, Y., Sintim, H.O., and Lee, V.T. (2012). A rapid assay for affinity and kinetics of molecular interactions with nucleic acids. *Nucleic Acids Res.* 40, e48–e48. <https://doi.org/10.1093/nar/gkr1299>.
155. Roelofs, K.G., Wang, J., Sintim, H.O., and Lee, V.T. (2011). Differential radial capillary action of ligand assay for high-throughput detection of protein-metabolite interactions. *Proc National Acad Sci* 108, 15528–15533. <https://doi.org/10.1073/pnas.1018949108>.
156. Patel, D.K., Gebbie, M.P., and Lee, V.T. (2014). Assessing RNA interactions with proteins by DRaCALA. *Methods Enzym.* 549, 489–512. <https://doi.org/10.1016/b978-0-12-801122-5.00021-0>.
157. Molodtsov, V., Wang, C., Zhang, J., Kaelber, J.T., Blaha, G., and Ebright, R.H. (2024). Structural basis of RfaH-mediated transcription–translation coupling. *Nat. Struct. Mol. Biol.* 31, 1932–1941. <https://doi.org/10.1038/s41594-024-01372-w>.
158. Seeman, N.C., Rosenberg, J.M., and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci.* 73, 804–808. <https://doi.org/10.1073/pnas.73.3.804>.

159. Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. *Genome Biol.* *1*, reviews001.1-reviews001.37. <https://doi.org/10.1186/gb-2000-1-1-reviews001>.
160. Saenger, W. (1984). Principles of Nucleic Acid Structure. Springer Advanced Texts in Chemistry, 105–115. https://doi.org/10.1007/978-1-4612-5190-3_5.
161. Weeks, K.M., and Crothers, D.M. (1993). Major Groove Accessibility of RNA. *Science* *261*, 1574–1577. <https://doi.org/10.1126/science.7690496>.
162. Masliah, G., Barraud, P., and Allain, F.H.-T. (2013). RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell. Mol. Life Sci.* *70*, 1875–1895. <https://doi.org/10.1007/s00018-012-1119-x>.
163. Corley, M., Burns, M.C., and Yeo, G.W. (2020). How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol. Cell* *78*, 9–29. <https://doi.org/10.1016/j.molcel.2020.03.011>.
164. Bernhardt, N.A., and Hansmann, U.H.E. (2018). Multifunnel Landscape of the Fold-Switching Protein RfaH-CTD. *J. Phys. Chem. B* *122*, 1600–1607. <https://doi.org/10.1021/acs.jpcc.7b11352>.
165. Porter, L.L., Kim, A.K., Rimal, S., Looger, L.L., Majumdar, A., Mensh, B.D., Starich, M.R., and Strub, M.-P. (2022). Many dissimilar NusG protein domains switch between α -helix and β -sheet folds. *Nat. Commun.* *13*, 3802. <https://doi.org/10.1038/s41467-022-31532-9>.
166. Galaz-Davison, P., Román, E.A., and Ramírez-Sarmiento, C.A. (2021). The N-terminal domain of RfaH plays an active role in protein fold-switching. *PLoS Comput. Biol.* *17*, e1008882. <https://doi.org/10.1371/journal.pcbi.1008882>.
167. 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. <https://doi.org/10.1016/j.jmb.2012.01.031>.
168. Kang, J.Y., Mishanina, T.V., Bellecourt, M.J., Mooney, R.A., Darst, S.A., and Landick, R. (2018). RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. *Mol. Cell* *69*, 802–815.e1. <https://doi.org/10.1016/j.molcel.2018.01.018>.
169. Yakhnin, H., Yakhnin, A.V., Mouery, B.L., Mandell, Z.F., Karbasiafshar, C., Kashlev, M., and Babitzke, P. (2019). NusG-Dependent RNA Polymerase Pausing and Tylosin-Dependent Ribosome Stalling Are Required for Tylosin Resistance by Inducing 23S rRNA Methylation in *Bacillus subtilis*. *mBio* *10*, 10.1128/mbio.02665-19. <https://doi.org/10.1128/mbio.02665-19>.
170. Nudler, E. (2012). RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell* *149*, 1438–1445. <https://doi.org/10.1016/j.cell.2012.06.003>.

171. Frey, S., and Görlich, D. (2014). A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. *J. Chromatogr. A* *1337*, 95–105. <https://doi.org/10.1016/j.chroma.2014.02.029>.
172. Burgess, R.R., and Jendrisak, J.J. (1975). Procedure for the rapid, large-scale purification of *Escherichia coli* DNA-dependent RNA polymerase involving polymyxin P precipitation and DNA-cellulose chromatography. *Biochemistry* *14*, 4634–4638. <https://doi.org/10.1021/bi00692a011>.
173. Burgess, R.R. (1991). [1] Use of polyethyleneimine in purification of DNA-binding proteins. *Methods Enzym.* *208*, 3–10. [https://doi.org/10.1016/0076-6879\(91\)08003-z](https://doi.org/10.1016/0076-6879(91)08003-z).
174. Burgess, R.R. (2009). Chapter 20 Protein Precipitation Techniques. *Methods Enzym.* *463*, 331–342. [https://doi.org/10.1016/s0076-6879\(09\)63020-2](https://doi.org/10.1016/s0076-6879(09)63020-2).
175. Scopes, R.K. (1994). Protein Purification, Principles and Practice. 293–309. https://doi.org/10.1007/978-1-4757-2333-5_11.
176. England, S., and Seifter, S. (1990). [22] Precipitation techniques. *Methods Enzym.* *182*, 285–300. [https://doi.org/10.1016/0076-6879\(90\)82024-v](https://doi.org/10.1016/0076-6879(90)82024-v).
177. Yang, X., and Lewis, P.J. (2008). Overproduction and purification of recombinant *Bacillus subtilis* RNA polymerase. *Protein Expr. Purif.* *59*, 86–93. <https://doi.org/10.1016/j.pep.2008.01.006>.
178. Mondal, S., Yakhnin, A.V., and Babitzke, P. (2017). Modular Organization of the NusA- and NusG-Stimulated RNA Polymerase Pause Signal That Participates in the *Bacillus subtilis* trp Operon Attenuation Mechanism. *J. Bacteriol.* *199*, 10.1128/jb.00223-17. <https://doi.org/10.1128/jb.00223-17>.
179. Yakhnin, A.V., Yakhnin, H., and Babitzke, P. (2006). RNA Polymerase Pausing Regulates Translation Initiation by Providing Additional Time for TRAP-RNA Interaction. *Mol. Cell* *24*, 547–557. <https://doi.org/10.1016/j.molcel.2006.09.018>.
180. Levin, J.R., Krummel, B., and Chamberlin, M.J. (1987). Isolation and properties of transcribing ternary complexes of *Escherichia coli* RNA polymerase positioned at a single template base. *J. Mol. Biol.* *196*, 85–100. [https://doi.org/10.1016/0022-2836\(87\)90512-2](https://doi.org/10.1016/0022-2836(87)90512-2).
181. Yanofsky, C. (1981). Attenuation in the control of expression of bacterial operons. *Nature* *289*, 751–758. <https://doi.org/10.1038/289751a0>.
182. Mitra, P., Ghosh, G., Hafeezunnisa, Md., and Sen, R. (2017). Rho Protein: Roles and Mechanisms. *Annual Review of Microbiology*.
183. Drögemüller, J., Strauß, M., Schweimer, K., Jurk, M., Rösch, P., and Knauer, S.H. (2015). Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci. Rep.* *5*, 16428. <https://doi.org/10.1038/srep16428>.

184. 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. <https://doi.org/10.1016/j.molcel.2011.05.026>.
185. Belogurov, G.A., Vassilyeva, M.N., Svetlov, V., Klyuyev, S., Grishin, N.V., Vassilyev, 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. <https://doi.org/10.1016/j.molcel.2007.02.021>.
186. Chen, X.-H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Koumoutsis, A., Hitzeroth, G., Grammel, N., Strittmatter, A.W., et al. (2006). Structural and Functional Characterization of Three Polyketide Synthase Gene Clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol.* 188, 4024–4036. <https://doi.org/10.1128/jb.00052-06>.
187. Fan, B., Li, L., Chao, Y., Förstner, K., Vogel, J., Borriss, R., and Wu, X.-Q. (2015). dRNA-Seq Reveals Genomewide TSSs and Noncoding RNAs of Plant Beneficial Rhizobacterium *Bacillus amyloliquefaciens* FZB42. *PLoS ONE* 10, e0142002. <https://doi.org/10.1371/journal.pone.0142002>.
188. Shankar, S., Hatoum, A., and Roberts, J.W. (2007). A Transcription Antiterminator Constructs a NusA-Dependent Shield to the Emerging Transcript. *Mol. Cell* 27, 914–927. <https://doi.org/10.1016/j.molcel.2007.07.025>.
189. Jayasinghe, O.T., Mandell, Z.F., Yakhnin, A.V., Kashlev, M., and Babitzke, P. (2022). Transcriptome-Wide Effects of NusA on RNA Polymerase Pausing in *Bacillus subtilis*. *J Bacteriol* 204, e00534-21. <https://doi.org/10.1128/jb.00534-21>.
190. Ferrero-Bordera, B., Bartel, J., Dijl, J.M. van, Becher, D., and Maaß, S. (2024). From the outer space to the inner cell: deconvoluting the complexity of *Bacillus subtilis* disulfide stress responses by redox state and absolute abundance quantification of extracellular, membrane, and cytosolic proteins. *Microbiol. Spectr.* 12, e02616-23. <https://doi.org/10.1128/spectrum.02616-23>.
191. Guo, X., Myasnikov, A.G., Chen, J., Crucifix, C., Papai, G., Takacs, M., Schultz, P., and Weixlbaumer, A. (2018). Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol. Cell* 69, 816-827.e4. <https://doi.org/10.1016/j.molcel.2018.02.008>.
192. Zhang, J., and Landick, R. (2016). A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem. Sci.* 41, 293–310. <https://doi.org/10.1016/j.tibs.2015.12.009>.
193. Yakhnin, A.V., and Babitzke, P. (2010). Mechanism of NusG-stimulated pausing, hairpin-dependent pause site selection and intrinsic termination at overlapping pause and termination sites in the *Bacillus subtilis* *trp* leader. *Mol Microbiol* 76, 690–705. <https://doi.org/10.1111/j.1365-2958.2010.07126.x>.

194. Greenblatt, J., McLimont, M., and Hanly, S. (1981). Termination of transcription by nusA gene protein of *Escherichia coli*. *Nature* 292, 215–220. <https://doi.org/10.1038/292215a0>.
195. Strauß, M., Vitiello, C., Schweimer, K., Gottesman, M., Rösch, P., and Knauer, S.H. (2016). Transcription is regulated by NusA:NusG interaction. *Nucleic Acids Res* 44, 5971–5982. <https://doi.org/10.1093/nar/gkw423>.
196. Miguel-Arribas, A., Val-Calvo, J., Gago-Córdoba, C., Izquierdo, J.M., Abia, D., Wu, L.J., Errington, J., and Meijer, W.J.J. (2021). A novel bipartite antitermination system widespread in conjugative elements of Gram-positive bacteria. *Nucleic Acids Res.* 49, 5553–5567. <https://doi.org/10.1093/nar/gkab360>.
197. Artsimovitch, I., and Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc National Acad Sci* 97, 7090–7095. <https://doi.org/10.1073/pnas.97.13.7090>.
198. Wang, C., Molodtsov, V., Firlar, E., Kaelber, J.T., Blaha, G., Su, M., and Ebright, R.H. (2020). Structural basis of transcription-translation coupling. *bioRxiv*, 2020.03.01.972380. <https://doi.org/10.1101/2020.03.01.972380>.
199. Behnken, S., and Hertweck, C. (2012). Anaerobic bacteria as producers of antibiotics. *Appl Microbiol Biot* 96, 61–67. <https://doi.org/10.1007/s00253-012-4285-8>.
200. Dunbar, K.L., Büttner, H., Molloy, E.M., Dell, M., Kumpfmüller, J., and Hertweck, C. (2018). Genome Editing Reveals Novel Thiotemplated Assembly of Polythioamide Antibiotics in Anaerobic Bacteria. *Angew. Chem. Int. Ed.* 57, 14080–14084. <https://doi.org/10.1002/anie.201807970>.
201. Dunbar, K.L., Dell, M., Molloy, E.M., Büttner, H., Kumpfmüller, J., and Hertweck, C. (2021). An Unexpected Split-Merge Pathway in the Assembly of the Symmetric Nonribosomal Peptide Antibiotic Closthoamide. *Angewandte Chemie Int Ed Engl* 60, 4104–4109. <https://doi.org/10.1002/anie.202011741>.
202. Dunbar, K.L., Dell, M., Molloy, E.M., Kloss, F., and Hertweck, C. (2019). Reconstitution of Iterative Thioamidation in Closthoamide Biosynthesis Reveals Tailoring Strategy for Nonribosomal Peptide Backbones. *Angewandte Chemie Int Ed Engl* 58, 13014–13018. <https://doi.org/10.1002/anie.201905992>.
203. Peters, L., Drechsler, M., Herrera, M.A., Liu, J., Pees, B., Jarstorff, J., Czerwinski, A., Lubbock, F., Angelidou, G., Salzer, L., et al. (2025). Polyketide synthase-derived sphingolipids mediate microbiota protection against a bacterial pathogen in *C. elegans*. *Nat. Commun.* 16, 5151. <https://doi.org/10.1038/s41467-025-60234-1>.
204. Yin, Q.-J., Ying, T.-T., Zhou, Z.-Y., Hu, G.-A., Yang, C.-L., Hua, Y., Wang, H., and Wei, B. (2023). Species-specificity of the secondary biosynthetic potential in *Bacillus*. *Front. Microbiol.* 14, 1271418. <https://doi.org/10.3389/fmicb.2023.1271418>.

205. Li, Z., Song, C., Yi, Y., and Kuipers, O.P. (2020). Characterization of plant growth-promoting rhizobacteria from perennial ryegrass and genome mining of novel antimicrobial gene clusters. *BMC Genom.* 21, 157. <https://doi.org/10.1186/s12864-020-6563-7>.
206. Timmusk, S., Kim, S.-B., Nevo, E., Daim, I.A.E., Ek, B., Bergquist, J., and Behers, L. (2015). Sfp-type PPTase inactivation promotes bacterial biofilm formation and ability to enhance wheat drought tolerance. *Front. Microbiol.* 6, 387. <https://doi.org/10.3389/fmicb.2015.00387>.
207. Baniulyte, G., Singh, N., Benoit, C., Johnson, R., Ferguson, R., Paramo, M., Stringer, A.M., Scott, A., Lapierre, P., and Wade, J.T. (2017). Identification of regulatory targets for the bacterial Nus factor complex. *Nat. Commun.* 8, 2027. <https://doi.org/10.1038/s41467-017-02124-9>.