Macromolecular assemblies supporting transcription-translation coupling

Michael W. Webster[1,2,3,4,*], and Albert Weixlbaumer[1,2,3,4,*]

[1]Department of Integrated Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)

[2]Université de Strasbourg

[3]CNRS UMR7104

[4]INSERM U1258, 67404 Illkirch Cedex, France

*Correspondence to: Michael W. Webster; Email: websterm@igbmc.fr, or Albert Weixlbaumer; Email albert.weixlbaumer@igbmc.fr

Contents

I. Background
II. Synchronization of transcription and translation rates
III. Regulation of transcription pausing by the ribosome
IV. Regulation of transcription termination by the ribosome
V. Bridging of the transcription and translation machineries
VI. The NusG-coupled expressome
VII. NusA-mediated coupling in *E. coli*
VIII. NusA-mediated coupling in *M. pneumoniae*
IX. The collided expressome
X. Establishment of coupling: RNAP is linked to translation initiation
XI. Perspectives

Abstract

Coordination between the molecular machineries that synthesize and decode prokaryotic mRNAs is an important layer of gene expression control known as transcription-translation coupling. While it has been long known that translation can regulate transcription and vice-versa, recent structural and biochemical work has shed light on its mechanistic basis. Complexes of RNA polymerase linked to a trailing ribosome (expressomes) have been structurally characterized in a variety of states at near-atomic resolution, and also directly visualized in cells. These data are complemented by recent biochemical and biophysical analysis of transcription-translation systems and the individual components within them. Here, we review our improved understanding of the molecular basis of transcription-translation coupling. These insights are discussed in relation to our evolving understanding of the role of coupling in cells.

I Background

In 1961, it was proposed that the transfer of genetic information could be delineated into two core stages, mRNA synthesis and mRNA decoding[1–3]. This profound insight raised the important question of whether these events are physically joined or separated: is the mRNA always released before the first ribosome associates? The team of Marshall Nirenberg showed in 1964 that the mRNA is not always released first in *Escherichia coli* (*E. coli*). Using the DNA-dependent translation system they recently developed to decipher the genetic code, they showed that DNA co-purifies with ribosomes, and that the mRNA is presumably linked to both[4]. It was later confirmed that ribosome binding can occur on endogenous mRNAs that are undergoing active transcription[5]. Electron micrographs of the nuclear material from burst *E. coli* cells revealed that the link between RNA polymerase (RNAP) and ribosomes is very close: the first ribosome bound to a growing mRNA chain is commonly immediately adjacent to the DNA, and sometimes appears to be in direct contact with RNAP[6]. Thus, while most ribosomes are associated with mRNAs that have been fully transcribed, the pioneering round of translation can take place co-transcriptionally. This provided the first evidence that supramolecular assemblies containing both the transcribing RNAP and a trailing ribosome occur in bacteria. Recently, these complexes were termed ‘expressomes’ as they contain all the machinery of gene expression[7]. The potential importance of processes that simultaneously control both core stages of bacterial gene expression was recognized more than 50 years ago[4].

Numerous examples of the mutual regulation of transcription and translation have since been discovered. These are collectively referred to as transcription-translation coupling (‘coupling’ hereafter). One aspect of coupling is the synchronization of rates of mRNA synthesis and decoding (Section II)[8,9]. This supports gene expression efficiency and likely minimizes the energetic excesses of unused mRNA or ribosome queueing[10]. It also allows the ribosome to relieve arrested transcription complexes that can generate collisions with the DNA replication machinery and cause genome instability (Section III)[11]. The regulation of transcription pausing by translation also supports the gene regulatory mechanism of attenuation[12], and likely contributes to genome-wide synchronization[8] (Section III). The regulation of

transcription termination by the ribosome, a form of mRNA quality control and the basis of the polarity effect, is a further important aspect of coupling[13] (Section IV).

Underlying these varied forms of coupling is a significant mechanistic question: is there a physical connection between the transcription and translation machineries and, if so, what does it look like? Regulatory proteins that interact with both the transcription and translation machineries represent an attractive hypothesis for the mechanistic basis of some forms of coupling. A molecular bridge is formed by NusG (Section V)[14,15], a universally conserved transcription factor that performs several important roles. Recent structural analysis of the NusG-coupled expressome has shed new light on how NusG likely contributes to coupling (Section VI)[16,17]. NusA, another important prokaryotic transcription factor, was also recently identified to play a role in coupling (Section VII)[17,18]. In light of the differences between the architecture of NusA-coupled expressomes of different species, the conservation of physical coupling mechanisms is discussed (Section VIII). A short intervening mRNA between the ribosome and RNAP produces the 'collided expressome', an assembly that does not depend on bridging factors[7,16–18]. The possibility this complex performs an important biological role is discussed in relation to its structural features (Section IX). Finally, we review evidence of transcription promoting translation initiation (Section X)[19–21]. We direct interested readers to recent reviews on the physiological roles of coupling and the evolutionary conservation of its mechanisms[22–24].

II Synchronization of transcription and translation rates

Maintenance of ribosome-RNAP proximity depends on coordination between rates of transcription and translation: the ribosome keeps pace with RNAP. More than 50 years ago, it was established that the rates of *lac* gene transcription and translation are coordinated in *E. coli*[25]. It was noted that this equality may either reflect direct coupling between the processes, or an independent fine-tuning of the rates of each over evolution. In support of direct coupling, both total mRNA synthesis and the global transcription elongation rate are halved upon amino acid starvation or treatment with antibiotics that inhibit translation[25–27]. General coordination between rates of transcription and translation in *E. coli* has been confirmed more recently under different growth conditions and for a variety of mutant strains[8,9]. Rates of translation are therefore communicated to the transcription machinery.

Synchronization is supported by multiple molecular mechanisms (**Figure 1A**). Molecular interactions between RNAP and the ribosome occur on the same mRNA ('physical coupling'). The ribosome suppresses pausing of RNAP, and the translation rate can therefore influence the transcription rate (Section III). Physical coupling on individual genes is supported by indirect coordination genome-wide mediated by secondary messengers. It has long been understood that global rates of transcription and translation are coordinated with growth rates and nutrient availability. In some bacteria, including *E. coli*, the primary mediator of this connection is the stringent response, which involves the alarmones (p)ppGpp[28–30]. (p)ppGpp levels affect numerous aspects of gene expression, including transcription of ribosomal RNAs, transcription of genes encoding ribosomal proteins[31], and translation initiation[32]. Upon amino acid starvation, ribosomes become associated with uncharged

tRNAs and are detected by RelA. Activated RelA synthesizes (p)ppGpp, which binds to two sites on *E. coli* RNAP and mediates allosteric regulation of transcription[29,33,34]. This transcription-translation connection was first described as a feedback loop that represses production of rRNA when demand for new ribosomes is low. Yet (p)ppGpp also allosterically regulates transcription elongation of mRNAs, and has therefore emerged as a mechanism that synchronizes rates of transcription with translation[35–37]. While (p)ppGpp production in response to translation stress is highly conserved among bacteria, the regulation of RNAP activity by (p)ppGpp is not. *Bacillus subtilis* (*B. subtilis*) RNAP does not bind (p)ppGpp and is likely affected indirectly[29,38].

Consistent with the effects of alarmones on RNAP, the transcription elongation rate of a reporter gene was recently found to be closely linked to levels of (p)ppGpp[9]. However, this study found no obligate association between transcription rates and rates of translation elongation when protein synthesis was slowed without an increase in (p)ppGpp. Thus, (p)ppGpp can mediate coordination between the rates of ribosomes and RNAP in the absence of physical coupling. The evident importance of alarmones could be interpreted as a challenge to the role of direct contact between RNAP and the ribosome as a general mechanism of coordination[10]. Alternatively, physical coupling may be prevalent when conditions favor efficient translation elongation, whereas conditions of stress and slow translation elongation yield a dependence on alarmones for synchronization of rates. This question has been discussed in detail in a recent review[22].

III Regulation of transcription pausing by the ribosome

Transcription is processive, but interrupted by pauses. The position that RNAP pauses is commonly dictated by the underlying DNA sequence. Concurrent translation regulates transcriptional pausing[39], and the release of paused RNAP by the ribosome likely supports genome-wide coupling of transcription and translation rates (Section II). Transcriptional pauses can be categorized into different classes: mRNA hairpin stabilized pauses[40–42], backtracked pauses[43,44] and consensus motif pauses[45,46]. Each pausing mechanism may occur independently, or represent an intermediate state on a pathway towards a stabilized pause that can be further modulated by transcription factors. The duration of RNAP pausing at each pause class is thought to be reduced by the activity of the ribosome.

The first examples of transcription regulation by the ribosome involved the gene-specific process of attenuation. Some *E. coli* operons contain a short open reading frame near the start of the transcription unit. Translation of the encoded ‘leader peptide’ is used by the cell as a sensor to dictate whether the downstream genes should be transcribed or terminated prematurely (‘attenuated’)[12,47,48]. For example, the leader peptide of the *E. coli* operon for tryptophan biosynthesis contains consecutive tryptophan codons that are only efficiently translated if the availability of tryptophan is sufficiently high (**Figure 1B**). The translation elongation rate alters the formation of mutually exclusive mRNA secondary structures, one of which is an intrinsic transcription terminator. At the terminator, the link between transcription and translation depends on the intervening mRNA, but leader sequences also contain a transcriptional pause sequence upstream that is more directly modulated by the ribosome. Pausing is

stabilized by an mRNA hairpin that forms in the RNAP exit channel[40–42]. It is thought the pause provides time for translation to initiate so that the formation of the downstream terminator depends on the subsequent rate of translation elongation, which in turn reflects the concentration of a specific aminoacylated tRNA[12]. The mechanism of ribosome-mediated pause release within attenuators is not known, but it is likely that translation causes unwinding of the mRNA hairpin that stabilizes the transcriptional pause as the ribosome displays intrinsic helicase activity[49,50].

The effect of translation on transcription pausing is essential to *E. coli* viability[11]. In dividing cells, replication and transcription complexes occupy the DNA concurrently. The replisome travels faster than RNAP, which pauses often, and the machineries therefore collide frequently. RNAP arrests commonly involve backtracking, and collision of the replisome with backtracked RNAP can result in a double strand DNA break[11]. Backtracked transcription complexes are reactivated by Gre factors[51,52]. Strains deficient in Gre factors consequently display genome instability due to accumulation of stalled RNAP. Strikingly, the presence of an open reading frame and actively translating ribosomes abolishes DNA breakage in Gre deficient cells[11]. The translation machinery therefore contributes to the rescue of transcription stalled by backtracking.

Backtracked transcription complexes are also produced when RNAP encounters a DNA-binding protein in its path: a roadblock. Roadblocks can be overcome (transcription readthrough) by cooperation between RNAP molecules that are transcribing the same gene[53]. Yet this only occurs on genes with a strong promoter that supports high RNAP density, while on most genes RNAP will terminate if a roadblock is not resolved[54]. Translation can assist with transcription readthrough of a roadblock[8]. The fraction of elongation complexes that traversed a roadblock increased from 10% to 40% when a ribosome binding site was inserted in a reporter gene. Thus, a trailing ribosome can act like a cooperating RNAP molecule in favoring forward translocation out of a backtracked state.

Recently, biochemical reconstitution of a transcription-translation reaction has supported a model in which the ribosome pushes RNAP out of a backtracked pause[55]. Transcription of DNA immobilized on streptavidin beads produced complexes backtracked by more than 12 base pairs. Upon translation of the mRNA being transcribed, the backtracked states were shortened by at least 6 base pairs. Importantly, the length of the mRNA between the RNAP active site and the ribosome decoding center is the same after backtrack reversal as when the ribosome was translocated into a non-backtracked arrested RNAP. It is therefore likely that direct physical contact between the machineries occurs during backtrack reversal, consistent with a pushing mechanism. The molecular assembly that is expected to occur during ribosome-mediated pause release is the collided expressome (Section IX).

A third type of transcriptional pause occurs at DNA sequences resembling the consensus motif identified by NET-Seq[45,46]. While variants of this motif may trigger the initial interruption of the nucleotide addition cycle leading to stabilized pauses, they do not necessarily contain inhibitory mRNA structures or induce RNAP backtracking. The effect of translation on the transcription of a gene containing this type of pause sequence was recently examined[21]. Whereas pause escape was reduced by the presence of the 30S without translation, it was slightly increased when translation occurred. Thus, there

is evidence that this variety of pause, which is very common across the *E. coli* genome, is also regulated by concurrent translation.

IV Regulation of transcription termination by the ribosome

Polarity is a form of transcription-translation coordination in which increased distance between RNAP and the trailing ribosome causes transcription termination. In 1965, the codons that signal translation termination were identified, revealing that nonsense mutants block gene expression by triggering a premature stop of protein synthesis[56]. It had previously been observed, however, that nonsense mutations not only inactivate the gene in which they are located, but also the genes downstream in the same operon[57]. These were termed polar mutants because their proximity to the start of the operon correlated with the magnitude of their repressive effect on the downstream genes[58].

To understand this interdependence between the expression of genes within a cistron, the molecular mechanism of polarity became a subject of intensive study. Biochemical reconstitution revealed that polarity depends on the presence of a sequence that supports the activity of the transcription termination factor Rho[59]. A coherent picture of polarity thereby emerged in which transcription termination mediated by Rho is suppressed by translation because ribosomes prevent access of Rho to the nascent mRNA[60]. The position of premature stop codons displays polarity because a longer region that is transcribed but not translated increases the potential for Rho to bind and premature termination to occur.

Polarity is not only a product of mutation, but also a form of quality control that halts synthesis of defective mRNAs, and a mechanism of global suppression of gene expression in response to stress. Intragenic terminators that are latent when a gene is being translated become active when translation is suppressed by metabolic stress, such as amino acid starvation[13,61]. The transcription-translation coordination related to the polarity effect allows transcription to also be rapidly halted in response to these environmental signals, likely to save resources in conditions of scarcity. The absence of coupled translation on antisense transcripts makes them targets of Rho[62].

Transcription termination occurs by Rho-dependent and Rho-independent (intrinsic) mechanisms[63]. Rho-mediated termination of transcription involves unstructured C-rich sequences termed *Rho utilization* (*rut*) sites, whereas intrinsic terminators contain an mRNA hairpin followed by a U-rich sequence[64,65]. By trailing closely behind RNAP, the pioneering ribosome on an mRNA masks intragenic *rut* sites, and therefore suppresses the association of Rho. Although polarity generally refers to termination involving Rho, intrinsic terminators are also suppressed by translation[66–69]. Bioinformatic analyses have identified numerous intrinsic terminator motifs within coding sequences that, if active, would generate truncated mRNAs[70]. The measured efficiency of intrinsic terminators is substantially lower in the coding region, however, likely due to shielding by the trailing ribosome[68]. Slow transcription in the leader sequence of the *pyrBI* operon supports coupling with translation that masks a regulatory intrinsic terminator[67]. Synchronization of transcription and translation rates therefore plays an important role in suppressing transcription termination[9].

Transcription and translation regulators, such as NusG and EF-P, support the suppression of transcription termination on genes that are translated efficiently. A competition exists between the binding of NusG to the ribosome and Rho (Section VI)[14,16,71]. Transient separation between the ribosome and RNAP may not lead to termination if NusG maintains contact with the ribosome and is therefore inaccessible to Rho. Strains that lack NusG display reduced Rho-dependent termination at a subset of sites genome-wide that contain sequences less likely to support Rho binding[62]. Inhibition of Rho was unable to rescue the expression of a reporter gene in an *E. coli* cell-free transcription-translation system containing NusG mutants[72]. Collectively, these studies suggest the significance of the interaction between NusG and the ribosome to Rho-mediated termination is context-dependent. The polarity effect is expected to be more pronounced in long, multicistronic transcripts. Slowed translation elongation occurs at sequences encoding polyproline in the absence of translation elongation factor EF-P. The resulting separation between the ribosome and RNAP causes Rho-mediated termination, indicating EF-P contributes to the maintenance of transcription-translation coupling[73].

V Bridging of the transcription and translation machineries

The formation of a molecular bridge between the ribosome and RNAP represents a key mechanism by which translation and transcription is coordinated. Although numerous proteins have long been known to regulate either transcription or translation, the existence of factors that concurrently regulate both transcription and translation was identified relatively recently. NusG and its paralog RfaH contain two domains: an N-terminal domain (NusG N-terminal domain or NGN) that interacts with RNAP[74,75], and a C-terminal Kyrpides-Ouzounis-Woese domain (KOW)[76] that interacts with ribosomal protein uS10[14,19]. It was proposed that if these contacts occurred simultaneously, a bridge would form that could account for the coupling of transcription and translation rates[14].

The interaction between NusG and uS10 was first characterized with isolated proteins rather than intact ribosomes or ribosomal subunits. Despite its potential importance, the role of the NusG-uS10 interaction in transcription-translation coupling therefore remained uncertain given that uS10 is not only a component of ribosomes but has an additional role as a transcription factor. uS10 (also known as NusE) cooperates with transcription factors NusA, NusG and NusB in the transcription of ribosomal RNAs[77], and is recruited by the phage-encoded protein λN to promote expression of the phage genome[78]. In this context, uS10 binds the transcription machinery via NusA and NusG-KOW[79]. However, uS10 may also bind directly to RNAP via the same interface that can bind the NusG-KOW[80]. Other ribosomal proteins (bS1, uS4 and uL2) were previously shown to bind and regulate the transcription machinery[81–83], and it was therefore possible that uS10 similarly performs a ‘moonlighting’ role in isolation from the ribosome[84].

Yet recent data supports a specific functional interaction between NusG and ribosomes. Firstly, NusG co-purifies with endogenous ribosomes, and purified components form a complex with a 1:1 stoichiometry[15]. Furthermore, in a cryo-EM reconstruction of the NusG-bound ribosome, density was

identified adjacent to uS10 that is consistent with the NusG binding site determined with isolated uS10[85]. Co-purification of NusG with ribosomes can be reduced by mutation of the NusG-uS10 interface[15]. In support of a functional role in coupling, mutation of the NusG-uS10 interface increased sensitivity to the translation inhibitor chloramphenicol[15]. The same mutation (NusG F165A) decreased the expression of a luciferase reporter gene in *E. coli* S30 extracts more than three-fold[72]. This study confirmed this is due to regulation of transcription, as a translation-only system, in which mRNA rather than DNA was added, was insensitive to the mutation in the NusG-KOW domain. The critical role of NusG in this system is its interaction with the ribosome rather than with Rho (Section IV), as inhibition of Rho was unable to rescue the effect of the NusG mutation. By contrast, addition of an RNAP mutant that is resistant to backtracking was less sensitive to mutation of the NusG-KOW domain. This strongly supports a model in which NusG aids transcription elongation through its interaction with the ribosome.

VI The NusG-coupled expressome

The mechanistic role of NusG in coupling has been recently clarified by high-resolution structures of RNAP-70S complexes ('expressomes') coupled by NusG[16,17]. NusG binds to both RNAP and uS10 on the ribosome, confirming its predicted role as a molecular bridge (**Figure 2A,B**). In the absence of NusG, the architecture of the expressome is dynamic, with RNAP adopting a wide range of positions relative to the ribosome[16]. The presence of NusG significantly reduces this freedom as RNAP is tethered to the 30S head domain. The linker between the two domains of NusG that forms the bridge consists of 8 residues without secondary structure (*E. coli* 118-125). The bridge is not resolved to high-resolution in the cryo-EM maps, consistent with a degree of structural disorder[16,17]. Variation in the distance between the two domains of NusG is observed among the particles in the dataset, and corresponds to expansion and contraction of the linker residues within the range ~13-30 Å[16]. The RNAP β′ subunit zinc finger domain (β′ ZF) is close to the ribosome surface, but makes no contacts that are consistent across the particles within the cryo-EM dataset[16]. Consequently, the NusG-coupled expressome is dynamic with rotations of up to 30˚ observed. We therefore consider NusG to act as an elastic molecular tether rather than a rigid link.

NusG-mediated coupling does not produce structural changes within the ribosome or RNAP that account for the effects of translation and transcription on each other. A key consequence of NusG-mediated coupling is instead the binding of the intervening mRNA to the surface of the ribosome. The location of the NusG bridge aligns the nascent mRNA emerging from RNAP to a basic surface of ribosomal protein uS3 (**Figure 2C**)[16]. This is expected to minimize formation of mRNA secondary structures that are inhibitory to transcription and translation rates. Pausing and termination of transcription commonly involves formation of mRNA hairpins in the exit channel of RNAP. The potential for the ribosome to conceal part of the mRNA sequence involved in this base-pairing has been long appreciated, such as in mechanisms of attenuation[12]. Furthermore, the ribosome has intrinsic helicase activity and can melt mRNA hairpins using residues lining the mRNA entrance channel[49,50]. The identification of a more extended surface for mRNA docking supported by NusG seems to add a further

tool to the arsenal of the ribosome: the ability to prevent mRNA structures from forming in the first place. This structure-based prediction requires directed mutagenesis experiments to confirm.

Formation of the coupled expressome is also important in what it excludes: transcription termination mediated by the termination factor Rho (Section IV). Rho activity is stimulated by NusG[86], and NusG is essential to *E. coli* due to its role in suppressing unwanted transcription via Rho[87]. NusG interacts with Rho and enhances its activity[62,88]. Each of the NusG domains participates in Rho-dependent termination. Firstly, the NusG-KOW interacts with Rho to promote closure of the hexameric ring around RNA and relieve the mRNA sequence dependence of this activity[71,89]. This interaction is mutually exclusive with coupled expressome formation: the interfaces of the NusG-KOW that interact with uS10 and Rho significantly overlap[14,16]. In particular, mutation of *E. coli* NusG residue Phe165 abolishes interactions with both the ribosome and Rho[15,71]. It is therefore likely that the trailing ribosome captures the NusG-KOW to conceal it from Rho. mRNAs that are being efficiently translated are thereby protected from premature transcription termination, while those that are not efficiently translated are targets of Rho. Through this connection, gene-specific regulators of translation can act as gene-specific regulators of transcription. The second interaction between NusG and Rho, identified recently by biochemical reconstitution and structural analysis, has been proposed to occur during the engagement and inactivation of RNAP[90,91]. Rho binds to the NusG-NGN, RNAP and NusA during this pathway. We observed in structural overlays that simultaneous binding of Rho and the ribosome to RNAP is impossible in all RNAP:Rho states and all expressome states reported to date. Thus, recent structural data support the long-standing model of competition between ribosome and Rho, and between translation and transcription termination[13].

Binding of the NusG-NGN to RNAP contributes to formation of a channel that guides the upstream DNA away from RNAP. It has been suggested that the positively-charged surface of NusG restricts the conformational freedom of the emerging DNA strands, promotes their reannealing, and thereby favors forward translocation of RNAP[75,92]. In the NusG-coupled expressome, the positions of uS10 and NusG-KOW further extend this channel directing the upstream DNA, and this likely enhances reannealing[16]. Further study is needed to assess the presence and magnitude of this effect on transcription elongation rate.

The architecture of NusG-coupled expressomes have been characterized with different intervening mRNA lengths between the start of the ribosome P-site and the RNAP active site: 38 nt and 42 nt in one study[16], and 41 nt, 44 nt and 47 nt in another[17]. The most equivalent samples in the two studies (41nt and 42 nt) are similar but not identical in the relative positions of RNAP and the ribosome. This likely reflects differences in the RNA sequence or differences in data analysis strategy, which is expected because the models represent a consensus of structurally heterogeneous particles. Collectively, the structures indicate that increases in intervening mRNA length produce small changes in a consistent direction: the side of RNAP that is tethered by NusG does not significantly move, while the far side moves away from the ribosome (**Figure 2D**). This is consistent with the notion that NusG and the mRNA each tether RNAP to the ribosome, and as the tension on the mRNA side is decreased with longer intervening sequence this side becomes more mobile[16]. Critically, the interaction between

the mRNA and the basic surface of uS3 is maintained even with the longest intervening mRNA examined[17]. The suppression of mRNA secondary structure formation by the NusG-coupled ribosome therefore likely occurs at distances even greater than has been investigated thus far. The proximity between the machineries necessary for NusG-coupling remains to be ascertained. Based on the available structures, we propose that additional mRNA sequence will spool from the mRNA exit channel of RNAP in a way not incompatible with NusG coupling (**Figure 2D**).

Coupling likely enhances the intrinsic positive effect of NusG on transcription. NusG promotes transcription elongation, and this is due to the binding of the NusG-NGN to RNAP[74,75,92]. The interaction between the NusG-KOW and the ribosome maintains a second point of contact that reduces the frequency of dissociation of the NusG-NGN from RNAP: a type of cooperative binding termed avidity. The proportion of time that RNAP is associated with NusG is therefore expected to be larger in the presence of a trailing ribosome. The interaction between the NGN and RNAP is substantially higher affinity than that between the NusG-KOW and isolated uS10[14]. It is therefore likely that binding of NusG to RNAP precedes the interaction between NusG and the ribosome.

Genome-wide chromatin immunoprecipitation studies have indicated, however, that the occupancy of NusG on transcription units peaks substantially downstream of the start site[93]. A recent study proposes this may be because the ribosome helps deliver NusG to RNAP[85]. Given the important role of NusG to Rho activity (Section IV), the authors used Rho-dependent termination as a readout of NusG occupancy. Disruption of translation produced only partial Rho-dependent termination of transcription. It was hypothesized that this is due to reduced NusG occupancy on RNAP in the absence of a coupled ribosome. To test this, the potential for another mechanism of NusG delivery (the λN Nus complex) was examined, and found to be sufficient to compensate for the absence of translation. Whether the delivery of NusG to RNAP is commonly mediated by the ribosome, or whether the ribosome stabilizes NusG that is already associated with RNAP remains to be fully explored.

VII NusA-mediated coupling in *E. coli*

NusA is a well-characterized transcription regulator that stably associates with RNAP. Two recent studies have shown that NusA, like NusG, can concurrently interact with the ribosome and RNAP to form a molecular bridge[17,18]. The first study examined the structural basis of NusA association with NusG-coupled expressomes reconstituted with purified *E. coli* factors[17], whereas the second visualized NusA-coupled expressomes in *Mycoplasma pneumoniae* (*M. pneumoniae*) cells[18]. Surprisingly, the way in which NusA links the transcription and translation machineries is distinct in each.

NusA is a multi-domain protein that adopts different, context-dependent structural arrangements[42,79]. Two regions of *E. coli* NusA interact with RNAP: the N-terminal domain of NusA (NusA-NTD) binds the β subunit flap-tip helix (β-FTH) to enhance transcriptional pausing[42,94,95], and the C-terminal AR2 domain of NusA (NusA-AR2) binds the α1 subunit C-terminal domain (α1-CTD) to promote RNA-binding by NusA[94,96] (**Figure 2E**). In the context of a hairpin-mediated pause, the NusA KH domains (NusA-KH) also contact the ω subunit C-terminal helix, and the NusA-NTD binds the α2

subunit C-terminal domain (α2-CTD)[42]. The N-terminal portion of NusA is conserved across bacteria and archaea[97,98]. The C-terminal AR repeats present in *E. coli* are not universally conserved, and the NusA-AR2:α1-CTD interaction is therefore not expected in many species, including *M. pneumoniae* (Section VIII)[99].

Structures of *E. coli* expressomes containing both NusA and NusG closely resemble the structures containing only NusG (**Figure 2F**)[17]. The lack of substantial changes in the relative positions of RNAP and the ribosome indicates NusG is sufficient for the adoption of this molecular state. In the expressome, the primary interactions between RNAP and NusA previously characterized are visible (NusA-NTD:β-FTH, NusA-NTD:α2-CTD and NusA-AR2:α1-CTD), but the contact NusA-KH:ω is not. The interaction between NusA and the ribosome is driven by insertion of the NusA KH1 domain (NusA-KH1) into a cleft between ribosomal proteins uS2 and uS5 (**Figure 2E**). The interaction is likely supported by the proximity of acidic residues of NusA-KH1 (E218, E219, D242, D246) to basic residues of uS2 (K105, R108) and of uS5 (R45, R68, R69). The resolution of the interface is limited, however, and the atomic details remain to be determined. The expressome is conformationally heterogeneous, and the ribosome is better resolved than RNAP in cryo-EM maps. Notably, the region of NusA that interacts with the ribosome is significantly better resolved and visible at higher contour levels compared to the remainder of NusA (**Figure 2G**). This is strong evidence that this domain moves with the ribosome rather than with RNAP, as is the case for the NusG-KOW bound to uS10. Thus, the interaction between NusA and the ribosome appears stable and specific. Like the bridge formed by NusG, the NusA connection is flexible rather than rigid. As in previous structures[42], NusA adopts an extended configuration that undergoes internal flexing, and the regions of RNAP contacted by NusA (β-FTH, α1-CTD and α2-CTD) are connected flexibly to the core of RNAP. This flexibility is likely essential as it allows the transcription and translation machineries to perform internal conformational changes necessary for their activity without steric interference. In this respect, NusA was likened to a railway pantograph, which maintains contact between a train and overhead power lines despite unpredictable movements[17].

Given the structural similarity between NusG-coupled expressomes with and without NusA, the role of the interaction between NusA and the ribosome is not obvious. It was observed that a higher fraction of imaged particles was assigned to structurally well-defined classes when NusA was present[17]. This suggests NusA may promote the formation, or maintain the stability, of NusG-mediated coupling. Alternatively, or in addition, the activity of NusA may be modulated by a change in its mode of binding RNAP. Comparison of the RNAP:NusA complex in the expressome to that of NusA stabilizing a nascent mRNA hairpin indicates the β-FTH must rotate to allow the NusA-NTD to bind to it without clashing with the ribosome (**Figure 2H**). The global change in NusA position is likely stabilized by the interaction between KH1 and the ribosome. The consequences of NusA remodeling in the coupled expressome to transcription pausing, termination and antitermination remain to be examined. Finally, NusA may modulate mRNA hairpin formation in the intervening mRNA. The path of the mRNA along the surface of uS3 is shielded on one side by the S1 domain of NusA (NusA-S1), and this shielding may reduce access by other RNA-binding proteins. Although NusA has the potential to bind RNA[99], the cryo-EM

density indicates the nascent mRNA still contacts uS3 as it does in the coupled expressome lacking NusA. It remains to be tested whether NusA suppresses or promotes mRNA hairpin formation in the context of the expressome. A major limitation in completing our understanding of the role of NusA in coupling is the lack of an *E. coli* expressome structure with NusA but without NusG.

NusA and NusG are essential in wild-type *E. coli*[100–102]. NusG is essential in *E. coli* due to its role in suppressing the *kil* gene derived from the prophage *rac*[87]. This indicates that, surprisingly, the varied roles of NusG in regulation of transcription elongation, Rho-mediated transcription termination at other sites and transcription-translation coupling are dispensable to viability under ideal growth conditions. NusG is not essential in *B. subtilis* and *Staphylococcus aureus*[97,103]. Although deletion of NusA is lethal, an *E. coli* strain that expresses only the first N-terminal 343 residues is viable, indicating that the C-terminal residues 344-495 are dispensable under typical growth conditions[104]. This is consistent with the known domain architecture of NusA, as the conserved domains are present in the N-terminal portion (**Figure 4B**).

VIII NusA-mediated coupling in *Mycoplasma Pneumoniae*

A NusA-coupled expressome was identified in *M. pneumoniae* by a combination of in-cell cross-linking mass spectrometry and cellular electron tomography (cryo-ET)[18]. The formation of chemical cross-links is an indicator of physical proximity of proteins in the cell, and it was found that NusA, but not NusG or core RNAP, cross-linked to ribosomal proteins. A subset of the cryo-electron subtomograms of ribosomes in *M. pneumoniae* cells were found to be physically associated with RNAP, NusA and NusG. The regions of the map containing RNAP and NusA were resolved to ~10 Å, allowing a model of the expressome complex to be built that is consistent with the cross-linking data (**Figure 3A**).

The architecture of this *M. pneumoniae* expressome is different to all *E. coli* expressome states characterized to date. Compared to the *E. coli* expressome coupled by NusG and NusA[17], RNAP in the *M. pneumoniae* expressome is rotated almost 180˚ and is adjacent to the 30S body rather than the 30S head (**Figure 3B**). The NusG-NGN is bound to RNAP, but the position does not permit the NusG-KOW to bind uS10. NusA, on the other hand, interacts with both RNAP and the ribosome. The complex is therefore in an exclusively NusA-coupled state. NusA-KH1 interacts with uS3 adjacent to the mRNA entrance channel of the ribosome. While relatively close to the NusA-binding site observed in *E. coli* formed by uS2 and uS5, this represents an entirely distinct interface (**Figure 3C**). The binding site occludes the mRNA path on the surface of uS3 observed in the *E. coli* coupled states[16,17], and redirects the intervening mRNA away from the 30S head. While density for the mRNA was not directly observed, the RNA-binding surface of NusA is oriented toward the channel connecting the transcription and translation machineries expected to contain the mRNA. Despite the structural differences, sequestration of the intervening mRNA therefore appears to be a shared feature of coupled expressomes.

The dissimilarity between models of *E. coli* and *M. pneumoniae* expressomes raises important questions regarding the conservation of coupling mechanisms. An analysis of the protein sequences of

NusG and NusA across bacterial species indicates it is perhaps not surprising that coupling in *M. pneumoniae* is unlike that observed in *E. coli* (**Figure 4A,B**). Firstly, the amino acid linker that separates NGN and KOW domains of NusG is substantially different between the species: eight amino acids in *E. coli* and more than 100 in *M. pneumoniae* (**Figure 4A**). Analysis of the *M. pneumoniae* NusG sequence with AlphaFold[105] indicates this expansion is predicted to fold into an additional domain that extends the NGN. It is therefore unlikely that *M. pneumoniae* NusG would act as a tight molecular tether in the way that *E. coli* NusG does in the coupled expressome.

The NusG-KOW domain contains a curved sheet with four β-strands. The residues that contact uS10 in the coupled *E. coli* expressome are located within loops between strands: *E. coli* residues 139-141 and 164-167[14,16]. Phenylalanine residue 165 (F165) plays a key role in the interaction by inserting into a hydrophobic cavity of uS10, and its mutation to alanine is sufficient to abolish the association between NusG and ribosomes and impair functional coupling[15,72]. Notably, the residues that contact uS10 in *E. coli* are all different in *M. pneumoniae*, including the equivalent of F165 (**Figure 4A**). To examine whether variation of these residues is common across bacterial species, we determined a consensus sequence for the NusG-KOW domain using representative species from 74 different taxonomic classes. This indicated that the relevant loops are highly conserved across the bacterial kingdom, and *M. pneumoniae* is a notable exception. Further analysis of the consensus sequence of NusG-KOW domains from 81 species of *Mycoplasma* indicated that even within its own genus *M. pneumoniae* is unusual in lacking these conserved residues (**Figure 4A**). Similarly, *E. coli* uS10 residues M88 and D97, which have been shown to support the interaction of uS10 and NusG-KOW *in vivo*[15] are not conserved in *M. pneumoniae* (**Figure 4C**). Hence, *M. pneumoniae* NusG is not expected to interact with uS10 in the way that has been characterized. Conservation of NusG F165 is not necessarily a reliable marker of species expected to display NusG-mediated coupling, however, as it is also important for the binding of NusG to Rho and stimulation of Rho ring closure[71].

An equivalent analysis of why the interaction between the ribosome and NusA observed in *E. coli* was not seen in *M. pneumoniae* was less conclusive (**Figure 4B**). The interaction interface between NusA-KH1 and the ribosome is less well resolved than that of NusG. The conservation of the regions that were modelled closest to the ribosome[17] nonetheless indicate they are mostly different in sequence between *E. coli* and *M. pneumoniae*, but are poorly conserved across bacterial species in general. The converse question, why the NusA-coupling arrangement observed in *M. pneumoniae* was not observed in *E. coli* is potentially easier to explain. In-cell cross-link mapping in *M. pneumoniae* identified contacts between the NusA proline-rich and disordered domains and the ribosome, which are not present in *E. coli*[18]. If these are essential to expressome formation in this arrangement, this state will not occur in *E. coli*.

Physical coupling mechanisms may have evolved more than once, or diverged from a common ancestor to their present states. Our current understanding of coupling across bacteria and archaea is limited[23]. The best prediction of the existence of physical coupling in different bacterial species comes from a recent analysis of genomic signatures of its absence[106]. Transcription outpaces translation in *B. subtilis* ('runaway transcription'), and coupling is therefore unlikely. Proximity of intrinsic transcription

terminators to the end of open reading frames is suggestive of runaway transcription, as these would not be functional if the ribosome were closely trailing RNAP when they are transcribed. By this definition, approximately half of Firmicutes (which includes *B. subtilis* and *M. pneumoniae*) display runaway transcription[106]. *M. pneumoniae* is difficult to assign by this method, however as it contains few intrinsic terminators. Given the differences in expressome architecture, the primary sequences of NusG and NusA of *M. pneumoniae*, and that it is a member of a taxonomic class that largely does not display coupling, an independent mechanism of NusA-mediated coupling may have evolved in *M. pneumoniae* from an ancestor that displayed runaway transcription. Future identification of the amino acids involved in the interaction between NusA and the ribosome in *M. pneumoniae* and *E. coli* are needed to better predict species that contain each mode of coupling.

The possibility that additional forms of coupling occur should not be discounted. These might involve different arrangements of NusG and NusA or the binding of other regulatory factors. Reconstitution of *E. coli* expressomes depends on selection of the purified components likely involved[16,17], and a systematic characterization of possible contribution by other proteins is warranted. The cryo-ET data on *M. pneumoniae* revealed only one coupled state, which involved NusA-mediated coupling[18], but it is likely not the only expressome complex that occurs. Besides the possibility that alternative expressome states occur under different growth conditions, the cryo-ET method relies on classification of particles to identify those that are sufficiently uniform in composition and conformation for resolution of structural features. In the analysis of 70S ribosome subtomograms, it was identified that 27% contained additional density adjacent to the mRNA entrance channel. The size of the density is consistent with that of RNAP, and it is therefore likely that most of these are expressomes. Only ~15% of these particles (~4% of the total) were subsequently assigned to a state with clear density for NusA. Thus, ~85% of expressome particles are in states different in conformation or composition to the NusA-expressome described so far. Further analysis will be needed to characterize these complexes. It also remains to be determined whether NusA produces the structurally well-defined state, or if a structurally well-defined state is produced for other reasons and this coincides with the positioning of NusA observed.

IX The collided expressome

When the mRNA connecting the transcription and translation complexes is short, an expressome complex with a distinct architecture occurs (**Figure 5A**). Compared to the NusG-coupled state, RNAP is rotated by almost 180° relative to the ribosome within a plane approximately parallel to the ribosome surface (**Figure 5B**). NusG is unable to contact uS10 in this orientation, and the binding site for NusA on the ribosome is occluded by the RNAP core. This state is therefore not coupled by additional factors.

This complex has been structurally characterized in four independent studies[7,16–18]. In the first study, *in vitro* translation of an mRNA bound to a stalled RNAP was performed[7]. This allowed the ribosome to translate until it was physically impeded by RNAP: a molecular collision. We therefore

termed this state the 'collided expressome' to distinguish it from those where bridging factors mediate coupling. A cryo-EM reconstruction of the resulting complex at 7.6 Å resolution allowed unambiguous docking of the relative positions of the constituents. A structurally equivalent assembly was identified in *M. pneumoniae* cells treated with the transcription inhibitor pseudouridimycin (**Figure 5C**)[18]. By halting transcription, the trailing ribosome collides in a similar manner to the sample prepared *in vitro*. The cryo-ET reconstruction was similarly resolved to approximately 7 Å. High-resolution reconstructions (3-4 Å resolution) were determined by biochemical reconstitution with *E. coli* factors and a nucleic acid scaffold with a short intervening mRNA[16,17].

In the collided expressome, the mRNA exit channel of RNAP is adjacent to, and aligned with, the mRNA entrance channel of the ribosome (**Figure 5D**). The nascent mRNA therefore travels directly from the site of synthesis to decoding. The tension of the mRNA likely drives the relative orientation of RNAP and the ribosome, as the mRNA path length is minimized by the observed architecture[16]. RNAP and the ribosome are in close contact as their surfaces display shape complementarity. The center of RNAP is located within the cleft formed by uS3, uS4 and uS5 that surrounds the mRNA entrance channel. The position is further confined by the position of the RNAP α1 subunit within the cleft formed by uS3 and uS10 (**Figure 5D**). While it is possible that direct contacts between RNAP and the ribosome support the relative positions of the ribosome and RNAP in the collided expressome, they are not stable. This was evident from an analysis of molecular dynamics of the complex by cryo-EM: the position of RNAP relative to the ribosome varies by rotation of up to ~90° about the mRNA axis[16]. No specific contacts are maintained by this degree of heterogeneity, and the molecular model represents a consensus position only. The structures of the ribosome and RNAP within the collided expressome are not detectably altered by formation of the collided expressome with one exception: the RNAP ω subunit is consistently absent[16,17]. The cause and consequence of this remains to be assessed.

While this expressome state is the most structurally well-characterized, its role is debated. We propose that this structural state occurs frequently but transiently in cells, when a ribosome catches up to an RNAP that is paused. The architecture therefore represents an important but short-lived event: the release of RNAP from a paused state by the ribosome (see Section III)[8,21,55]. That this state was observed only in cells treated with transcription inhibitor supports the notion that it is infrequent or transient[18].

It remains to be determined, however, whether this molecular state occurs in the cell exclusively upon collision, or if active transcription and translation proceeds with this tight association between the machineries. The state is sterically incompatible with binding of NusA to RNAP and the formation of mRNA secondary structures that support transcription termination and pausing. For the collided state to also occur during active transcription and translation, it must therefore regularly form and separate for these important regulatory events to occur. A precise match of transcription and translation rates at the level of individual molecules would be required to maintain the tight association of RNAP and the ribosome during elongation. No evidence of such a consistent intervening mRNA length has been presented to date.

Incompatibility of the collided expressome with NusA binding raises a key unanswered question: when the ribosome approaches a paused RNAP, does NusA prevent formation of the collided expressome, or does the action of the ribosome displace NusA from RNAP? The addition of NusA and NusG did not alter ribosome-mediated reversal of backtracking *in vitro*, indicating NusA may be readily displaced when the transcription and translation machineries come into close contact[55]. Structural studies suggest translocation of the ribosome disrupts NusG-mediated coupling when the intervening mRNA length becomes short, and displacement of NusA might similarly occur. Further experiments are needed to determine the effect of translation on the association and dissociation of transcription factors.

Steric incompatibility with the formation of mRNA secondary structure could, by contrast, be a key attribute of the collided expressome. The ribosome can suppress termination of transcription at intrinsic terminators that depend on mRNA hairpins[66–68], and may aid transcription elongation by preventing formation of pause hairpins (Section III). Prevention of the formation of mRNA hairpins that support transcription pause and termination likely supporting this role of the ribosome. Depending on the kinetics of translation and transcription, the nascent mRNA could be sequestered by the ribosome in the collided expressome, and this state could also follow melting of pause mRNA hairpins prior to resumption of transcription.

It has been proposed that the collided state is not, as it was first described, a transcribing-translating assembly, as the ribosome may not be able to translate[17]. This was based on the potential for steric incompatibility between the collided expressome and an aligned model of the ribosome undergoing EF-G catalyzed translocation, which is accompanied by swiveling of the small subunit head domain. Yet structural superposition reveals no clashes between the ribosome in a swiveled state and the position of RNAP in the collided expressome (**Figure 5E**). In fact, swiveling increases the separation between ribosome and RNAP. This is true for each deposited collided expressome model[7,16–18] and different models of swiveled 30S conformations[107–109]. If the collided expressome is supported by direct physical interactions between RNAP and the ribosome, ribosomal movements associated with translocation would be hindered by disruption of the interface. As detailed above, this is unlikely as direct contacts were not resolved in EM density maps and the assembly is highly dynamic. Translation inhibition of this form should therefore not be seen as a challenge to the physiological relevance of the collided expressome.

There is reason, however, to expect translation inactivation could be a key attribute of ribosome-RNAP collisions. In some instances, such as upon nucleotide misincorporation, halted RNAP complexes are formed that must be protected from interference by translation. The forward force of ribosome translocation in the absence of transcription could conceivably promote resumption of transcription from a backtracked state before proof-reading of the transcript has occurred, which would be disadvantageous to the cell. The formation of a collided expressome in which translation is halted would resolve this potential problem. Consistent with the notion of translation inactivation upon collision, the in-cell architecture of expressomes formed by transcription inhibition shows the ribosome to be in a stalled, intermediate translocation state with EF-G bound[18]. In the collided expressome formed by translation *in vitro*, tRNA was not observed in the A-site, which also indicates an impediment to further

translation[7]. It is therefore possible that the architecture of the expressome has evolved to control contact between RNAP and the ribosome so as to support ribosome stalling. The existence and mechanism of this type of regulation remains to be explored.

X Establishment of coupling: RNAP is linked to translation initiation

Stimulation of transcription by translation elongation is not the only form of transcription-translation coupling. In the past 10 years, a number of studies have indicated that the close relationship between the processes can begin earlier, when the transcription machinery aids initiation of translation. The first evidence of this followed analysis of the transcription factor RfaH. RfaH is a paralog of NusG, and can similarly interact with both RNAP and ribosomal protein uS10[19,75]. Like NusG, RfaH accelerates transcription elongation by suppressing pausing[110,111]. The binding sites of RfaH and NusG on RNAP are overlapping, consistent with the structural similarity of their N-terminal domains[75]. RfaH interacts with RNAP with higher affinity than NusG, and consequently the binding of RfaH excludes binding of NusG[112]. RfaH activates expression of a small number of operons associated with cell wall biosynthesis, conjugation and virulence[112], each of which has a DNA sequence downstream of the promoter termed the *ops* element[113]. This sequence induces pausing of RNAP, and is recognized by RfaH when it is within the transcription bubble.

In a study where the *ops* element was inserted upstream of a reporter gene, it was observed that RfaH partially compensates for the removal of the ribosome binding sequence[19]. Thus, RfaH likely aids mRNA recruitment by tethering the 30S subunit to RNAP, and thereby promotes translation initiation. Such a role is consistent with the sequences of operons that contain *ops* elements, as they commonly lack strong ribosome binding sites and contain non-canonical start codons due to their origin as horizontally acquired genes. The *ops* element acts as a transcriptional pause primarily to allow time for RfaH to stably associate. It is possible, however, that it also provides time for RfaH to subsequently recruit the 30S subunit and translation to commence – at least in cases where the *ops* element is located downstream of the start codon. By guiding the translation initiation machinery to nascent mRNAs rather than mature mRNAs, RfaH may thereby support the establishment of transcription-translation coupling.

Whether transcription-directed initiation of translation is limited to RfaH and its target genes has yet to be fully explored. Two recent studies, however, provide evidence that it may be a general mechanism of prokaryotic gene expression. *In vitro* biochemical and single-molecule methods have indicated that even in the absence of additional transcription factors, the 30S subunit binds approximately two-fold more rapidly to an mRNA associated with a paused RNAP than to the corresponding free mRNA[21]. The dissociation rate of the 30S was also slightly lower when RNAP was present. These observations are consistent with a direct physical interaction between RNAP and the 30S, which would increase the probability of 30S subunits locating the mRNA. Such an interaction has been identified through the co-purification and cross-linking analysis with purified factors in the absence of mRNA[20,114].

A structure has been determined of a complex containing 30S and RNAP in the absence of DNA and mRNA (**Figure 6A,B**)[20]. The zinc finger domain of the RNAP β′ subunit (β′ ZF) makes extensive contacts with ribosomal protein uS2. The β′ flap-tip helix (FTH) is also close to the 30S, but no cryo-EM density was observed for this domain. This indicates it remains disordered, unlike in complexes in which the FTH contributes to binding of other factors[42]. Notably, additional density was observed bound to RNAP, and was assigned to ribosomal protein bS1 (**Figure 6C**). bS1 was previously found to interact with RNAP[81], and chemical cross-linking experiments are consistent with the identified interaction site[20]. Given that bS1 interacts with both 30S and RNAP, it is possible that it contributes to complex formation. Yet bS1 was found to be dispensable for the interaction, and a structurally similar 30S-RNAP complex was identified following depletion of bS1[20]. Thus, the architecture of the RNAP-30S complex is driven primarily by the interaction between the RNAP domain β′ ZF and ribosomal protein uS2. The position of the β′ ZF overlaps with the binding site of the bS1 OB1 domain, which is not visible in the map[115]. Binding of bS1 domains to the ribosome surface is dynamic[116], and complete dissociation of bS1 commonly occurs during isolation of ribosomes. The binding of RNAP to the 30S subunit presumably depends on the transient dissociation of the OB1 domain from uS2.

The 30S-RNAP structural model may represent a state that occurs during recruitment of 30S to RNAP to aid translation initiation. If this is the case, it shows that transcription factors such as RfaH may aid recruitment but are not absolutely required. In the structure, the mRNA exit channel of RNAP is aligned with the 30S cavity that contains the 16S anti-Shine-Dalgarno sequence (**Figure 6B**). Once the newly-synthesized Shine Dalgarno sequence emerges from RNAP it would therefore be delivered to the recognition site of the 30S. Further investigation is needed to confirm this hypothesized role of the 30S-RNAP complex. The use of RNAP without bound DNA and mRNA in this study raises the question of whether this structural state can occur during active transcription. Superimposition of a structural model of RNAP in an elongating state showed no clashes between RNAP and 30S. However, the position of one of the bS1 OB domains is incompatible with the expected position of DNA (**Figure 6D**). Thus, the binding of bS1 to RNAP may not be relevant to transcription-translation coupling. It could instead represent the proposed contribution of bS1 to RNAP recycling[117]. This does not detract from the likelihood of this state supporting 30S recruitment, however, as bS1 was dispensable to the 30S-RNAP interaction.

NusG promotes 30S binding to mRNA *in vitro*[21]. While the 30S association rate was increased by 70% by the presence of RNAP, it was increased 115% by the presence of RNAP and NusG. This suggests that NusG may bridge RNAP and 30S in a manner similar to that observed in the coupled expressome. By contrast, RfaH had no effect on the 30S association rate[21]. The model gene examined in this study did not, however, contain the *ops* element that is needed to relieve autoinhibition of RfaH[118]. The role of RfaH therefore needs to be further examined in the context of its target genes. Despite this, the addition of RfaH decreased the dissociation rate of the 30S from RNA approximately 2.5-fold[21]. It is possible that, after the association of 30S with RNA, the autoinhibition of RfaH that is mediated by the binding of the two domains of RfaH is stochastically overcome and RfaH-mediated coupling stabilizes the complex.

XI Perspectives

In the almost 60 years since coupling of transcription and translation was first proposed, our understanding of it has progressed on many fronts. RNAP coupled to a translating ribosome has been reconstituted *in vitro*[8,55,119], structurally characterized close to atomic resolution[16,17], directly observed *in situ*[18], and corroborated by *in vivo* super-resolution data[120,121]. Nevertheless, many important and fundamental questions remain. For example, it is not clear how frequently, and in what contexts, physical coupling occurs *in vivo*. Likewise, the functional role of coupling remains under debate[10,122]. Although both bacteria and archaea appear to be able to couple transcription and translation, it is likely not true for all prokaryotic species[106,123]. Key insights into the importance of coupling could be gained through transcriptomics and ribosome profiling of modified strains in which coupling has been compromised in various ways.

High-resolution structures have clarified the likely molecular mechanisms of coupling. In particular, a distinction between expressomes coupled by transcription factors and a collided state, in which the core machineries are in close contact, is now evident. Yet the biological role of collisions between the ribosome and RNAP remains an important question. Is a trailing ribosome able to distinguish RNAP molecules that are stalled due to nucleotide misincorporation from those that are either arrested and backtracked or halted by a regulatory pause? In the first case, the ribosome may pause translation until the erroneous base has been removed to maintain transcription fidelity. In the other two cases, RNAP may benefit from the mechanical push of the ribosome to be released from the arrested state or regulatory pause. Precise reconstitution followed by high-resolution cryo-EM reconstructions could reveal possible allosteric effects of the ribosome on RNAP. These studies will complement biochemical and single-molecule approaches that measure base misincorporation and pause escape rates, and complex dynamics in a coupled *in vitro* transcription-translation system. Precise reconstitution using subsets of charged tRNA and NTP substrates may also allow us to recapitulate and gain insights into specific states that occur during attenuation[39,47,48,124].

NusG is not the only transcription factor able to form a physical bridge between RNAP and the ribosome[17,18]. Understanding if, and how, NusA or RfaH couple RNAP to the ribosome in absence of NusG is an important unanswered question. In association with RNAP, NusG or RfaH may also support recruitment of the small ribosomal subunit to the start codon or stabilizing a 30S initiation complex[19,21]. This distinct mode of coupling is only poorly understood.

Finally, the visualization of RNAP coupled to the ribosome using cryo-ET represents a technological breakthrough that will serve as a roadmap for the study of molecular machineries *in situ*[18]. In the future, combining single particle cryo-EM approaches with cryo-ET will provide a powerful approach to investigating this type of complex and dynamic structural question. Whereas the former allows precise control over complex composition and the state of the assembly in question, the latter provides the advantage of studying the complex in its native environment.

Studies on the bacterial expressome reiterate a general lesson in biology: the whole can be greater than the sum of its parts. The expressome combines the two key steps of gene expression in one molecular assembly that facilitates direct interplay between RNAP and the ribosome. It therefore represents a higher level of structural organization, in which enzymes with fundamental roles in biology are organized in supramolecular complexes for better coordination of their activities[125].

Acknowledgments

We thank members of the Weixlbaumer laboratory for critical reading of the manuscript. We thank Robert Landick, Max E. Gottesman and Ruben L. Gonzalez, Jr. for their feedback on the manuscript. Funding: The authors were supported by the French Infrastructure for Integrated Structural Biology (FRISBI ANR-10-INBS-05, Instruct-ERIC, and grant ANR-10-LABX-0030- INRT, a French State fund managed by the Agence Nationale de la Recherche under the program Investissements d'Avenir ANR-10- IDEX-0002-02). The work was supported by an EMBO long-term fellowship to M.W.W. and the European Research Council starting grant TRANSREG (679734) to A.W.

M.W.W. prepared the manuscript and figures with input from A.W.

Figures

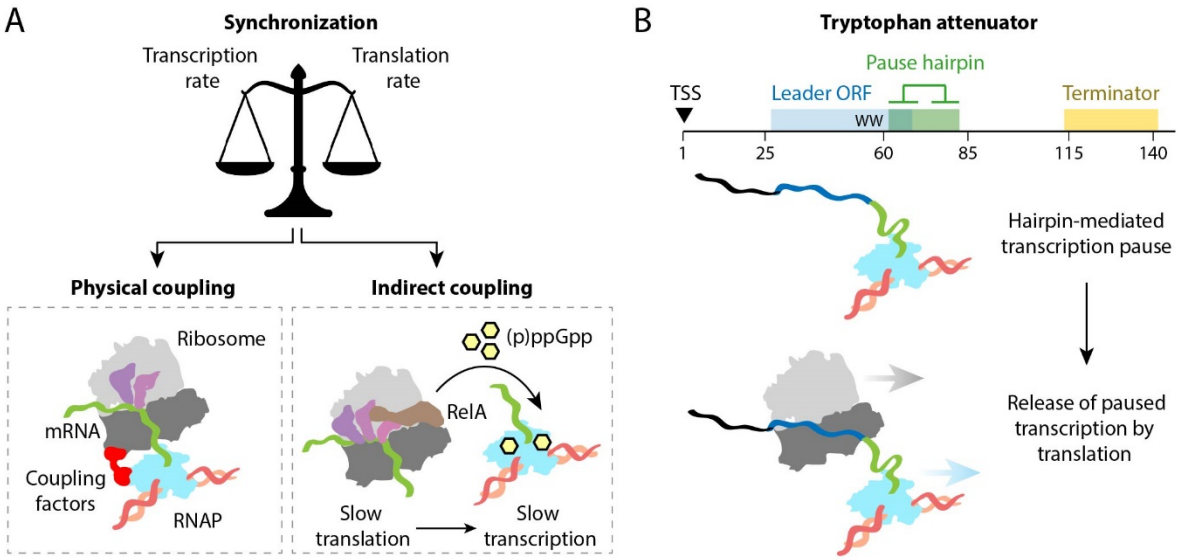

Figure 1: Regulation of transcription by translation. **(A)** The close correlation between rates of transcription and translation observed in some bacteria is supported by both physical coupling and indirect coupling mechanisms. Concurrent synthesis and decoding of prokaryotic mRNAs allow direct contact between the ribosome and RNAP that can be mediated by the coupling factors NusG and NusA. Reduced rates of translation can also suppress transcription rates more generally, through the activity of the alarmone (p)ppGpp. **(B)** Regulated release of transcription pausing by the ribosome in the tryptophan operon. Like several other biosynthetic operons, the sequence downstream of the transcription start site (TSS) contains a short open reading frame (leader ORF). Formation of a hairpin in the nascent mRNA at the end of this sequence causes a programmed transcription pause that provides time for translation initiation. Translation is thought to then promote resumption of transcription, allowing formation of the downstream intrinsic terminator structure to be conditional on the coordination of transcription with translation. Coordination is modulated in the case of the tryptophan attenuator by the availability of tryptophan, which affects the efficiency of translation of consecutive codons encoding tryptophan in the leader ORF (indicated by WW).

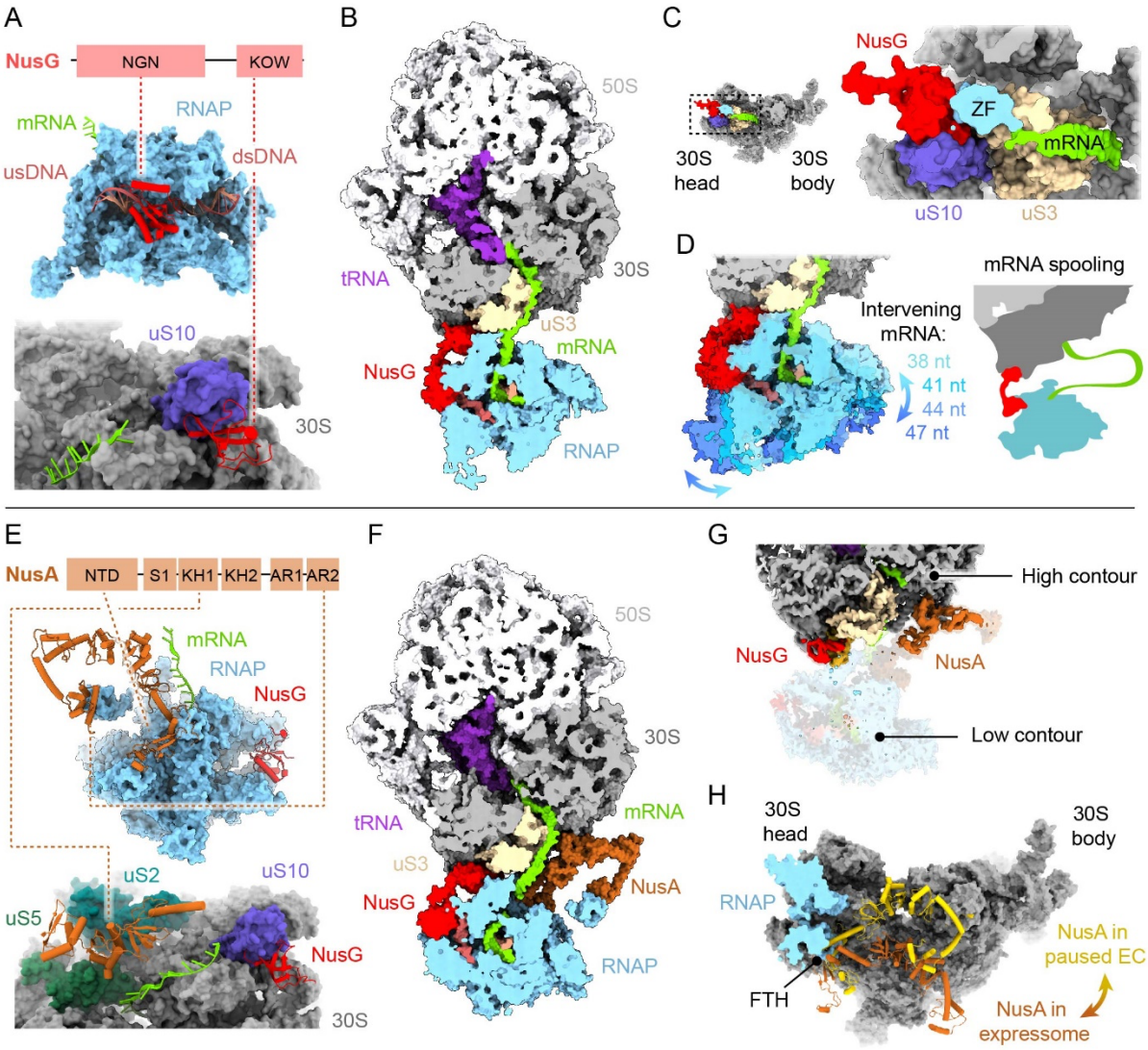

Figure 2: Structural details of the *E. coli* expressomes coupled by NusG (A-D) and coupled by both NusG and NusA (E-H). **(A)** NusG contains two structurally-independent domains, which interact with RNAP (top) and ribosomal protein uS10 (bottom). Upstream DNA (usDNA) and downstream DNA (dsDNA) are indicated. **(B)** Structure of the NusG-coupled expressome. Slice through atomic model (PDB: 6ZTJ) shown in surface representation. **(C)** RNAP binding site on the 30S head domain causes the intervening nascent mRNA to traverse ribosomal protein uS3. The RNAP β′ subunit zinc finger domain (ZF) is closest to the ribosome surface. **(D)** Coupling by NusG is supported upon increase in the length of the intervening mRNA. Small rotation of RNAP with increased mRNA length permits spooling of the intervening sequence while maintaining coupling by NusG. **(E)** Domain architecture of NusA, which interacts with RNAP through the NTD and AR2 domains, and with the ribosome through the KH1 domain. **(F)** Structure of the NusG-NusA-coupled expressome. Slice through atomic model (PDB: 6X7K) shown in surface representation. **(G)** Cryo-EM map of NusG-NusA-coupled expressome (EMDB: 22087) shown at two contour levels. Density for the domains of NusG and NusA that contact the ribosome is well-defined at high-contour levels, indicating they move with the ribosome rather than RNAP. **(H)** Movement of NusA associated with expressome binding. The position of NusA observed in hairpin-mediated pause (yellow) is incompatible with ribosome binding. NusA in the expressome

(brown) is rotated about its contact with the RNAP FTH domain, which is flexibly linked to the RNAP core. Molecular graphics and analyses performed with UCSF ChimeraX version 1.2.4[126].

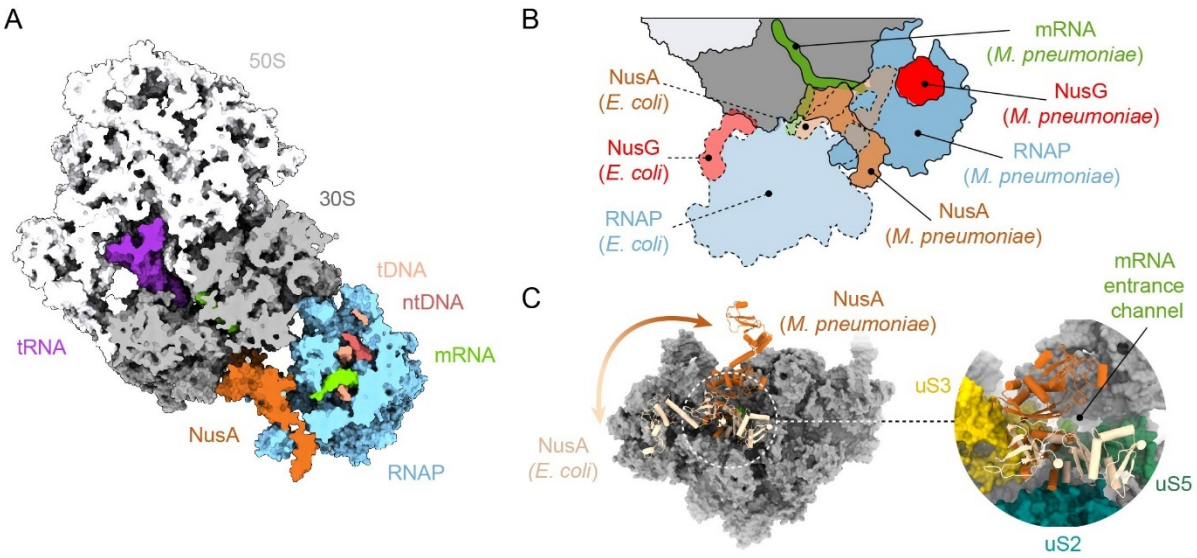

Figure 3: Structural details of the *M. pneumoniae* NusA-coupled expressome. **(A)** Structure of the NusA-coupled expressome. Slice through atomic model (PDB_Dev accession code: PDBDEV_00000049) shown in surface representation. **(B)** Comparison of NusA-coupled expressomes from *M. pneumoniae* (solid lines) and *E. coli* (dashed lines). The position of RNAP (blue) with associated factors NusA (brown) and NusG (red) relative to the ribosome (gray) differs substantially between species. The path of the mRNA (green) is a prediction for *M. pneumoniae*, for which density was not visible in the cryo-ET data. **(C)** Binding of *M. pneumoniae* NusA (brown) to the ribosome involves a distinct interface to that of *E. coli* NusA (cream). Molecular graphics and analyses performed with UCSF ChimeraX version 1.2.4[126].

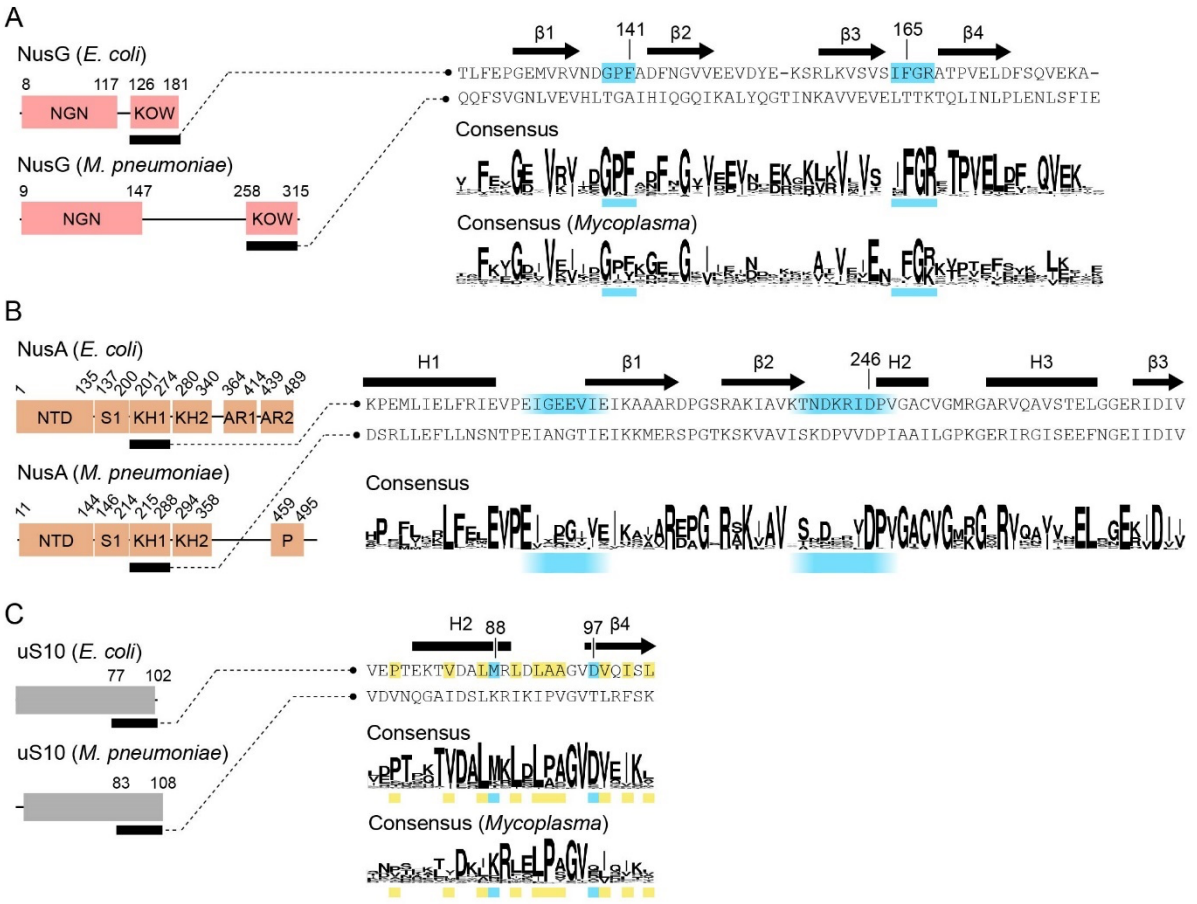

Figure 4: Conservation of sequence elements involved in transcription-translation coupling among bacterial species. **(A)** Conservation of NusG amino acid sequences. Left: domain architectures of NusG from *E. coli* and *M. pneumoniae* with numbers indicating residues at the boundary of the N-terminal domains (NGN) and C-terminal KOW domains (KOW). Right: comparison of the sequences of the KOW domains of *E. coli* and *M. pneumoniae*, and consensus sequence logos of NusG KOW domains generated from 74 representative species from different bacterial taxonomic classes (top) and 81 species of the genus *Mycoplasma* (bottom). Residue positions involved in contacting uS10 in the NusG-coupled *E. coli* expressome are indicated in blue. **(B)** Conservation of NusA amino acid sequences. Left: domain architectures of NusA from *E. coli* and *M. pneumoniae* with numbers indicating residues at the boundaries of domains. Right: comparison of amino acid sequences of NusA-KH1 domain of *E. coli* and *M. pneumoniae* and consensus sequence logo generated from 68 representative species from different bacterial taxonomic classes. Residue positions possibly involved in contacting the ribosome in the NusA-NusG-coupled *E. coli* expressome are indicated in blue and faded blue (indicating lower confidence). Acidic repeat domains (AR1 and AR2) are present in *E. coli* but not *M. pneumoniae*, and proline-rich region (P) is present in *M. pneumoniae* but not *E. coli*. **(C)** Conservation of ribosomal protein uS10 sequences. Comparison of indicated amino acid sequences of uS10 of *E. coli* and *M. pneumoniae*, and consensus sequence logos generated from 67 representative species from different bacterial taxonomic classes (top) and 84 species of the genus *Mycoplasma* (bottom). Residue positions involved in contacting NusG-KOW identified by mutagenesis[15] indicated in blue, and residue positions

of the hydrophobic surface contacting NusG-KOW[16] indicated in yellow (as well as *E. coli* residue 25, which is conserved but not shown). Sequence analyses performed with Clustal Omega[127] and WebLogo[128].

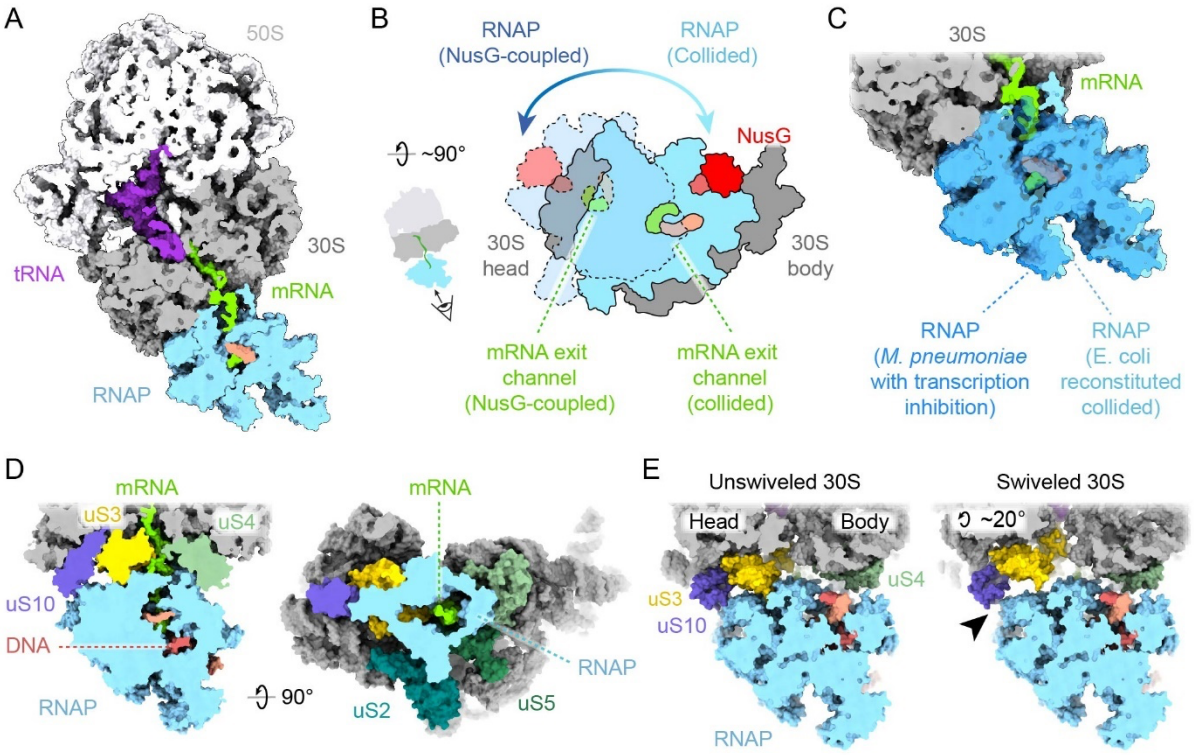

Figure 5: Structural details of the collided expressome. **(A)** Structure of the collided expressome. Slice through atomic model (PDB: 6ZTL) shown in surface representation. **(B)** Comparison of the position of RNAP (blue) bound to NusG (red) in the *E. coli* collided expressome (solid lines) and *E. coli* NusG-coupled expressome (dashed lines). The RNAP mRNA exit channel (green) is directly aligned with the mRNA entry channel of the 30S subunit in the collided expressome, which can no longer by coupled by NusG (red). **(C)** Equivalent relative positions of RNAP in the reconstituted *E. coli* expressome (pale blue; PDB: 6ZTL) and expressome identified following treatment of *M. pneumoniae* cells with the transcription inhibitor pseudouridimycin (medium blue; models fitted to EMDB: 10684). **(D)** Contact of RNAP and the ribosome in the collided expressome. RNAP inserts into the concave surface of the 30S subunit bordered by uS2, uS3, uS4, uS5 and uS10. Tension of the mRNA, which transits directly from RNAP to ribosome, drives the observed architecture. **(E)** Translocation of the ribosome involves rotation of the 30S head domain with respect to the 30S body domain by approximately 20°. A structural model of the ribosome in an intermediate state of translocation (PDB: 4V7B)[108] was aligned to the collided expressome (PDB: 6ZTL) to assess potential for steric clash with RNAP. The resulting model ('swiveled') shows increased separation between RNAP and ribosomal proteins uS3 and uS10 (arrow head) relative to the experimental collided expressome model ('unswiveled'). No structural overlap was observed. Molecular graphics and analyses performed with UCSF ChimeraX version 1.2.4[126].

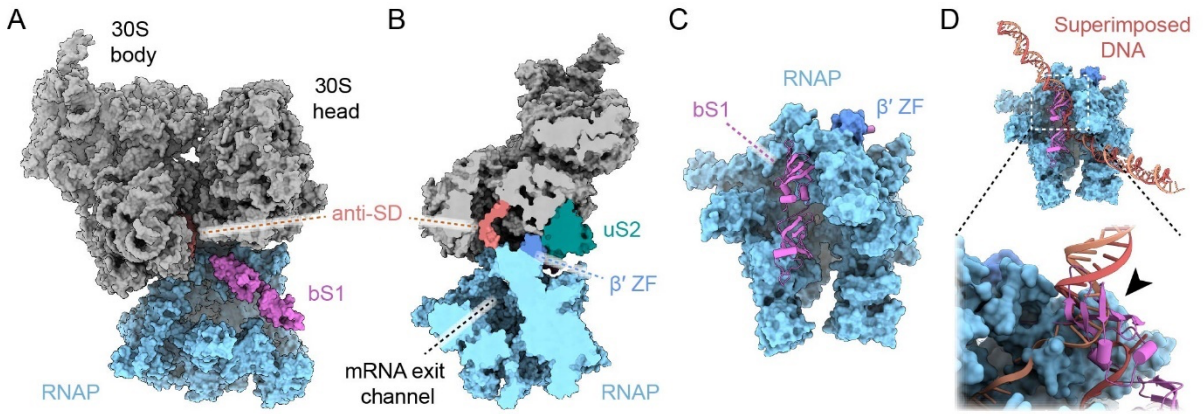

Figure 6: Structural details of RNAP:30S complex. **(A)** Atomic model of the RNAP:30S complex (PDB: 6AWB) shown in surface representation. Additional density assigned to two OB domains of ribosomal protein bS1 bound to RNAP shown (purple). **(B)** Slice through atomic model showing interaction between ribosomal protein uS2 and RNAP β′ ZF domain and alignment of the mRNA exit channel of RNAP with the anti-Shine-Dalgarno sequence of the 16S rRNA (anti-SD). **(C)** Interaction of bS1 with RNAP. **(D)** The RNAP:30S complex was determined in the absence of DNA-RNA bound to RNAP. Structural alignment of RNAP from the RNAP:30S complex with RNAP bound to DNA predicts a clash between bS1 and DNA (arrow head). Molecular graphics and analyses performed with UCSF ChimeraX version 1.2.4[126].

References

1. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 1961; 3:318–56.

2. Brenner S, Jacob F, Meselson M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 1961; 190:576–81.

3. Gros F, Hiatt H, Gilbert W, Kurland CG, Risebrough RW, Watson JD. Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. Nature 1961; 190:581–5.

4. BYRNE R, LEVIN JG, BLADEN HA, NIRENBERG MW. the in Vitro Formation of a Dna-Ribosome Complex. Proc Natl Acad Sci United States 1964; 52:140–8.

5. Das HK, Goldstein A, Lowney LI. Attachment of ribosomes to nascent messenger RNA in Escherichia coli. J Mol Biol 1967; 24:231–45.

6. Miller OL, Hamkalo BA, Thomas CA. Visualization of bacterial genes in action. Science (80-) 1970; 169:392–5.

7. Kohler R, Mooney RA, Mills DJ, Landick R, Cramer P. Architecture of a transcribing-translating expressome. Sci (New York, NY) 2017; 356:194–7.

8. Proshkin S, Rachid Rahmouni A, Mironov A, Nudler E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science (80-) 2010; 328:504–8.

9. Zhu M, Mori M, Hwa T, Dai X. Disruption of transcription–translation coordination in Escherichia coli leads to premature transcriptional termination. Nat Microbiol 2019; 4:2347–56.

10. Chen M, Fredrick K. RNA Polymerase's Relationship with the Ribosome: Not So Physical, Most of the Time. J Mol Biol 2020; 432:3981–6.

11. Dutta D, Shatalin K, Epshtein V, Gottesman ME, Nudler E. Linking RNA polymerase backtracking to genome instability in E. coli. Cell 2011; 146:533–43.

12. Turnbough CL. Regulation of Bacterial Gene Expression by Transcription Attenuation. Microbiol Mol Biol Rev 2019 [cited 2021 Apr 14]; 83.

13. Richardson JP. Preventing the synthesis of unused transcripts by rho factor. Cell 1991; 64:1047–9.

14. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rösch P. A NusE:NusG complex links transcription and translation. Science (80-) 2010; 328:501–4.

15. Saxena S, Myka KK, Washburn R, Costantino N, Court DL, Gottesman ME. Escherichia coli transcription factor NusG binds to 70S ribosomes. Mol Microbiol 2018; 108:495–504.

16. Webster MW, Takacs M, Zhu C, Vidmar V, Eduljee A, Abdelkareem M, Weixlbaumer A.

Structural basis of transcription-translation coupling and collision in bacteria. Science (80-) 2020; 369:1355–9.

17. Wang C, Molodtsov V, Firlar E, Kaelber JT, Blaha G, Su M, Ebright RH. Structural basis of transcription-translation coupling. Science (80-) 2020; 369:1359–65.

18. O'Reilly FJ, Xue L, Graziadei A, Sinn L, Lenz S, Tegunov D, Blötz C, Singh N, Hagen WJH, Cramer P, et al. In-cell architecture of an actively transcribing-translating expressome. Science (80-) 2020; 369:554–7.

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

20. Demo G, Rasouly A, Vasilyev N, Svetlov V, Loveland AB, Diaz-Avalos R, Grigorieff N, Nudler E, Korostelev AA. Structure of RNA polymerase bound to ribosomal 30S subunit. Elife 2017; 6:94.

21. Chatterjee S, Chauvier A, Dandpat SS, Artsimovitch I, Walter NG. A translational riboswitch coordinates nascent transcription-translation coupling. Proc Natl Acad Sci U S A 2021; 118.

22. Irastortza-Olaziregi M, Amster-Choder O. Coupled Transcription-Translation in Prokaryotes: An Old Couple With New Surprises. Front Microbiol 2021; 11:3532.

23. Weixlbaumer A, Grünberger F, Werner F, Grohmann D. Coupling of Transcription and Translation in Archaea: Cues From the Bacterial World. Front Microbiol 2021; 12:935.

24. Wang B, Artsimovitch I. A Growing Gap between the RNAP and the Lead Ribosome. Trends Microbiol 2021; 29:4–5.

25. Jacquet M, Kepes A. Initiation, elongation and inactivation of lac messenger RNA in Escherichia coli studied by measurement of its β-galactosidase synthesizing capacity in vivo. J Mol Biol 1971; 60:453–72.

26. Lazzarini RA, Dahlberg AE. The control of ribonucleic acid synthesis during amino acid deprivation in Escherichia coli. J Biol Chem 1971; 246:420–9.

27. Imamoto F. Diversity of regulation of genetic transcription. I. Effect of antibiotics which inhibit the process of translation on RNA metabolism in Escherichia coli. J Mol Biol 1973; 74:113–36.

28. Potrykus K, Murphy H, Philippe N, Cashel M. ppGpp is the major source of growth rate control in E. coli. Environ Microbiol 2011; 13:563–75.

29. Gourse RL, Chen AY, Gopalkrishnan S, Sanchez-Vazquez P, Myers A, Ross W. Transcriptional Responses to ppGpp and DksA. Annu. Rev. Microbiol.2018; 72:163–84.

30. Irving SE, Choudhury NR, Corrigan RM. The stringent response and physiological roles of (pp)pGpp in bacteria. Nat Rev Microbiol 2021; 19:256–71.

31. Burgos HL, O'Connor K, Sanchez-Vazquez P, Gourse RL. Roles of transcriptional and translational control mechanisms in regulation of ribosomal protein synthesis in Escherichia coli. J Bacteriol 2017; 199.

32. Diez S, Ryu J, Caban K, Gonzalez RL, Dworkin J. The alarmones (p)ppGpp directly regulate translation initiation during entry into quiescence. Proc Natl Acad Sci U S A 2020; 117:15565–72.

33. Molodtsov V, Sineva E, Zhang L, Huang X, Cashel M, Ades SE, Murakami KS. Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. Mol Cell 2018; 69:828-839.e5.

34. Sanchez-Vazquez P, Dewey CN, Kitten N, Ross W, Gourse RL. Genome-wide effects on Escherichia coli transcription from ppGpp binding to its two sites on RNA polymerase. Proc Natl Acad Sci U S A 2019; 116:8310–9.

35. Kingston RE, Nierman WC, Chamberlin MJ. A direct effect of guanosine tetraphosphate on pausing of Escherichia coli RNA polymerase during RNA chain elongation. J Biol Chem 1981; 256:2787–97.

36. Vogel U, Jensen KF. Effects of guanosine 3',5'-bisdiphosphate (ppGpp) on rate of transcription elongation in isoleucine-starved Escherichia coli. J Biol Chem 1994; 269:16236–41.

37. Furman R, Sevostyanova A, Artsimovitch I. Transcription initiation factor DksA has diverse effects on RNA chain elongation. Nucleic Acids Res 2012; 40:3392–402.

38. Krásný L, Gourse RL. An alternative strategy for bacterial ribosome synthesis: Bacillus subtilis rRNA transcription regulation. EMBO J 2004; 23:4473–83.

39. Landick R, Carey J, Yanofsky C. Translation activates the paused transcription complex and restores transcription of the trp operon leader region. Proc Natl Acad Sci U S A 1985; 82:4663–7.

40. Chan CL, Landick R. Dissection of the his leader pause site by base substitution reveals a multipartite signal that includes a pause RNA hairpin. J Mol Biol 1993; 233:25–42.

41. Kang JY, Mishanina T V., Bellecourt MJ, Mooney RA, Darst SA, Landick R. RNA Polymerase Accommodates a Pause RNA Hairpin by Global Conformational Rearrangements that Prolong Pausing. Mol Cell 2018; 69:802-815.e1.

42. Guo X, Myasnikov AG, Chen J, Crucifix C, Papai G, Takacs M, Schultz P, Weixlbaumer A. Structural Basis for NusA Stabilized Transcriptional Pausing. Mol Cell 2018; 69:816-827.e4.

43. Nudler E, Mustaev A, Lukhtanov E, Goldfarb A. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell 1997; 89:33–41.

44. Komissarova N, Kashlev M. Transcriptional arrest: Escherichia coli RNA polymerase

translocates backward, leaving the 3′ end of the RNA intact and extruded. Proc Natl Acad Sci U S A 1997; 94:1755–60.

45. Larson MH, Mooney RA, Peters JM, Windgassen T, Nayak D, Gross CA, Block SM, Greenleaf WJ, Landick R, Weissman JS. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. Science (80-) 2014; 344:1042–7.

46. Vvedenskaya IO, Vahedian-Movahed H, Bird JG, Knoblauch JG, Goldman SR, Zhang Y, Ebright RH, Nickels BE. Interactions between RNA polymerase and the “core recognition element” counteract pausing. Science (80-) 2014; 344:1285–9.

47. Yanofsky C. Attenuation in the control of expression of bacterial operons. Nature 1981; 289:751–8.

48. Yanofsky C. Transcription attenuation: Once viewed as a novel regulatory strategy. J Bacteriol 2000; 182:1–8.

49. Takyar S, Hickerson RP, Noller HF. mRNA helicase activity of the ribosome. Cell 2005; 120:49–58.

50. Qu X, Wen J Der, Lancaster L, Noller HF, Bustamante C, Tinoco I. The ribosome uses two active mechanisms to unwind messenger RNA during translation. Nature 2011; 475:118–21.

51. Borukhov S, Sagitov V, Goldfarb A. Transcript cleavage factors from E. coli. Cell 1993; 72:459–66.

52. Erie DA, Hajiseyedjavadi O, Young MC, Von Hippel PH. Multiple RNA polymerase conformations and GreA: Control of the fidelity of transcription. Science (80-) 1993; 262:867–73.

53. Epshtein V, Toulmé F, Rachid Rahmouni A, Borukhov S, Nudler E. Transcription through the roadblocks: The role of RNA polymerase cooperation. EMBO J 2003; 22:4719–27.

54. Hao N, Krishna S, Ahlgren-Berg A, Cutts EE, Shearwin KE, Dodd IB. Road rules for traffic on DNA - Systematic analysis of transcriptional roadblocking in vivo. Nucleic Acids Res 2014; 42:8861–72.

55. Stevenson-Jones F, Woodgate J, Castro-Roa D, Zenkin N. Ribosome reactivates transcription by physically pushing RNA polymerase out of transcription arrest. Proc Natl Acad Sci U S A 2020; 117:8462–7.

56. Brenner S, Stretton AOW, Kaplan S. Genetic code: The “nonsense” triplets for chain termination and their suppression. Nature 1965; 206:994–8.

57. Franklin NC, Luria SE. Transduction by bacteriophage P1 and the properties of the lac genetic region in E. coli and S. dysenteriae. Virology 1961; 15:299–311.

58. Newton WA, Beckwith JR, Zipser D, Brenner S. Nonsense mutants and polarity in the Lac

operon of Escherichia coli. J Mol Biol 1965; 14:290–6.

59. de Crombrugghe B, Adhya S, Gottesman M, Pastan I. Effect of rho on transcription of bacterial operons. Nat. New Biol.1973; 241:260–4.

60. Adhya S, Gottesman M. Control of transcription termination. Annu Rev Biochem 1978; 47:967–96.

61. Ruteshouser EC, Richardson JP. Identification and characterization of transcription termination sites in the Escherichia coli lacZ gene. J Mol Biol 1989; 208:23–43.

62. Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. Rho and NusG suppress pervasive antisense transcription in Escherichia coli. Genes Dev 2012; 26:2621–33.

63. Roberts JW. Mechanisms of Bacterial Transcription Termination. J Mol Biol 2019; 431:4030–9.

64. Chen CY, Richardson JP. Sequence elements essential for rho-dependent transcription termination at lambda tR1. J Biol Chem 1987; 262:11292–9.

65. Peters JM, Vangeloff AD, Landick R. Bacterial transcription terminators: The RNA 3′-end chronicles. J Mol Biol 2011; 412:793–813.

66. Wright JJ, Hayward RS. Transcriptional termination at a fully rho-independent site in Escherichia coli is prevented by uninterrupted translation of the nascent RNA. EMBO J 1987; 6:1115–9.

67. Roland KL, Liu C, Turnbough CL. Role of the ribosome in suppressing transcriptional termination at the pyrBI attenuator of Escherichia coli K-12. Proc Natl Acad Sci U S A 1988; 85:7149–53.

68. Li R, Zhang Q, Li J, Shi H. Effects of cooperation between translating ribosome and RNA polymerase on termination efficiency of the Rho-independent terminator. Nucleic Acids Res 2015; 44:2554–63.

69. Wang X, Monford Paul Abishek N, Jeon HJ, Lee Y, He J, Adhya S, Lim HM. Processing generates 3′ ends of RNA masking transcription termination events in prokaryotes. Proc Natl Acad Sci U S A 2019; 116:4440–5.

70. Lesnik EA, Sampath R, Levene HB, Henderson TJ, McNeil JA, Ecker DJ. Prediction of rho-independent transcriptional terminators in Escherichia coli. Nucleic Acids Res 2001; 29:3583–94.

71. Lawson MR, Ma W, Bellecourt MJ, Artsimovitch I, Martin A, Landick R, Schulten K, Berger JM. Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. Mol Cell 2018; 71:911-922.e4.

72. Bailey EJ, Gottesman ME, Gonzalez RL. NusG links transcription and translation in

Escherichia coli extracts. bioRxiv 2021; 48109:2021.07.31.454578.

73. Elgamal S, Artsimovitch I, Ibba M. Maintenance of transcription-translation coupling by elongation factor P. MBio 2016; 7.

74. Mooney RA, Schweimer K, Rösch P, Gottesman M, Landick R. Two Structurally Independent Domains of E. coli NusG Create Regulatory Plasticity via Distinct Interactions with RNA Polymerase and Regulators. J Mol Biol 2009; 391:341–58.

75. Kang JY, Mooney RA, Nedialkov Y, Saba J, Mishanina T V., Artsimovitch I, Landick R, Darst SA. Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. Cell 2018; 173:1650–62.

76. Kyrpides NC, Woese CR, Ouzounis CA. KOW: A novel motif linking a bacterial transcription factor with ribosomal proteins. Trends Biochem Sci 1996; 21:425–6.

77. Condon C, Squires C, Squires CL. Control of rRNA transcription in Escherichia coli . Microbiol. Rev.1995; 59:623–45.

78. Das A, Ghosh B, Barik S, Wolska K. Evidence that ribosomal protein S10 itself is a cellular component necessary for transcription antitermination by phage λ N protein. Proc Natl Acad Sci U S A 1985; 82:4070–4.

79. Krupp F, Said N, Huang YH, Loll B, Bürger J, Mielke T, Spahn CMT, Wahl MC. Structural Basis for the Action of an All-Purpose Transcription Anti-termination Factor. Mol Cell 2019; 74:143-157.e5.

80. Drögemüller J, Strauß M, Schweimer K, Jurk M, Rösch P, Knauer SH. Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. Sci Rep 2015; 5:16428.

81. Sukhodolets M V., Garges S. Interaction of Escherichia coli RNA polymerase with the ribosomal protein S1 and the Sm-like ATPase Hfq. Biochemistry 2003; 42:8022–34.

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

83. Rippa V, Cirulli C, Di Palo B, Doti N, Amoresano A, Duilio A. The ribosomal protein L2 interacts with the RNA polymerase α subunit and acts as a transcription modulator in Escherichia coli. J Bacteriol 2010; 192:1882–9.

84. Bhavsar RB, Makley LN, Tsonis PA. The other lives of ribosomal proteins. Hum Genomics 2010; 4:327–44.

85. Washburn RS, Zuber PK, Sun M, Hashem Y, Shen B, Li W, Harvey S, Acosta Reyes FJ, Gottesman ME, Knauer SH, et al. Escherichia coli NusG Links the Lead Ribosome with the

Transcription Elongation Complex. iScience 2020; 23:101352.

86. Mitra P, Ghosh G, Hafeezunnisa M, Sen R. Rho Protein: Roles and Mechanisms. Annu Rev Microbiol 2017; 71:687–709.

87. Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. Science (80-) 2008; 320:935–8.

88. Burns CM, Richardson JP. NusG is required to overcome a kinetic limitation to Rho function at an intragenic terminator. Proc Natl Acad Sci U S A 1995; 92:4738–42.

89. Valabhoju V, Agrawal S, Sen R. Molecular basis of NusG-mediated regulation of Rho-dependent transcription termination in bacteria. J Biol Chem 2016; 291:22386–403.

90. Said N, Hilal T, Sunday ND, Khatri A, Bürger J, Mielke T, Belogurov GA, Loll B, Sen R, Artsimovitch I, et al. Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase p. Science (80-) 2021; 371:eabd1673.

91. Hao Z, Epshtein V, Kim KH, Proshkin S, Svetlov V, Kamarthapu V, Bharati B, Mironov A, Walz T, Nudler E. Pre-termination Transcription Complex: Structure and Function. Mol Cell 2021; 81:281-292.e8.

92. Turtola M, Belogurov GA. NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. Elife 2016; 5.

93. Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, Landick R. Regulator Trafficking on Bacterial Transcription Units In Vivo. Mol Cell 2009; 33:97–108.

94. Mah TF, Li J, Davidson AR, Greenblatt J. Functional importance of regions in Escherichia coli elongation factor NusA that interact with RNA polymerase, the bacteriophage λ N protein and RNA. Mol Microbiol 1999; 34:523–37.

95. Ha KS, Toulokhonov I, Vassylyev DG, Landick R. The NusA N-Terminal Domain Is Necessary and Sufficient for Enhancement of Transcriptional Pausing via Interaction with the RNA Exit Channel of RNA Polymerase. J Mol Biol 2010; 401:708–25.

96. Schweimer K, Prasch S, Sujatha PS, Bubunenko M, Gottesman ME, Rösch P. NusA interaction with the α subunit of E. coli RNA polymerase is via the UP element site and releases autoinhibition. Structure 2011; 19:945–54.

97. Ingham CJ, Dennis J, Furneaux PA. Autogenous regulation of transcription termination factor Rho and the requirement for Nus factors in Bacillus subtilis. Mol Microbiol 1999; 31:651–63.

98. Shibata R, Bessho Y, Shinkai A, Nishimoto M, Fusatomi E, Terada T, Shirouzu M, Yokoyama S. Crystal structure and RNA-binding analysis of the archaeal transcription factor NusA. Biochem Biophys Res Commun 2007; 355:122–8.

99. Mah TF, Kuznedelov K, Mushegian A, Severinov K, Greenblatt J. The α subunit of E. coli RNA polymerase activates RNA binding by NusA. Genes Dev 2000; 14:2664–75.

100. Downing WL, Sullivan SL, Gottesman ME, Dennis PP. Sequence and transcriptional pattern of the essential Escherichia coli secE-nusG operon. J Bacteriol 1990; 172:1621–7.

101. Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, Lund PA, Cole JA, Henderson IR. The essential genome of Escherichia coli K-12. MBio 2018; 9.

102. Nakamura Y, Uchida H. Isolation of conditionally lethal amber mutations affecting synthesis of the nusA protein of Escherichia coli. Mol Gen Genet 1983; 190:196–203.

103. Xia M, Lunsford RD, McDevitt D, Iordanescu S. Rapid method for the identification of essential genes in Staphylococcus aureus. Plasmid 1999; 42:144–9.

104. Tsugawa A, Saito M, Court DL, Nakamura Y. nusA amber mutation that causes temperature-sensitive growth of Escherichia coli. J Bacteriol 1988; 170:908–15.

105. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;

106. Johnson GE, Lalanne JB, Peters ML, Li GW. Functionally uncoupled transcription–translation in Bacillus subtilis. Nature 2020 [cited 2020 Aug 28]; 585:124–8.

107. Ratje AH, Loerke J, Mikolajka A, Brünner M, Hildebrand PW, Starosta AL, Dönhöfer A, Connell SR, Fucini P, Mielke T, et al. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature 2010; 468:713–6.

108. Ramrath DJF, Lancaster L, Sprink T, Mielke T, Loerke J, Noller HF, Spahn CMT. Visualization of two transfer RNAs trapped in transit during elongation factor G-mediated translocation. Proc Natl Acad Sci U S A 2013; 110:20964–9.

109. Zhou J, Lancaster L, Donohue JP, Noller HF. How the ribosome hands the A-site tRNA to the P site during EF-G-catalyzed translocation. Science (80-) 2014; 345:1188–91.

110. Artsimovitch I, Landick R. The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. Cell 2002; 109:193–203.

111. Svetlov V, Belogurov GA, Shabrova E, Vassylyev DG, Artsimovitch I. Allosteric control of the RNA polymerase by the elongation factor RfaH. Nucleic Acids Res 2007; 35:5694–705.

112. Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I. Functional specialization of transcription elongation factors. EMBO J 2009; 28:112–22.

113. Bailey MJA, Hughes C, Koronakis V. RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. Mol Microbiol 1997; 26:845–51.

114. Fan H, Conn AB, Williams PB, Diggs S, Hahm J, Gamper HB, Hou YM, O'Leary SE, Wang Y, Blaha GM. Transcription-Translation coupling: Direct interactions of RNA polymerase with ribosomes and ribosomal subunits. Nucleic Acids Res 2017; 45:11043–55.

115. Byrgazov K, Grishkovskaya I, Arenz S, Coudevylle N, Temmel H, Wilson DN, Djinovic-Carugo K, Moll I. Structural basis for the interaction of protein S1 with the Escherichia coli ribosome. Nucleic Acids Res 2015; 43:661–73.

116. Loveland AB, Korostelev AA. Structural dynamics of protein S1 on the 70S ribosome visualized by ensemble cryo-EM. Methods 2018; 137:55–66.

117. Sukhodolets M V., Garges S, Adhya S. Ribosomal protein S1 promotes transcriptional cycling. Rna 2006 [cited 2021 Jul 4]; 12:1505–13.

118. Zuber PK, Schweimer K, Rösch P, Artsimovitch I, Knauer SH. Reversible fold-switching controls the functional cycle of the antitermination factor RfaH. Nat Commun 2019; 10.

119. Castro-Roa D, Zenkin N. In vitro experimental system for analysis of transcription-translation coupling. Nucleic Acids Res 2012; 40:e45–e45.

120. Bakshi S, Siryaporn A, Goulian M, Weisshaar JC. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. Mol Microbiol 2012; 85:21–38.

121. Bakshi S, Choi H, Weisshaar JC. The spatial biology of transcription and translation in rapidly growing Escherichia coli. Front Microbiol 2015; 6:636.

122. Chen M, Fredrick K. Measures of single-versus multiple-round translation argue against a mechanism to ensure coupling of transcription and translation. Proc Natl Acad Sci U S A 2018; 115:10774–9.

123. French SL, Santangelo TJ, Beyer AL, Reeve JN. Transcription and translation are coupled in Archaea. Mol Biol Evol 2007; 24:893–5.

124. Winkler ME, Yanofsky C. Pausing of RNA Polymerase during in Vitro Transcription of the Tryptophan Operon Leader Region. Biochemistry 1981; 20:3738–44.

125. Cohen RD, Pielak GJ. A cell is more than the sum of its (dilute) parts: A brief history of quinary structure. Protein Sci 2017; 26:403–13.

126. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci 2021; 30:70–82.

127. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011; 7.

128. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator.

Genome Res 2004; 14:1188–90.