HHS Public Access

Author manuscript

Microbiol Spectr. Author manuscript; available in PMC 2020 June 29.

Published in final edited form as:

Microbiol Spectr. 2018 September ; 6(5): . doi:10.1128/microbiolspec.RWR-0031-2018.

Processive Antitermination

Jonathan R. Goodson, **Wade C. Winkler**[1]

The University of Maryland, Department of Cell Biology and Molecular Genetics, College Park, MD 20742

Abstract

Transcription is a discontinuous process, where each nucleotide incorporation cycle offers a decision between elongation, pausing, halting, or termination. Many *cis*-acting regulatory RNAs, such as riboswitches, exert their influence over transcription elongation. Through such mechanisms, certain RNA elements can couple physiological or environmental signals to transcription attenuation, a process where *cis*-acting regulatory RNAs directly influence formation of transcription termination signals. However, through another regulatory mechanism called processive antitermination (PA), RNA polymerase can bypass termination sites over much greater distances than transcription attenuation. PA mechanisms are widespread in bacteria, although only a few classes have been discovered overall. Also, although traditional, signal-responsive riboswitches have not yet been discovered to promote PA, it is increasingly clear that small RNA elements are still oftentimes required. In some instances, small RNA elements serve as loading sites for cellular factors that promote PA. In other instances, larger, more complicated RNA elements participate in PA in unknown ways, perhaps even acting alone to trigger PA activity. These discoveries suggest that what is now needed is a systematic exploration of PA in bacteria, to determine how broadly these transcription elongation mechanisms are utilized, to reveal the diversity in their molecular mechanisms, and to understand the general logic behind their cellular applications. This review covers the known examples of PA regulatory mechanisms and speculates that they may be broadly important to bacteria.

INTRODUCTION

An extraordinarily diverse range of genetic regulatory mechanisms has been discovered in the half century since Francois Jacob and Jacques Monod first proposed the operon model of gene regulation (1). Studies based on this model identified a soluble regulator, located distally from the targeted operon, that acts to repress transcription initiation of the *lac* operon. This discovery led to the identification and characterization of many more repressor proteins, each acting in modestly different ways to reduce the efficiency of transcription initiation. Soon followed discoveries of other types of transcriptional regulators, including those that activate gene expression by enhancing transcription initiation. And now, in an era where bacterial genome sequences can be acquired and draft-annotated in mere days and at low cost, it is clear that all bacteria encode for dozens or hundreds of proteins that regulate

[1]Corresponding author information: The University of Maryland, Department of Cell Biology and Molecular Genetics, 4066 Campus Drive, 3112 Biosciences Research Building, College Park, MD 20742, wwinkler@umd.edu.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

transcription initiation and that this 'layer' of genetic regulation is both ubiquitous and profoundly important. However, perhaps because transcription initiation is so universally recognized as a key point of regulatory influence (2), later stages of transcription elongation have not yet been sufficiently analyzed for genetic regulation. While the molecular mechanisms of transcription have been, and continue to be, intensively investigated, the biological extent of post-initiation regulatory mechanisms has been incompletely analyzed. Transcription initiation is only the first stage of gene expression. The stages that follow include transcription elongation, transcription termination, translation and mRNA degradation; each of these stages can be subjected to genetic regulatory control (3).

While riboswitches, which control transcription attenuation in a signal-dependent manner, are widely used by bacteria, their initial discoveries have been significantly aided by the extensive conservation of their sequences and secondary structures (4, 5). This level of sequence conservation is not observed for many other types of transcription elongation regulatory strategies, a limitation that may have slowed discovery of the latter. How, then, may other transcription elongation-based regulatory strategies be systematically discovered if experimentalists cannot rely primarily on bioinformatics searches of highly conserved regulatory RNAs? And what kinds of transcription elongation regulatory mechanisms have not yet been found? One type of regulatory mechanism that might still be understudied, but yet has been identified through a variety of experimental approaches, is called processive antitermination (PA). These systems offer a convenient and powerful mechanism for altering the efficiency of transcription elongation (6–8).

In PA mechanisms, antitermination factors associate with a bacterial RNA polymerase (RNAP) elongation complex, leading to read-through of termination sites (6). Termination signals normally induce rapid dissociation of the transcription elongation complex (TEC) and are most often located at the ends of operons (9). However, when placed within operons, they can serve as key points of regulatory control (10). In bacteria, there are two known classes of termination signals: intrinsic and Rho-dependent terminators (9). In many bacteria intrinsic terminators consist of a GC-rich RNA hairpin followed by a poly-uridine tract. Alone (11), or enhanced by a factor such as NusA (12, 13), these RNA elements promote pausing of the TEC, followed by release of the nascent transcript and dissociation of polymerase (14). In contrast, Rho-dependent termination depends upon the adenosine triphosphate (ATP)-dependent translocase Rho associating with Rho-utilization (*rut*) sites on a nascent mRNA and translocating the RNA to eventually promote TEC dissociation (15, 16). Both classes of termination sites may be specifically regulated by signal-responsive riboswitches (5, 17) or trans-encoded small RNAs (18, 19). However, whereas riboswitches exert control over a single intrinsic terminator site, or a particular entry point for Rho, PA systems differ in that they modify TECs to render them generally resistant to downstream termination sites (8). PA systems, therefore, are capable of causing read-through of multiple termination sites, even over long genomic distances. While only a few classes of PA mechanisms have been discovered in the past four decades, they vary widely in the molecular mechanisms they utilize and in their biological applications. Several new examples of PA mechanisms have been discovered more recently, which appear to be broadly used by bacteria for regulation of diverse sets of genes. We extrapolate from these discoveries that many new PA mechanisms still await discovery.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

PROCESSIVE ANTITERMINATION

Termination of transcription at any given location is rarely 100% complete, with some proportion of elongation complexes proceeding past the point of termination. In general, two types of mechanisms can control transcription elongation to affect the efficiency of termination: transcription attenuation and processive antitermination (PA). For the former, regulatory mechanisms determine the formation of either Rho-dependent or Rho-independent termination sites (10). Importantly, transcription attenuation-based regulatory mechanisms exert their influence on only a single, defined terminator region. In other words, a regulatory RNA that promotes transcription attenuation by definition evolved in concert with the terminator region that it targets—it does not affect other terminator regions. Riboswitches, which are signal-responsive, *cis*-acting regulatory RNAs, oftentimes affect gene expression via transcription attenuation-based mechanisms (20). As discussed elsewhere in this book, riboswitches are widespread in bacteria and offer localized control of transcription termination sites throughout bacterial genomes. In many instances, these transcription attenuation-based regulatory elements can be considered modular, with a signal-responsive portion followed by a portion responsible for premature transcription termination (21).

In contrast, PA mechanisms do not necessarily target a specific terminator region, but instead manipulate elongating RNAP complexes to avoid termination signals throughout an individual transcript (6). These PA strategies do not take a single form and may reduce transcript termination through a variety of direct and indirect effects. For example, some PA strategies rely on direct interference with factor-mediated termination (22). Alternatively, they can modify recruitment of transcription elongation factors, such as NusA, to affect nascent RNA behavior (23, 24). Additionally, they may alter recruitment of ribosomes in a manner that affects termination within coding regions (25). Furthermore, some PA systems have evolved to utilize multiple strategies simultaneously (23, 26).

Phage Lambda Antitermination

During lytic growth, phage λ transcription temporally progresses from one large set of genes to another (27). In order to switch from intermediate-early gene expression to delayed-early gene expression, the phage utilizes a unique protein, λN, to promote antitermination, which enables expression of downstream genes (Figure 1A) (28). λN is a small protein that is intrinsically disordered alone (29) but is stabilized by protein and RNA contacts in the final, λN antitermination complex (Figure 2A) (23). Formation of the λN antitermination complex is triggered by synthesis of a *nut* sequence, composed of two RNA elements. The first, *boxB*, is a 15-nucleotide motif that resembles a GNRA tetraloop structure (Figure 3) (30, 31) and serves as the substrate for λN binding (23, 24). In addition to binding λN, *boxB* also interacts with NusA. Formation of the antitermination complex occurs in steps, with initial association of λN to *boxB* followed by binding of NusA to the λN:*boxB* complex (32). This minimal λN:*boxB*:NusA complex is sufficient for antitermination of *nut*-proximal terminator sequences (6), although it is generally believed that the full antitermination complex *in vivo* relies on additional elongation proteins loaded at the second RNA element. This second RNA element, *boxA*, acts as a loading site for the NusB:NusE

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(S10) complex (33). Binding of the NusB:NusE (S10) complex to *boxA* promotes additional contacts between λN and NusA. This results in a unique complex of factors that are associated with RNAP near the RNA exit channel and remain together as a ribonucleoprotein complex (Figure 2B) (23).

Binding of λN alone to RNAP modifies transcription elongation both *in vitro* and *in vivo,* promoting antitermination by modulating RNA exit channel elements and by suppressing melting of the RNA:DNA hybrid after terminator hairpin formation or in response to Rho activity (23, 24, 34, 35). However, formation of the complex with the full complement of transcription elongation factors is thought to further stabilize the interaction of λN with RNAP and increase its duration of occupancy—and, therefore, overall processivity—of λN antitermination (36). In "standard" transcription elongation complexes, NusA binds RNA polymerase near the RNA exit channel where it can enhance intrinsic termination (37). Indeed, NusA affects transcription termination at many locations across the genome and is even required for formation of some NusA-dependent termination sites (13). However, λN is thought to counteract the direct effects of NusA on terminator hairpin folding (24). A recent high-resolution structural model of the λN antitermination complex revealed that the C-terminal RNA-binding domains of NusA are repositioned such that they redirect nascent RNA away from the RNA exit channel (Figure 2C). This is predicted to reduce formation of terminator hairpins, thereby essentially reprogramming NusA into a transcription antitermination factor (23). Formation of the λN complex also inhibits Rho-dependent termination. In "standard" elongation complexes, NusG helps recruit Rho to nascent RNA and thereby aids in Rho termination (38, 39). In contrast, the λN antitermination complex is likely to restrict NusG-mediated recruitment of Rho by instead promoting association of factors that compete for binding to NusG (*e.g.,* S10:NusB), and also because of restricted access to the nascent RNA as it is looped out of the antitermination complex (23). Therefore, the λN complex acts as a physical roadblock to prevent Rho translocation and helps occlude access to Rho utilization (*rut*) sites.

Phage λ also contains a second antitermination system, which relies upon another unique protein (λQ) to promote antitermination of late-expressed genes (6, 40). However, unlike the N-antitermination system, λQ protein is a DNA-binding protein that associates with RNA polymerase within the promoter region during transcription initiation and triggers formation of an antitermination complex that is different from the N complex (41).

Ribosomal RNA Operon Antitermination

Dissociation of transcription elongation complexes by Rho helicase underlies the polarity which occurs when nonsense mutations reduce transcript abundance of downstream genes (42). Rho is capable of loading onto RNA molecules via C-rich binding sequences (*rut* sites), but the presence of ribosomes during coupled transcription-translation generally reduces Rho loading and translocation (43). Given that ribosomal RNA operons are not translated and are thereby not protected by ribosomes, their transcripts must be protected from Rho termination by other means. This protection may be partially explained by the extensive secondary structure of ribosomal RNAs, which acts to reduce loading of Rho at potential *rut* sites (44, 45). However, in *Escherichia coli* and many other bacteria, these

Author Manuscript

operons are also subjected to an antitermination system that resembles closely the λN-antitermination mechanism (44–46). For example, the 5' leader regions of *E. coli* rRNA operons contain *boxA* as well as a *boxB*-like hairpin, although only *boxA* appears to be essential for antitermination activity (Figure 1C) (33, 47). Binding of the NusB:NusE (S10) complex to *boxA* RNA occurs in a manner similar to N-mediated antitermination, ultimately promoting a conformational state that strongly disfavors association of Rho (33).

In contrast to λN antitermination, which requires N protein in addition to host Nus proteins, rRNA antitermination requires an additional host factor, SuhB (48). The complete elongation complex containing NusB:NusE, NusA, NusG, and SuhB is required not only for full rRNA antitermination activity *in vitro* but for correct rRNA maturation *in vivo (48)*. In addition to regulation of rRNA transcription, *boxA* and Nus factors directly repress *suhB* translation in enterobacteria in a manner reminiscent of λN autoregulation and have been implicated in regulation of additional genes (49). Therefore, the rRNA antitermination system relies exclusively on general transcription elongation factors and their recruitment to the *boxA* RNA element. This system serves a dual purpose in rRNA operons, promoting both antitermination and RNA folding, and may regulate yet additional transcripts. Together, these observations suggest that N-antitermination may have arisen as a modification of the host Nus protein antitermination system, where λN protein evolved to reconfigure and further manipulate host transcription elongation factors.

RNA Elements that Promote Processive Antitermination

In addition to the role that RNA elements (*boxA* and *boxB*) play in antitermination of phage λ and rRNA operons, a few PA systems have been discovered that involve larger and more complicated RNA elements. Many if not most lambdoid phages utilize PA systems related to both N- or Q-antitermination (6). However, phage HK022 differs in that it encodes for λQ yet lacks λN, despite the fact it still requires antitermination of early-expressed genes (50). Moreover, HK022 does not utilize *nut* sites for antitermination. Instead, early gene antitermination is mediated directly by a larger RNA motif called *put*, found in regions analogous to λ *nut* sites (Figure 1B) (51). HK022 *put* forms a two-hairpin RNA element of approximately 65 nucleotides in length that is critical for antitermination activity (Figure 3) (51, 52). This element appears to directly affect RNAP elongation activity through pause suppression, potentially requiring no additional elements to promote antitermination (50). Evolution of this mechanism is likely interrelated with the evolution of a λN-like protein, Nun, which is also produced by HK022 (53, 54). Nun, found in the same relative genomic position as λN in phage λ, instead promotes Nun-termination at *nut* elements by binding to *boxB* and inhibiting RNAP translocation (55, 56). HK022 *put* promotes antitermination of both Rho-dependent termination and Nun-dependent transcription arrest in the HK022 early transcripts (55) as well as intrinsic terminators (57). While some mechanistic details of *put*-mediated antitermination are still lacking, its discovery was significant as it demonstrated proof-in-principle that PA could be driven primarily by RNA elements.

More recently, an even larger and more structurally complicated RNA element was discovered to trigger PA in bacteria. This RNA element, which is at least ~125 nucleotides in length and is constructed from an array of at least five helical elements and a characteristic

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

pseudoknot, was discovered to be broadly conserved in Bacillales (Figure 3) (58). Coined the EAR element, for *eps*-associated RNA, it is almost always associated with operons that encode for biosynthesis of biofilm or capsule exopolysaccharides (Figure 1D). Either mutagenesis of conserved residues or deletion of EAR resulted in incomplete transcription of the *Bacillus subtilis eps* operon. Instead, transcripts were found to be prematurely truncated at the site of intrinsic terminators, located in the middle region of the *eps* operon. Indeed, placement of EAR directly upstream of this terminator site resulted in nearly complete read-through of the terminators *in vivo*, whereas, conversely, mutagenesis of conserved EAR residues resulted in termination. Moreover, placement of EAR upstream of unrelated intrinsic terminators, originating from sources other than the *eps* operon, still resulted in their read-through, demonstrating that EAR promotes general PA of intrinsic terminators. That EAR promoted read-through of intrinsic terminators is strikingly different than the biological utilization of the λN and rRNA PA systems, which are believed to function primarily for read-through of Rho termination. However, EAR PA has not yet been recapitulated *in vitro* or in a heterologous host, indicating that at least one additional factor may be required for its antitermination activity, in contrast to HK022 *put*. Regardless, discovery of EAR demonstrated that structurally complicated RNAs, with the size and apparent complexity resembling that of riboswitches, are sometimes used to promote PA. Moreover, the distribution of EAR PA determinants further showcases how PA mechanisms can be broadly important for biologically important functions such as biofilm formation.

SPECIALIZED NUSG PARALOGS

RfaH

Although most known PA systems are found in phage genomes or are reliant on general transcription elongation factors, some Gammaproteobacteria encode for the specialized PA and translation factor RfaH (26). RfaH is a paralog of NusG. NusG is an elongation factor generally associated with transcription elongation complexes and is an integral component of the λ and rRNA PA systems (59). RfaH, encoded by an essential gene in *E. coli*, is required for the expression of a regulon of virulence-related pathways—including synthesis of haemolysin, lipopolysaccharide, and the F-factor sex pilus (59, 60)—as well as additional targets involved in the production of membrane or extracellular components (61).

As a paralog of NusG, RfaH is a small protein containing two conserved domains. In general, the core domains of NusG homolog proteins exhibit strongly conserved structure (62, 63) and interface with RNAP in a similar fashion (63–65). The first domain is an N-terminal domain (NTD) unique to the NusG/Spt5 family of proteins (66). This domain is responsible for binding of RfaH to RNAP at the same site normally occupied by NusG. The C-terminal domain (CTD) contains a KOW (Kyprides, Ouzounis, Woese) motif found in several ribosomal proteins in addition to NusG (67). This characteristic CTD is shared among nearly all NusG homologs as well as several ribosomal proteins (67), and is believed to function as a tether that can interact with additional proteins (68).

While RfaH and NusG have distinct regulatory consequences, they rely on similar mechanisms to improve transcriptional processivity (65). The NTD of both proteins share highly similar sequences and structures (61, 63) and suppress pausing at many sites when

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

added to purified transcription complexes *in vitro* (22, 69, 70). Both proteins are believed to suppress pausing by binding to the β' clamp and β pincer and stabilizing the active closed conformation of RNAP (63, 71). Recently, single molecule cryo-EM studies have clarified how stabilization of RNAP structure can promote processive elongation. Certain types of transcriptional pauses are affected by a swivelling of the RNAP β' pincer elements, resulting in an increase in pause lifetimes (72). However, binding of NusG or RfaH to RNAP disfavors this "swiveled" conformation, thereby suppressing pausing (65). Additional mechanisms for anti-pausing activity of NusG proteins have been proposed, including stabilization of the elongation complex by direct binding to non-template DNA (22, 70, 73) as well as upstream DNA (74–77). Indeed, both NusG and RfaH interact with the upstream DNA fork and promote re-annealing of the upstream DNA, although the specific effects of this activity on RNAP activity are unclear (65, 74). These mechanisms are conserved between NusG and RfaH, are are likely to be shared to varying degrees with other NusG paralogs.

RfaH is specifically recruited to the operons that comprise its regulon by a DNA element called the operon polarity suppressor, or *ops* (Figure 4A). Deletion of this 8-bp conserved element reduces downstream gene expression (60); correspondingly, introduction of *ops* to other transcripts increases their expression (59). Depletion of RfaH mirrors these results, indicating that RfaH and *ops* are both required for expression of target operons (78). RfaH is specifically recruited to transcription elongation complexes by binding to the non-template DNA strand of the *ops*-element; this occurs during the lifetime of a programmed transcriptional pause (22). The *ops*-element forms both a consensus pause sequence as well as a DNA hairpin loop that makes specific, direct contacts to the RfaH NTD (79). RfaH and NusG are mutually exclusive, as both homologs share the same binding site on RNAP (80, 81). Moreover, once recruited, RfaH exhibits increased affinity for RNAP relative to NusG, allowing for extended association of RfaH with TECs (82). This increased affinity may also be responsible for the more pronounced effects of RfaH NTD on RNAP as compared to NusG (65). In this way, RfaH exerts its regulatory effects specifically on those operons that include the *ops* element.

RfaH in solution differs from RNAP-associated RfaH. Instead of the common β-barrel fold found in most high-resolution structures of KOW domains, the CTD of free RfaH forms a dramatically different α-helical structure (80). This α-helical CTD interacts with the NTD, partially masking the RNAP-binding portion and thereby resulting in an autoinhibited form of the protein (Figure 5) (82). After a conformational change is triggered, the NTD can associate fully with the transcription complex, which in turn promotes re-folding of the CTD to the β-barrel structure found in NusG (Figure 5) (26). Because of this structural mechanism, RfaH adopts the classical NusG KOW domain structure only after the NTD has fully associated with RNA polymerase.

Though NusG and RfaH display nearly identical anti-pausing effects on transcription complexes *in vitro*, their overall regulatory outcomes are different. In some instances, NusG may promote pausing *in vivo (83)*, perhaps as a result of increased affinity for certain non-template DNA strand sequences (70). More importantly, NusG is known to directly bind Rho (68). This interaction is likely to broadly promote Rho-dependent termination activity,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

possibly by increasing the rate at which Rho successfully binds RNA and forms a closed translocation-capable conformation (39). Ultimately, association of NusG results in Rho-dependent termination and suppression of transcription, particularly in genomic regions that feature foreign DNA (25). This activity is essential in most *E. coli* strains primarily due to suppression of toxic genes in prophage DNA (25). However, in addition to its interaction with Rho, the NusG CTD can associate with NusE (S10), as well as NusA (84, 85). Similar to NusG, RfaH can associate with NusE (S10); however, in contrast to NusG, RfaH is incapable of binding Rho (26, 84). Because of this, RfaH strongly discourages Rho termination within its targeted operons (86).

Finally, the remaining mechanism by which RfaH may promote antitermination is through recruitment of ribosomes to nascent transcripts. NusG proteins are thought to couple transcription and translation by facilitating macromolecular interactions between both of these machines (84). RfaH in particular has been shown to exhibit much stronger polarity effects *in vivo* than its effects on transcription *in vitro (82)*. Also, genes that are known to be regulated by RfaH display particularly poor ribosome binding sites, suggesting that translational enhancement is likely to be a key feature of RfaH regulation (86). It is possible that binding of NusG or RfaH to ribosomal S10 (NusE) may assist ribosome recruitment, thereby increasing local concentration of ribosomes and promoting translation initiation on nascent RNA (26, 87). This functional interaction might also affect transcription processivity. Indeed, recent data suggest that the leading ribosome—which conducts translation immediately upstream of RNAP, and that may participate in the RNAP-ribosome “expressome” (88, 89)—improves transcription processivity by directly blocking RNAP backtracking (90) and by obstructing Rho access (84, 91).

Through these aggregate mechanisms, RfaH acts as a specialized elongation factor that exhibits anti-pausing activity, prevents NusG-mediated Rho termination, and encourages ribosome recruitment, for each of the operons that display *ops* elements.

Other NusG Paralogs (ActX, TaA, UpxY, LoaP)

Although RfaH is the most prominent and best studied NusG paralog, other examples have been identified, several of which have been predicted to function in transcription antitermination (92–95). All of these homologs share significant sequence similarity to NusG and RfaH and undoubtedly share conserved structural features. Moreover, for those NusG paralogs where a functional role has been demonstrated, they have inevitably been found to affect transcription of certain targeted transcripts, suggesting that NusG paralogs are broadly used by bacteria as specialized transcription regulators (93–95).

ActX and TraB proteins are most phylogenetically similar to RfaH (92, 96) and are found in a variety of conjugative plasmids conferring antibiotic resistance in Gammaproteobacteria (92, 97). Though a function has not been demonstrated for these proteins, they are often transcribed as the first open reading frame in long pilus biosynthesis operons and are suspected to be involved in the transcription of conjugation genes (98).

Myxococcus xanthus, a Gram-negative soil bacterium, produces the well-studied polyketide antibiotic TA (also called myxovirescin) (99, 100). The first open reading frame of the TA-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

producing gene cluster is *taA*, which encodes for a NusG paralog (93). Disruption of the *taA* gene eliminated antibiotic production, suggesting a regulatory relationship. However, the specific role of TaA in expression of the TA gene cluster is unknown, although as a NusG paralog and relative of other known NusG specialized paralogs, it has been proposed to regulate transcription elongation, perhaps through PA.

More recently, a NusG paralog called UpxY has been proposed to function as a family of regulators for complex polysaccharide pathways in Bacteroidetes (94, 101). They are widely used by these microorganisms. Indeed, many Bacteroides encode between six and nine copies of the UpxY proteins. The genes encoding these proteins, initially described in *Bacteroides fragilis*, are each associated with a different capsular polysaccharide gene cluster (Figure 4C) (101). These proteins have been shown to affect transcription of their associated gene cluster and it has been proposed that they participate in antitermination-based regulatory mechanisms that involve unique sequence features located within the 5' leader regions of their respective operons. Additionally, while these regulators might be co-transcribed with the operons they affect, they can also affect gene expression when moved to a distal location, supporting the claim that they are recruited to their targeted operons, perhaps via sequence elements within the transcript leader regions. Yet, despite these observations, little is known regarding the molecular mechanisms of UpxY proteins. Adding a new wrinkle to the overall family of NusG paralogous proteins, *Bacteroides fragilis* also encodes a set of unique proteins (UpxZ) alongside genes encoding UpxY proteins. The UpxZ proteins can act as *trans*-inhibitors of UpxY proteins, and have been hypothesized to hierarchically regulate the expression of different sets of capsular polysaccharides, although the underlying mechanism of this inhibition is also unknown (94).

The most recently described NusG paralog is LoaP. Genes encoding LoaP are consistently positioned adjacent to polyketide biosynthesis pathways in Firmicutes, or near polysaccharide biosynthesis gene clusters in certain Firmicutes, Actinobacteria and Spirochaetes (95). While proximity alone is not evidence of a functional relationship, *Bacillus velezensis loaP* was shown to affect expression of an adjacent polyketide synthesis gene cluster. More specifically, *Bacillus velezensis loaP* is situated adjacent to a gene cluster (*dfn*) that encodes for production of the polyketide antibiotic difficidin (102). Deletion of *loaP* resulted in low abundance across the *dfn* transcript, whereas complementation of *loaP* from an ectopic locus restored *dfn* expression. Moreover, these global expression experiments revealed that the difficidin operon is not the only region affected by LoaP. LoaP is also required for transcription elongation of a second polyketide gene cluster, which encodes for production of another antibiotic, macrolactin, indicating that LoaP controls a regulon of antibiotic biosynthesis operons in *B. velezensis* (Figure 4B).

Upon depletion of LoaP, transcript abundance dramatically decreases at intrinsic terminator sites located within the targeted operons, whereas induction of LoaP restored read-through of these terminators (95). These initial data suggest that LoaP antitermination may function primarily on intrinsic termination, in contrast to the suppression of Rho-dependent termination known for RfaH (22, 95). However, while this might be due to the preference of Gram-positive bacteria for Rho-independent antitermination, it is also true that Rho termination has been insufficiently characterized in *B. velezensis* and other Firmicutes (103).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Therefore, the relationship between LoaP proteins and Rho termination is still unknown. Nor have the full determinants for LoaP PA yet been described. While LoaP affects transcript abundance across the length of the targeted operon, it appears to require sequence elements located somewhere within the 5' leader region. Indeed, when placed upstream of terminator signals and upstream of a reporter gene, the *dfn* leader region alone is sufficient for promoting LoaP-dependent PA activity (95). Therefore, the recruitment signals for LoaP are located fully within this leader region. Interestingly, a small UNCG-type hairpin was identified in the leader regions of both the difficidin and macrolactin operons in a sequence region required for antitermination, although its exact relationship to LoaP regulation has not yet been investigated (Figure 3) (95).

The discovery of LoaP, along with the initial description of TaA, suggests that transcription elongation may be a broad point of regulatory control for secondary metabolite gene clusters in bacteria. Therefore, it is important to study PA mechanisms in order to improve discovery and production of new natural products from bacteria.

Phylogenetic Overview of the NusG Family of Proteins

NusG paralogs putatively involved in antitermination have been identified in a variety of bacteria, including but not limited to Alpha-, Gamma- and Delta-proteobacteria, Bacteroidetes, and most recently Firmicutes, Actinobacteria, and Spirochaetes (86, 95). Of the general transcription elongation factors, only NusG is found in all three domains of life, suggesting its function is important in all organisms. Therefore, essentially all bacteria encode for a core NusG protein, while archaea and eukaryotes encode for a similar protein, Spt5 (80). As a result, all NusG family proteins share core conserved sequence and structure features (80).

Although analysis of the paralogs supports grouping them within the overall NusG family, each sub-grouping displays significant sequence diversity, with some subgroups displaying very limited overall sequence identity despite sharing remarkably conserved structural elements (95). In recent work, the phylogenetic analyses of the NusG family was extended to include as many distant homologs as are detectable by HMM-based homology modeling (95). This large-scale phylogenetic analysis utilized structural modeling to efficiently align specialized NusG paralogs with limited sequence similarity, and focused on comprehensively covering the diversity of paralog sequences without restriction to the known subgroups. The resulting phylogenetic tree (Figure 6) confirmed that each set of NusG paralogs forms its own distinct group, separate from core bacterial NusG and archaeal Spt5, while also revealing a few new candidate subgroups (82, 95). It is likely that each subgroup will be defined by specific sequence differences. Indeed, a number of characteristic differences between sequences—such as between RfaH or UpxY and core NusG—have been identified as being important for the distinct activities of those specialized paralogs (82, 101, 104).

As NusG paralogs were found in a variety of distinct genetic contexts (82), it was important to systematically identify associations between these genes and potential target pathways. Overall, they were found in diverse genomic contexts, with some positioned alone, at the beginning of complex polysaccharide or secondary metabolite gene clusters, at the end of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

operons, or in unique contexts (82, 95). For example, NusG paralogous sequences from Betaproteobacteria and Bacteroides are located in or near large polysaccharide pathways. TaA and LoaP sequences are generally present in or near large polyketide biosynthesis pathways, which suggests they share a broad relationship to secondary metabolites (82, 95). Indeed, there appears to be a general association of NusG specialized paralogs with polysaccharide biosynthesis gene clusters, and to a lesser extent polyketide synthase gene clusters. In fact, of all the paralog groups, only the Gammaproteobacterial RfaH and its related ActX gene sequences were not frequently identified near or in these classes of gene clusters (95).

There also appear to be several subgroups of NusG paralogs with interesting genomic association and evolutionary distribution, but that have not been characterized or named. For example, a group of sequences closely related to RfaH and found in Alpha-, Beta-, and Gamma-proteobacteria is oftentimes associated with polysaccharide gene clusters. Similarly, at least two more uncharacterized and unnamed putative groups of sequences are consistently associated with polysaccharide and polyketide biosynthesis gene clusters. From this, it can be tentatively speculated that NusG specialized paralogs evolved as regulators of these long operons (polysaccharides and secondary metabolite biosynthesis genes) and became further specialized into RfaH in Gammaproteobacteria. Finally, an additional set of paralog sequences in Alphaproteobacteria was not found in a consistent genomic context, and remains unnamed. Ultimately, the evolutionary relationship between all these different NusG paralogs remains unclear, as bootstrap support for early branches after divergence from core NusG is low, likely due to the extensive sequence divergence in this family. Elucidating the true history of this family may require different approaches, integrating more information about the structural changes and sequence insertions and deletions during evolution of the NusG paralogs. However, it is already clear that the NusG family of proteins is widely used in bacteria as specialized transcription elongation regulatory factors, and that they regulate expression of fundamentally important pathways, albeit through largely unexplored molecular mechanisms.

OUTLOOK

The past few years have uncovered a few new examples of PA mechanisms, as well as remarkable new insight into the structural basis of antitermination activity. However, it is possible that these findings still only represent a small proportion of what remains to be discovered. Therefore, what is now needed is a systematic exploration of the molecular mechanisms used by NusG paralogs, combined with new bioinformatics searches for RNA elements that promote PA. From this, an accurate portrayal of the extent of PA usage in bacteria will emerge, which will resolve whether transcription elongation is a much broader point of regulatory control than has historically been perceived. Furthermore, studying new PA systems will help uncover the diversity of their molecular mechanisms and shed important light on when and why PA mechanisms are employed by bacteria.

Author Manuscript

References

1. Jacob F, Monod J. 1961 Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3:318–356. [PubMed: 13718526]
2. Walsh C 2003 Regulation of Antibiotic Biosynthesis in Producer Organisms, p. 159–174. *In* Antibiotics: Actions, Origins, Resistance. ASM Press.
3. Waters LS, Storz G. 2009 Regulatory RNAs in bacteria. Cell 136:615–628. [PubMed: 19239884]
4. Barrick JE, Breaker RR. 2007 The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol 8:R239. [PubMed: 17997835]
5. Breaker RR. 2011 Prospects for riboswitch discovery and analysis. Mol Cell 43:867–879. [PubMed: 21925376]
6. Weisberg RA, Gottesman ME. 1999–1 Processive Antitermination. J Bacteriol 181:359–367. [PubMed: 9882646]
7. Roberts JW, Shankar S, Filter JJ. 2008 RNA polymerase elongation factors. Annu Rev Microbiol 62:211–233. [PubMed: 18729732]
8. Santangelo TJ, Artsimovitch I. 2011–5 Termination and antitermination: RNA polymerase runs a stop sign. Nat Rev Microbiol 9:319–329. [PubMed: 21478900]
9. Ray-Soni A, Bellecourt MJ, Landick R. 2016 Mechanisms of Bacterial Transcription Termination: All Good Things Must End. Annu Rev Biochem 85:319–347. [PubMed: 27023849]
10. Merino E, Yanofsky C. 2005 Transcription attenuation: a highly conserved regulatory strategy used by bacteria. Trends Genet 21:260–264. [PubMed: 15851059]
11. Gusarov I, Nudler E. 1999 The mechanism of intrinsic transcription termination. Mol Cell 3:495–504. [PubMed: 10230402]
12. Greenblatt J, McLimont M, Hanly S. 1981 Termination of transcription by nusA gene protein of Escherichia coli. Nature 292:215–220. [PubMed: 6265785]
13. Mondal S, Yakhnin AV, Sebastian A, Albert I, Babitzke P. 2016 NusA-dependent transcription termination prevents misregulation of global gene expression. Nature Microbiology 1:15007.
14. Epshtein V, Cardinale CJ, Ruckenstein AE, Borukhov S, Nudler E. 2007 An allosteric path to transcription termination. Mol Cell 28:991–1001. [PubMed: 18158897]
15. Holmes WM, Platt T, Rosenberg M. 1983 Termination of transcription in E. coli. Cell 32:1029–1032. [PubMed: 6188534]
16. Epshtein V, Dutta D, Wade J, Nudler E. 2010 An allosteric mechanism of Rho-dependent transcription termination. Nature 463:245–249. [PubMed: 20075920]
17. Proshkin S, Mironov A, Nudler E. 2014 Riboswitches in regulation of Rho-dependent transcription termination. Biochim Biophys Acta 1839:974–977. [PubMed: 24731855]
18. DebRoy S, Gebbie M, Ramesh A, Goodson JR, Cruz MR, van Hoof A, Winkler WC, Garsin D a. 2014 A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. Science 345:937–940. [PubMed: 25146291]
19. Sedlyarova N, Shamovsky I, Bharati BK, Epshtein V, Chen J, Gottesman S, Schroeder R, Nudler E. 2016 sRNA-Mediated Control of Transcription Termination in E. coli. Cell 167:111–121.e13. [PubMed: 27662085]
20. Montange RK, Batey RT. 2008 Riboswitches: emerging themes in RNA structure and function. Annu Rev Biophys 37:117–33. [PubMed: 18573075]
21. Ceres P, Garst AD, Marcano-Velázquez JG, Batey RT. 2013 Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. ACS Synth Biol 2:463–472. [PubMed: 23654267]
22. Artsimovitch I, Landick R. 2002 The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. Cell 109:193–203. [PubMed: 12007406]
23. Said N, Krupp F, Anedchenko E, Santos KF, Dybkov O, Huang Y-H, Lee C-T, Loll B, Behrmann E, Bürger J, Mielke T, Loerke J, Urlaub H, Spahn CMT, Weber G, Wahl MC. 2017 Structural basis for λN-dependent processive transcription antitermination. Nat Microbiol 2:17062. [PubMed: 28452979]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

24. Gusarov I, Nudler E. 2001 Control of intrinsic transcription termination by N and NusA: the basic mechanisms. Cell 107:437–449. [PubMed: 11719185]

25. Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. 2008 Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. Science 320:935–938. [PubMed: 18487194]

26. Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney R a., Landick R, Artsimovitch I, Rösch P. 2012 An α helix to β barrel domain switch transforms the transcription factor RfaH into a translation factor. Cell 150:291–303. [PubMed: 22817892]

27. Friedman DI. 1988 Regulation of phage gene expression by termination and antitermination of transcription. The bacteriophages 2:263–319.

28. Patterson TA, Zhang Z, Baker T, Johnson LL, Friedman DI, Court DL. 1994 Bacteriophage lambda N-dependent transcription antitermination. Competition for an RNA site may regulate antitermination. J Mol Biol 236:217–228. [PubMed: 8107107]

29. Mogridge J, Legault P, Li J, Van Oene MD, Kay LE, Greenblatt J. 1998 Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. Mol Cell 1:265–275. [PubMed: 9659923]

30. Legault P, Li J, Mogridge J, Kay LE, Greenblatt J. 1998 NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. Cell 93:289–299. [PubMed: 9568720]

31. Thapar R, Denmon AP, Nikonowicz EP. 2014 Recognition modes of RNA tetraloops and tetraloop-like motifs by RNA-binding proteins. Wiley Interdiscip Rev RNA 5:49–67. [PubMed: 24124096]

32. Mogridge J, Mah T-F, 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. [PubMed: 9461609]

33. Nodwell JR, Greenblatt J. 1993 Recognition of boxA antiterminator RNA by the E. coli antitermination factors NusB and ribosomal protein S10. Cell 72:261–268. [PubMed: 7678781]

34. Mason SW, Li J, 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. [PubMed: 1388170]

35. Rees WA, Weitzel SE, Das A, von Hippel PH. 1997 Regulation of the elongation-termination decision at intrinsic terminators by antitermination protein N of phage lambda. J Mol Biol 273:797–813. [PubMed: 9367773]

36. Nudler E, Gottesman ME. 2002 Transcription termination and anti-termination in E. coli. Genes Cells 7:755–768. [PubMed: 12167155]

37. Liu K, Zhang Y, Severinov K, Das A, Hanna MM. 1996 Role of Escherichia coli RNA polymerase alpha subunit in modulation of pausing, termination and anti-termination by the transcription elongation factor NusA. EMBO J 15:150–161. [PubMed: 8598198]

38. Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. 2012 Rho and NusG suppress pervasive antisense transcription in Escherichia coli. Genes Dev 26:2621–2633. [PubMed: 23207917]

39. Valabhoju V, Agrawal S, Sen R. 2016 Molecular Basis of NusG-mediated Regulation of Rho-dependent Transcription Termination in Bacteria. J Biol Chem 291:22386–22403. [PubMed: 27605667]

40. Herskowitz I, Signer ER. 1970 A site essential for expression of all late genes in bacteriophage lambda. J Mol Biol 47:545–556. [PubMed: 5418171]

41. Yarnell WS, Roberts JW. 1992 The phage lambda gene Q transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. Cell 69:1181–1189. [PubMed: 1535556]

42. Lowery C, Richardson JP. 1977 Characterization of the nucleoside triphosphate phosphohydrolase (ATPase) activity of RNA synthesis termination factor p. II. Influence of synthetic RNA …. J Biol Chem.

43. Guérin M, Robichon N, Geiselmann J, Rahmouni AR. 1998 A simple polypyrimidine repeat acts as an artificial Rho-dependent terminator in vivo and in vitro. Nucleic Acids Res 26:4895–4900. [PubMed: 9776750]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

44. Li SC, Squires CL, 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. [PubMed: 6091902]
45. Squires CL, Greenblatt J, Li J, Condon C, Squires CL. 1993 Ribosomal RNA antitermination in vitro: requirement for Nus factors and one or more unidentified cellular components. Proc Natl Acad Sci U S A 90:970–974. [PubMed: 8430111]
46. Arnvig KB, Zeng S, Quan S, Papageorge A, Zhang N, Villapakkam AC, Squires CL. 2008 Evolutionary comparison of ribosomal operon antitermination function. J Bacteriol 190:7251–7257. [PubMed: 18757535]
47. Berg KL, Squires C, Squires CL. 1989 Ribosomal RNA operon anti-termination. Function of leader and spacer region box B-box A sequences and their conservation in diverse micro-organisms. J Mol Biol 209:345–358. [PubMed: 2479752]
48. Singh N, Bubunenko M, Smith C, Abbott DM, Stringer AM, Shi R, Court DL, Wade JT. 2016 SuhB Associates with Nus Factors To Facilitate 30S Ribosome Biogenesis in Escherichia coli. MBio 7:e00114. [PubMed: 26980831]
49. Baniulyte G, Singh N, Benoit C, Johnson R, Ferguson R, Paramo M, Stringer AM, Scott A, Lapierre P, Wade JT. 2017 Identification of regulatory targets for the bacterial Nus factor complex. Nat Commun 8:2027. [PubMed: 29229908]
50. Clerget M, Jin DJ, Weisberg RA. 1995 A zinc-binding region in the beta' subunit of RNA polymerase is involved in antitermination of early transcription of phage HK022. J Mol Biol 248:768–780. [PubMed: 7752239]
51. King RA, Banik-Maiti S, Jin DJ, Weisberg RA. 1996 Transcripts that increase the processivity and elongation rate of RNA polymerase. Cell 87:893–903. [PubMed: 8945516]
52. Banik-Maiti S, King RA, Weisberg RA. 1997 The antiterminator RNA of phage HK022. J Mol Biol 272:677–687. [PubMed: 9368650]
53. Robert J, Sloan SB, Weisberg RA, Gottesman ME, Robledo R, 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. [PubMed: 2822258]
54. Hung SC, Gottesman ME. 1997 The Nun protein of bacteriophage HK022 inhibits translocation of Escherichia coli RNA polymerase without abolishing its catalytic activities. Genes Dev 11:2670–2678. [PubMed: 9334329]
55. King RA, Weisberg RA. 2003 Suppression of factor-dependent transcription termination by antiterminator RNA. J Bacteriol 185:7085–7091. [PubMed: 14645267]
56. Vitiello CL, Kireeva ML, Lubkowska L, Kashlev M, Gottesman M. 2014 Coliphage HK022 Nun protein inhibits RNA polymerase translocation. Proc Natl Acad Sci U S A 111:E2368–75. [PubMed: 24853501]
57. Oberto J, Clerget M, Ditto M, Cam K, Weisberg RA. 1993 Antitermination of early transcription in phage HK022. Absence of a phage-encoded antitermination factor. J Mol Biol 229:368–381. [PubMed: 8429552]
58. Irnov I, Winkler WC. 2010 A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. Mol Microbiol 76:559–575. [PubMed: 20374491]
59. Bailey MJ, Hughes C, Koronakis V. 1996 Increased distal gene transcription by the elongation factor RfaH, a specialized homologue of NusG. Mol Microbiol 22:729–737. [PubMed: 8951819]
60. Bailey MJ, Koronakis V, Schmoll T, Hughes C. 1992 Escherichia coli HlyT protein, a transcriptional activator of haemolysin synthesis and secretion, is encoded by the rfaH (sfrB) locus required for expression of sex factor and lipopolysaccharide genes. Mol Microbiol 6:1003–1012. [PubMed: 1584020]
61. Bailey MJ, Hughes C, Koronakis V. 1997 RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. Mol Microbiol 26:845–851. [PubMed: 9426123]
62. Reay P, Yamasaki K, Terada T, Kuramitsu S, Shirouzu M, Yokoyama S. 2004 Structural and sequence comparisons arising from the solution structure of the transcription elongation factor NusG from Thermus thermophilus. Proteins 56:40–51. [PubMed: 15162485]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

63. Martinez-Rucobo FW, Sainsbury S, Cheung ACM, Cramer P. 2011 Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. EMBO J 30:1302–1310. [PubMed: 21386817]
64. Liu B, Steitz TA. 2017 Structural insights into NusG regulating transcription elongation. Nucleic Acids Res 45:968–974. [PubMed: 27899640]
65. Kang JY, Mooney RA, Nedialkov Y, Saba J, Mishanina TV, Artsimovitch I, Landick R, Darst SA. 2018 Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. Cell 173:1650–1662.e14. [PubMed: 29887376]
66. Ponting CP. 2002 Novel domains and orthologues of eukaryotic transcription elongation factors. Nucleic Acids Res 30:3643–3652. [PubMed: 12202748]
67. Kyrpides NC, Woese CR, Ouzounis CA. 1996 KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. Trends Biochem Sci 21:425–426. [PubMed: 8987397]
68. Mooney RA, Schweimer K, Roesch P, Gottesman M, Landick R. 2009 Two structurally independent domains of E. coli NusG create regulatory plasticity via distinct interactions with RNA polymerase and regulators. J Mol Biol 391:341–358. [PubMed: 19500594]
69. Burova E, Hung SC, Sagitov V, Stitt BL, Gottesman ME. 1995 Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. J Bacteriol 177:1388–1392. [PubMed: 7868616]
70. Yakhnin AV, Murakami KS, Babitzke P. 2016 NusG is a Sequence-specific RNA Polymerase Pause Factor that Binds to the Non-template DNA Within the Paused Transcription Bubble. J Biol Chem.
71. Weixlbaumer A, Leon K, Landick R, Darst SA. 2013 Structural basis of transcriptional pausing in bacteria. Cell 152:431–441. [PubMed: 23374340]
72. Kang JY, Mishanina TV, Bellecourt MJ, Mooney RA, Darst SA, Landick R. 2018 RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. Mol Cell 69:802–815.e1. [PubMed: 29499135]
73. Crickard JB, Fu J, Reese JC. 2016 Biochemical Analysis of Yeast Suppressor of Ty 4/5 (Spt4/5) Reveals the Importance of Nucleic Acid Interactions in the Prevention of RNA Polymerase II Arrest. J Biol Chem 291:9853–9870. [PubMed: 26945063]
74. Nedialkov Y, Svetlov D, Belogurov GA, Artsimovitch I. 2018 Locking the non-template DNA to control transcription. Mol Microbiol.
75. Guo G, Gao Y, Zhu Z, Zhao D, Liu Z, Zhou H, Niu L, Teng M. 2015 Structural and biochemical insights into the DNA-binding mode of MjSpt4p:Spt5 complex at the exit tunnel of RNAPII. J Struct Biol 192:418–425. [PubMed: 26433031]
76. Ehara H, Yokoyama T, Shigematsu H, Yokoyama S, Shirouzu M, Sekine S-I. 2017 Structure of the complete elongation complex of RNA polymerase II with basal factors. Science 357:921–924. [PubMed: 28775211]
77. Turtola M, Belogurov GA. 2016 NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. Elife 5.
78. Beutin L, Manning PA, Achtman M, Willetts N. 1981 sfrA and sfrB products of Escherichia coli K-12 are transcriptional control factors. J Bacteriol 145:840–844. [PubMed: 7007352]
79. Zuber PK, Artsimovitch I, NandyMazumdar M, Liu Z, Nedialkov Y, Schweimer K, Rösch P, Knauer SH. 2018 The universally-conserved transcription factor RfaH is recruited to a hairpin structure of the non-template DNA strand. Elife 7.
80. Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin NV, Vassylyev DG, Artsimovitch I. 2007 Structural basis for converting a general transcription factor into an operon-specific virulence regulator. Mol Cell 26:117–129. [PubMed: 17434131]
81. Yakhnin AV, Yakhnin H, Babitzke P. 2008 Function of the Bacillus subtilis transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the trp leader. Proc Natl Acad Sci U S A 105:16131–16136. [PubMed: 18852477]
82. Belogurov G a., Mooney R a., Svetlov V, Landick R, Artsimovitch I. 2009 Functional specialization of transcription elongation factors. EMBO J 28:112–22. [PubMed: 19096362]
83. Yakhnin AV, Babitzke P. 2014 NusG/Spt5: are there common functions of this ubiquitous transcription elongation factor? Curr Opin Microbiol 18:68–71. [PubMed: 24632072]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

84. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rösch P. 2010 A NusE:NusG complex links transcription and translation. Science 328:501–504. [PubMed: 20413501]
85. Strauß M, Vitiello C, Schweimer K, Gottesman M, Rösch P, Knauer SH. 2016 Transcription is regulated by NusA:NusG interaction. Nucleic Acids Res 44:5971–5982. [PubMed: 27174929]
86. Tomar SK, Artsimovitch I. 2013 NusG-Spt5 proteins-Universal tools for transcription modification and communication. Chem Rev 113:8604–8619. [PubMed: 23638618]
87. Saxena S, Myka KK, Washburn R, Costantino N, Court DL, Gottesman ME. 2018 Escherichia coli transcription factor NusG binds to 70S ribosomes. Mol Microbiol 108:495–504. [PubMed: 29575154]
88. Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P. 2017 Architecture of a transcribing-translating expressome. Science 356:194–197. [PubMed: 28408604]
89. Demo G, Rasouly A, Vasilyev N, Svetlov V, Loveland AB, Diaz-Avalos R, Grigorieff N, Nudler E, Korostelev AA. 2017 Structure of RNA polymerase bound to ribosomal 30S subunit. Elife 6.
90. Proshkin S, Rahmouni AR, Mironov A, Nudler E. 2010 Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science 328:504–508. [PubMed: 20413502]
91. Banerjee S, Chalissery J, Bandey I, Sen R. 2006 Rho-dependent transcription termination: more questions than answers. J Microbiol 44:11–22. [PubMed: 16554712]
92. Núñez B, Avila P, de la Cruz F. 1997 Genes involved in conjugative DNA processing of plasmid R6K. Mol Microbiol 24:1157–1168. [PubMed: 9218765]
93. Paitan Y, Orr E, Ron EZ, 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 170:221–7. [PubMed: 9919671]
94. Chatzidaki-Livanis M, Weinacht KG, Comstock LE. 2010 Trans locus inhibitors limit concomitant polysaccharide synthesis in the human gut symbiont Bacteroides fragilis. Proc Natl Acad Sci U S A 107:11976–11980. [PubMed: 20547868]
95. Goodson JR, Klupt S, Zhang C, Straight P, Winkler WC. 2017 LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in *Bacillus amyloliquefaciens*. Nat Microbiol 2:17003. [PubMed: 28191883]
96. Arutyunov D, Arenson B, Manchak J, Frost LS. 2010 F plasmid TraF and TraH are components of an outer membrane complex involved in conjugation. J Bacteriol 192:1730–1734. [PubMed: 20081027]
97. Jones CS, Osborne DJ, Stanley J. 1993 Molecular comparison of the IncX plasmids allows division into IncX1 and IncX2 subgroups. J Gen Microbiol 139:735–741. [PubMed: 8390556]
98. NandyMazumdar M, Artsimovitch I. 2015 Ubiquitous transcription factors display structural plasticity and diverse functions: NusG proteins - Shifting shapes and paradigms. Bioessays 37:324–334. [PubMed: 25640595]
99. Varon M, Fuchs N, Monosov M, Tolchinsky S, Rosenberg E. 1992 Mutation and mapping of genes involved in production of the antibiotic TA in Myxococcus xanthus. Antimicrob Agents Chemother 36:2316–21. [PubMed: 1332595]
100. Simunovic V, Zapp J, Rachid S, Krug D, Meiser P, Müller R. 2006 Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases. Chembiochem 7:1206–20. [PubMed: 16835859]
101. Chatzidaki-Livanis M, Coyne MJ, Comstock LE. 2009 A family of transcriptional antitermination factors necessary for synthesis of the capsular polysaccharides of Bacteroides fragilis. J Bacteriol 191:7288–7295. [PubMed: 19801412]
102. Chen X-H, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Grammel N, Strittmatter AW, Gottschalk G, Süssmuth RD, Borriss R. 2006 Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. J Bacteriol 188:4024–36. [PubMed: 16707694]
103. Mitra P, Ghosh G, Hafeezunnisa M, Sen R. 2017 Rho Protein: Roles and Mechanisms. Annu Rev Microbiol 71:687–709. [PubMed: 28731845]

104. Shi D, Svetlov D, Abagyan R, Artsimovitch I. 2017 Flipping states: a few key residues decide the winning conformation of the only universally conserved transcription factor. Nucleic Acids Res 45:8835–8843. [PubMed: 28605514]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

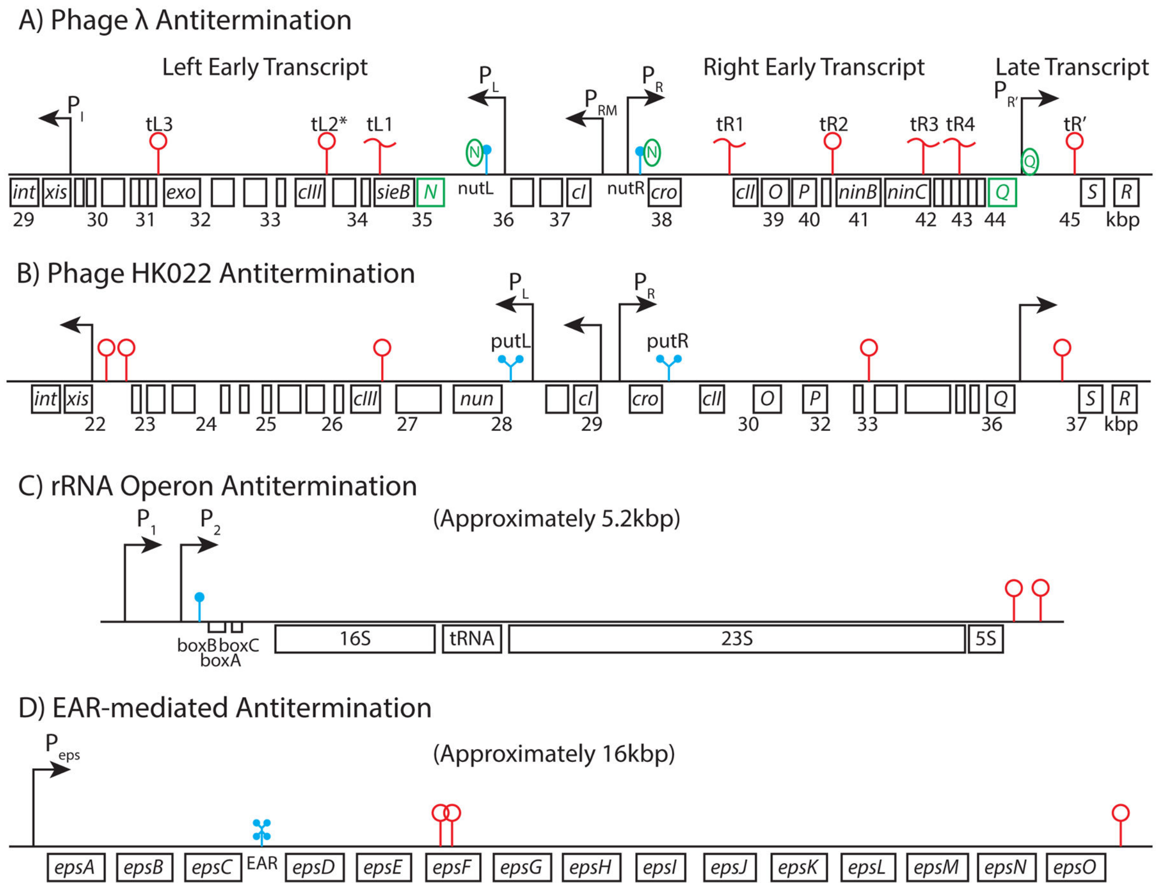

Figure 1: Genomic Context of Processive Antitermination Systems.

This figure schematically illustrates the transcripts regulated by the λN, *put*, rRNA, and EAR RNA-based antitermination systems. (A) Phage λ early transcripts are initiated from two divergently facing promoters with *boxA/B nut* elements found early in the transcripts. The λN protein is encoded by the first gene in the left early transcript. RNA polymerase (RNAP) complexes associated with λN bypass multiple terminators in both transcripts. Using a different mechanism, the λQ protein promotes antitermination of the late transcript by binding to DNA near the late promoter and promoting a terminator-resistant configuration of RNAP. (B) Phage HK022 early transcripts are similar to phage λ, although they include *put* elements early in each transcript, which trigger λN-independent antitermination. Additional Rho-dependent terminators are likely present in these transcripts, although they have not been specifically characterized and are therefore not indicated on this particular diagram. (C) A representative *E. coli* rRNA operon is shown, containing *boxA/B/C* elements immediately downstream of the P_2 promoter. These elements promote read-through of Rho-dependent termination in the non-coding rRNA genes. (D) Several intrinsic terminators have been demonstrated in the *Bacillus subtilis eps* operon, which encodes for biosynthesis of biofilm exopolysaccharides. The *eps*-associated RNA (EAR) is found within the *epsBC* intergenic region and promotes read-through of the terminators within the operon. Intrinsic terminators are shown as sticks with empty circles, and Rho termination regions are shown as sticks with wavy lines, both in red. RNA elements involved

in antitermination are show in blue, and proteins and protein-coding genes involved in antitermination are shown in green. Elements are not shown to scale.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

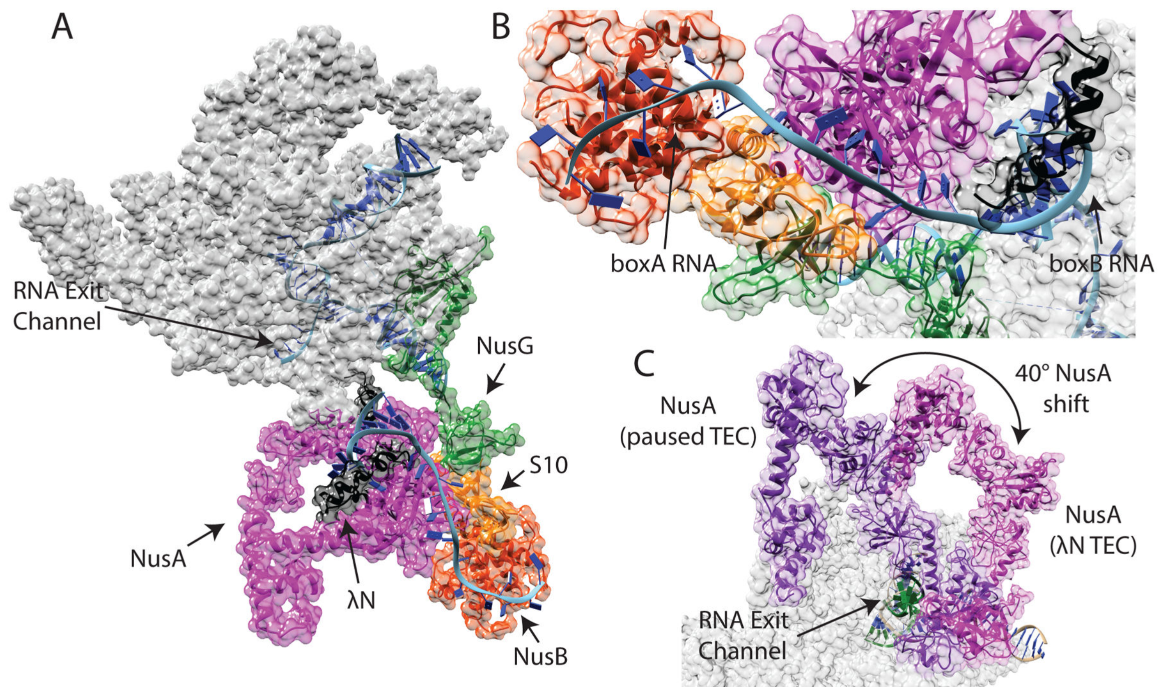

Figure 2: Cryo-electron Microscopy Reveals Details of Antitermination Mechanism.
This figure contains structural models generated from cryo-EM data of transcription elongation complexes (PDB 5MS0, PDB 6FLQ). (A) This panel shows the λN antitermination complex (PDB 5MS0) comprising λN (black), NusA (magenta), NusB (red), S10 (orange), NusG (green), and *boxA/B* RNA (blue), in addition to RNA polymerase (gray). (B) A zoom-in on the *boxA/B* and λN complex shows an extended binding of the *nut* RNA sequence with multiple protein components, with *boxA* bound to the NusB/S10 dimer and the *boxB* hairpin bound to λN and NusA. (C) Formation of the λN antitermination complex shifts the position of NusA (magenta) by 40 degrees away from the RNA exit channel, as compared to NusA (purple) in a transcription elongation complex constructed with the *E. coli his* hairpin-mediated pause sequence (PDB 6FLQ). Nascent RNA is shown in green.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

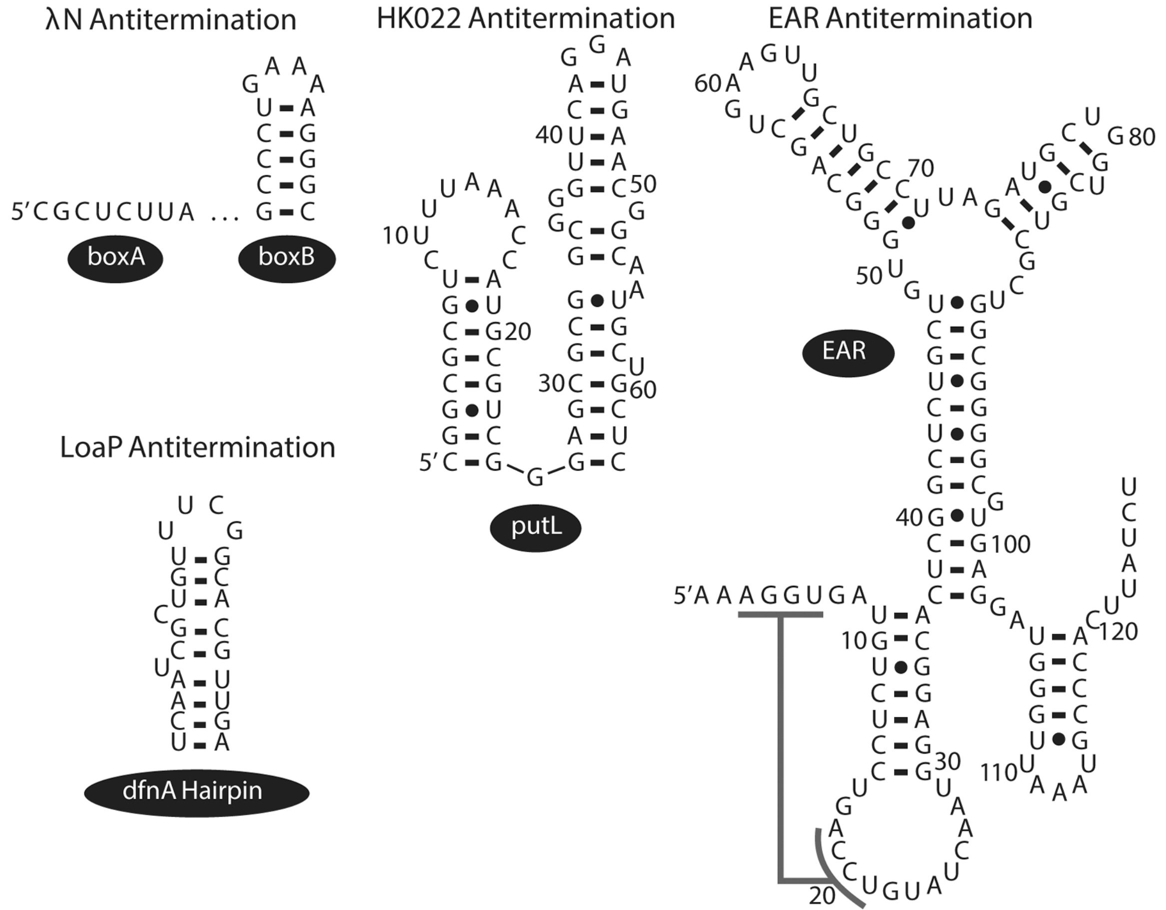

Figure 3: RNA Elements Involved in Processive Antitermination.

This figure shows the sequence and secondary structure of RNA elements known or predicted to be utilized in processive antitermination mechanisms. Shown are the *boxA* and *boxB* elements forming the λN *nut* sequence as well as rRNA antitermination signal, the *put* RNA element from phage HK022, EAR from the *B. subtilis* exopolysaccharide pathway, and a UNCG-type hairpin implicated in LoaP antitermination.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

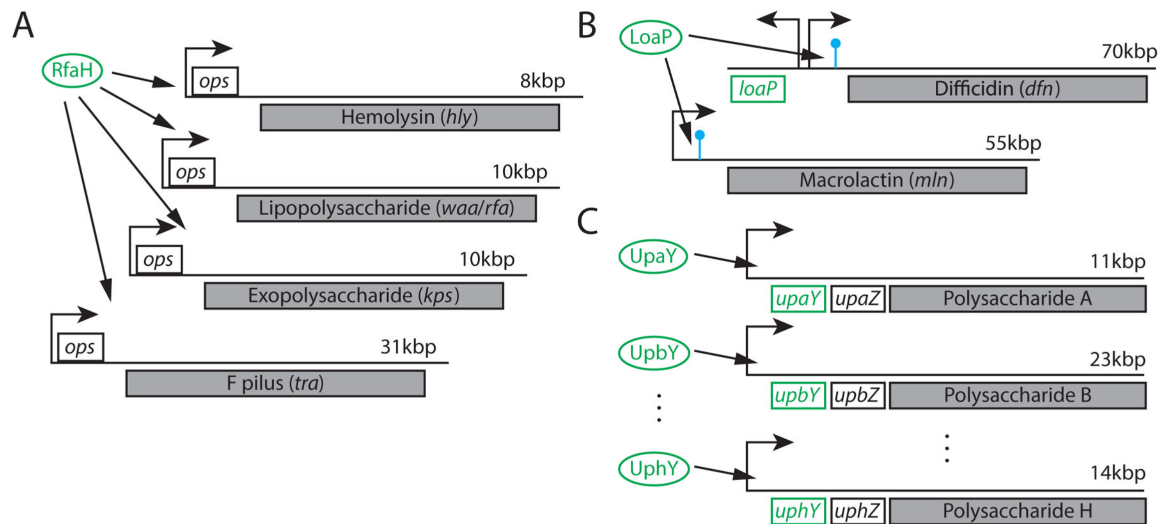

Figure 4: Genomic Context of NusG Paralog Antitermination Systems.
This figure illustrates the transcripts regulated by the RfaH, LoaP, and UpxY antitermination systems. (A) RfaH regulates multiple pathways in *E. coli* including the hemolysin, F pilus, and lipo- and exopolysaccharide operons. Each regulated transcript includes the DNA *ops* element for RfaH recruitment. RfaH promotes antitermination of Rho-dependent promoters. (B) LoaP regulates two polyketide antibiotic operons in *B. velezensis*: the *dfn* difficidin operon and the *mln* macrolactin operon. LoaP is found divergently oriented upstream of the *dfn* operon. Each transcript includes a required sequence region in the 5' leader region, which might include a functionally important hairpin followed by an intrinsic terminator. Additional intrinsic terminator sites have been implicated within the *dfn* and *mln* operons, although they are not shown in this figure. (C) UpxY proteins regulate multiple capsular polysaccharide pathways in *B. fragilis*. Each polysaccharide operon includes both a UpxY and UpxZ protein involved in targeted regulation, with 5' leader sequence required for antitermination. *B. fragilis* has eight distinct polysaccharide operons containing UpxY proteins. Grey rectangles represent multi-gene operons. RNA elements potentially involved in antitermination are show in blue, and proteins and protein-coding genes involved in antitermination are shown in green.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

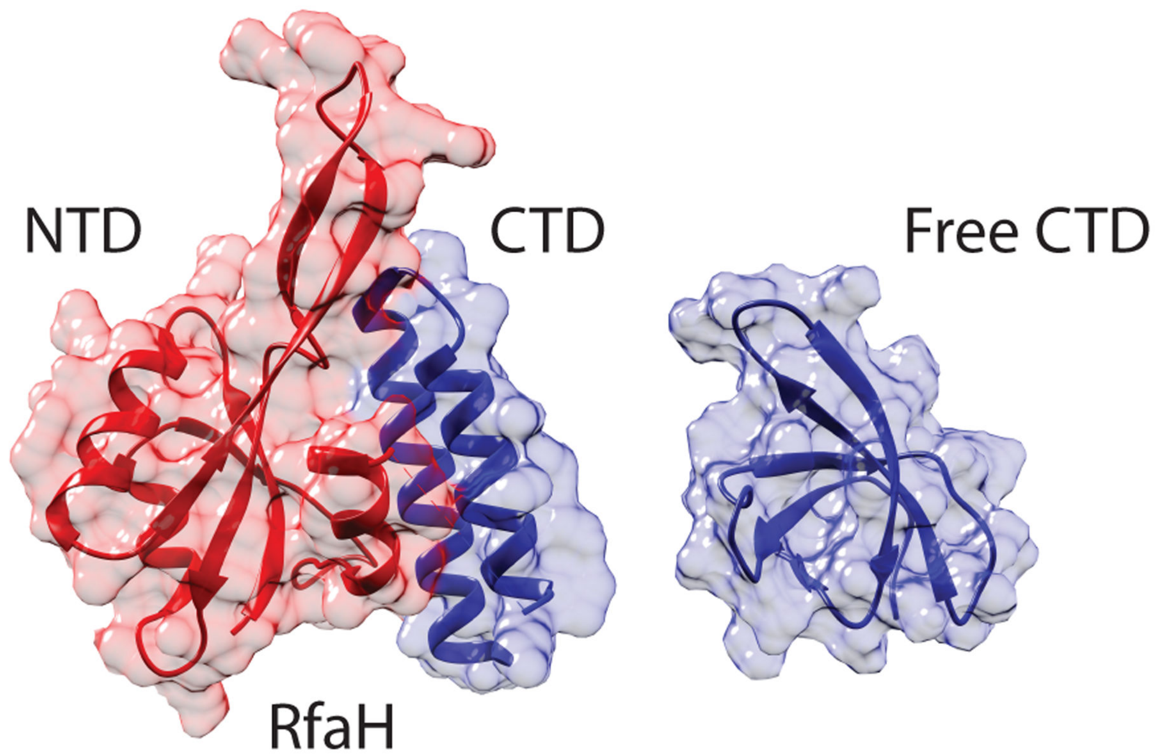

Figure 5: RfaH-CTD Undergoes a Large Conformational Shift from an α helix to β barrel.
Full length RfaH (left) exists as an autoinhibited structure with the CTD (blue) in an α helix conformation bound to the NTD (red) (PDB: 2OUG). Upon binding to RNAP and the *ops* DNA, the CTD (right) is released and forms the β barrel conformation characteristic of NusG KOW domains (PDB: 2LCL).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

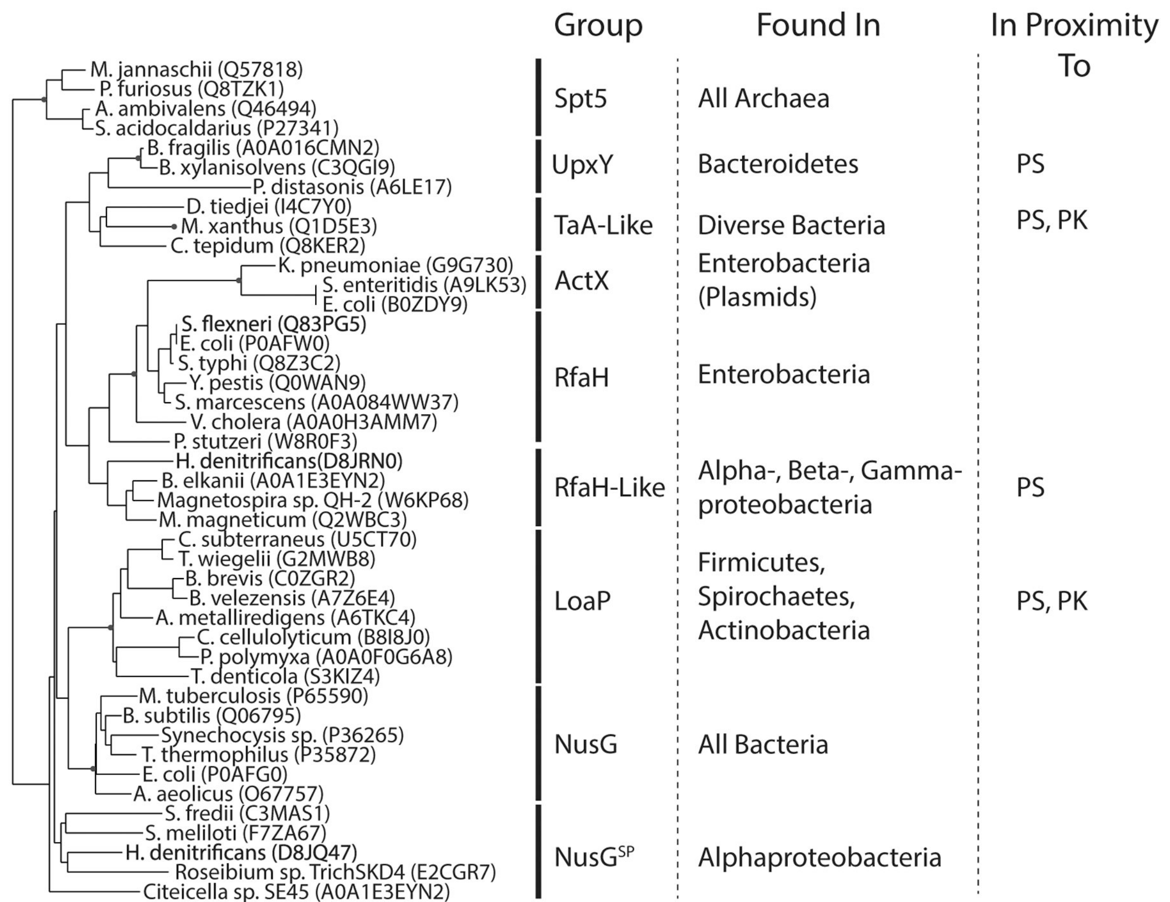

Figure 6: Phylogenetic Tree of NusG, Spt5 and Specialized NusG Paralog Groups.

Represented NusG sequences selected from the subgroups discussed in the chapter form the NusG family. Bacterial sequences from core NusG proteins are found in all bacteria, while a variety of paralogs are found in diverse bacteria phyla. Some groups of NusG paralogs are commonly found in or adjacent to large gene clusters encoding for production of polysaccharides (PS) or polyketides (PK).