Published in final edited form as:

Chem Rev. 2013 November 13; 113(11): 8604–8619. doi:10.1021/cr400064k.

NusG-Spt5 proteins - universal tools for transcription modification and communication

Sushil Kumar Tomar and Irina Artsimovitch

Department of Microbiology and The Center for RNA Biology, The Ohio State University, Columbus, OH, USA

1. INTRODUCTION

DNA-dependent RNA polymerases (RNAPs) carry out synthesis of RNAs in all domains of life. Each round of transcription can be divided into three major steps¹. Initiation encompasses RNAP recruitment to double-stranded promoter DNA. This leads to DNA strand separation to form a transcription bubble in which a region around the start site (denoted as +1) is melted. After reiterative synthesis and release of short abortive RNAs, RNAP escapes from the promoter. During elongation, RNAP adds one nucleotide (nt) to the nascent RNA chain and moves forward along the DNA by one base, repeating this cycle for thousands of times while remaining bound to both the DNA template and the growing RNA chain in a stable ternary transcription elongation complex (TEC; Figure 1). At a terminator, the TEC becomes massively destabilized by nucleic acid signals or proteins, which triggers the release of the RNA message. The transcription cycle is completed when RNAP dissociates from the DNA, after which it can rebind at another promoter.

Some RNAPs are composed of a single polypeptide chain that can carry out the entire cycle either in the absence (as is the case in bacteriophage T7² and N4³ RNAPs) or with the help of a single promoter-specificity factor (e.g., in the case of the yeast mitochondrial RNAP⁴). However, all cellular genomes are transcribed by multisubunit enzymes that differ in their size and complexity (bacterial RNAPs commonly have five subunits, whereas eukaryotic and archaeal enzymes are composed of 12–17 polypeptide chains) but that have evolved from a common ancestor and share overall architecture and active-site organization, types of interactions with the nucleic acid chains, and molecular mechanism of catalysis^{5–6}. They also share requirements for regulatory proteins that define the boundaries of the transcription units, enable processive RNA synthesis, and rescue arrested transcription complexes⁷.

Catalytically-competent multisubunit core RNAPs can transcribe DNA templates faithfully and efficiently but do not recognize specific nucleic acid sequences with high affinity; instead, they rely on transiently bound sequence-specific factors to initiate (and sometimes terminate) transcription at defined positions. In *Bacteria*, a family of σ factors recognizing different consensus DNA elements direct core enzymes to specific sets of promoters and facilitate DNA strand separation^{8–9}. Although bacterial σ factors are not homologous to

initiation factors utilized by more complex RNAPs, structural studies reveal some similarities in the details of their molecular interactions^{10–11}.

To begin processive RNA chain elongation, RNAP must break the specific interactions with the promoter DNA, which would otherwise prevent the transition from initiation to elongation¹². Since these interactions are mediated by the σ factor, the simplest mechanism entails abrupt dissociation of σ from the core triggered by a clash with the growing nascent RNA¹³. However, σ release is not an obligatory step during promoter escape: while structural studies are consistent with such a clash, they also reveal that σ /core contacts could be broken sequentially (reviewed in¹⁴), leading to slow and stochastic release of σ from the elongating RNAP¹⁵. Among many σ /core contacts, which encompass 10^4 Å of the surface area, interactions between σ region 2 and the clamp helices domain of the β' subunit (β' CH) contribute the most to σ /core affinity and are structurally compatible with both the initiation and the elongation complexes. During initiation, interactions of σ region 2 with the –10 promoter element mediate DNA melting^{11,16}. During elongation, σ can bind to the β' CH and the –10 sequence exposed in the nontemplate DNA on the TEC surface, inducing strong RNAP pausing far away from a promoter¹⁷. Similarly, translation initiation signals stall the ribosome within coding sequences¹⁸. Thus, the absence of strong interactions with the template ensures unimpeded progression during chain synthesis. The lack of conservation among initiation factors may suggest that in the last universal common ancestor a linear genome was transcribed end-to-end. In addition to targeted initiation, sequence-specific initiation factors may also suppress spurious initiation from the end of a linear template¹⁹.

The mechanistic requirements for termination are simpler, and RNA release can be triggered by nucleic acid signals such as an RNA hairpin followed by a run of U residues (in *Escherichia coli* and *Bacillus subtilis*²⁰) or by a U-track alone (in *Methanothermobacter thermautotrophicus*²¹); yet accessory factors that have a helicase or a translocase activity can be required for termination at other sites. In *Bacteria*, these classes of factors are represented by Rho, which limits expression of deleterious foreign DNA²² and suppresses antisense transcription²³, and Mfd, a transcription-repair coupling factor which recruits nucleotide-excision repair machinery to RNAPs stalled at sites of DNA lesions²⁴. Unrelated proteins that play analogous roles have been characterized in other systems^{25–26}.

Surprisingly, despite the remarkable similarities in the structure, mechanism, and many regulatory challenges faced by all multisubunit RNAPs, the initiation, termination, and most of other types of accessory factors utilized by these enzymes have no common evolutionary origin⁷. The only exception is a class of NusG-like proteins (called Spt5 in *Archaea* and yeast and DSIF in humans) whose sequences, structures, binding sites on RNAP and mechanism of RNAP modification are universally conserved. These regulators are commonly viewed as transcription elongation factors^{27–28}, but they also regulate termination^{22–23,29} and, as suggested by recent studies in *Archaea*, may modulate initiation³⁰. Another key role of these proteins is to tether the elongating RNAP to other macromolecular complexes (e.g., the leading ribosome^{31–32} in *Bacteria* and *Archaea* vs mRNA capping enzyme³³ in eukaryotes), thereby enabling coordinated regulation of transcription and coupled processes.

This review is focused on regulation of gene expression by bacterial NusG and its paralogs, among which RfaH is the best characterized. Studies of RfaH and NusG identified the binding sites on the TEC^{34–35} and the molecular mechanism of processivity²⁹ that are common to all members of their class^{28,36}. Analysis of the cellular targets and comparison of the regulatory properties of RfaH and NusG^{37–38} revealed opposite effects on gene expression: while NusG, together with Rho, silences foreign DNA²², RfaH strongly activates a subset of horizontally-transferred genes^{29,39}. Most recently, demonstration that the C-terminal domain (CTD) of RfaH undergoes a complete fold conversion led to a discovery of a class of transformer proteins⁴⁰, in which a dramatic change in protein structure³¹ defines a new function. We will briefly describe the TEC structure and mechanisms of pausing and termination as these are essential for understanding how NusG works; see recent reviews for in-depth coverage of these topics^{7,41–43}. We will then focus on structural and regulatory properties of *E. coli* RfaH and NusG, drawing comparison to their homologs in other organisms to illustrate functional diversity within this family.

2. TRANSCRIPTION ELONGATION

2.1 TEC structure

All multisubunit RNAPs must synthesize RNAs with high fidelity and processivity while participating in regulatory interactions with diverse auxiliary factors; these interactions control initiation, elongation, and termination of RNA synthesis and determine the gene expression program of each cell. The highly conserved structures of the TECs reflect similar challenges faced by RNAPs during elongation: the TEC must be highly stable and processive at most template positions, transcribing through various obstacles (such as DNA-bound proteins), yet should halt RNA synthesis at specific regulatory pause signals and readily fall apart at terminators. Recent structural studies of the TEC revealed an atomic-resolution map of RNAP contacts to the nucleic acid chains and offered insights into the molecular basis of TEC dynamics⁴³. In the TEC, ~15 base-pairs of the DNA duplex are bound in the downstream DNA-binding channel (Figure 1). The duplex is melted just ahead of the active site (at +2) to form an ~14-base-pair transcription bubble, in which the non-template DNA strand is partially exposed on surface of the enzyme and the template DNA strand lies within the active-site cleft, with the acceptor DNA template (+1) available for base pairing with the incoming NTP substrate. The RNAP active site is accessible through a secondary channel, also called the substrate entry pore.

2.2 RNAP clamp dynamics

The RNAP structure resembles a crab claw⁴⁴ in which two pincers encircle a DNA-binding cleft with the RNAP active center positioned at its base. The clamp domain of the β' subunit forms one pincer and the β lobe forms the other. In various crystal structures, the clamp has been observed in open and closed conformational states^{11,45–47} arising as a result of an ~20° rotation around the switch region located at the base of the clamp. Conformational dynamics of the β' clamp have been proposed to play key roles throughout the transcription cycle. The clamp may open to permit loading of DNA duplex into the RNAP active-site cleft during initiation⁹, close to establish the tight grip on the template DNA during elongation⁴⁸, and then open again to allow release of DNA during termination⁴⁹.

Recent single-molecule FRET analysis⁵⁰ demonstrated that the clamp adopts different states in solution: the clamp is predominantly open in free RNAP and unstable closed promoter complexes but is closed in catalytically-competent open promoter complexes and TECs. These results supported the long-held view that clamp closure is required for the high stability of transcription complexes and that accessory factors that maintain the clamp in the closed state would enable synthesis of long RNAs⁵¹. All RNAPs are obligatorily processive and cannot rebind a prematurely released nascent transcript, which can be tens of thousands of nucleotides long. Thus, regulation of the clamp dynamics is essential for productive transcription.

2.3 Discontinuous RNA chain elongation

A complete cycle of nucleotide addition consists of NTP binding, catalysis, pyrophosphate (PP_i) release, and translocation. RNAP repeats this cycle many thousands of times to complete synthesis of the nascent RNA chain, while remaining bound to both the DNA template and the growing transcript. The cycle begins with NTP binding to the *i+1* sub-site (also referred to as the insertion site or substrate site) of the post-translocated TEC, in which the 3' end of the RNA is positioned in the *i* (or product) sub-site (Figure 1). Catalysis occurs by Mg²⁺ dependent, S_N2-type nucleophilic attack of the RNA 3' hydroxyl on the NTP α-phosphorus atom, displacing PP_i. Catalysis involves stabilization of a trigonal-bipyramidal transition state⁵² by two Mg²⁺ ions, a high-affinity Mg²⁺I that resides in the active site and Mg²⁺II that is delivered bound to the substrate NTP for each round of catalysis. After catalysis, the 3' end is positioned in the pre-translocated state; forward translocation driven by thermal motions⁵³ gives rise to the post-translocated state, and the cycle is repeated.

While *in vitro* *E. coli* RNAP can extend the nascent RNA at 500 nt/sec⁵⁴, *in vivo* it moves only at 20–90 nt/sec⁵⁵ because RNA synthesis is hindered by various obstacles, including DNA-bound proteins, DNA lesions, and intrinsic signals encoded in the DNA template and the nascent RNA¹. Even at saturating NTP concentrations on naked intact templates *in vitro*, RNAP moves in leaps, with its fast movement along the template punctuated by short-lived pauses⁵⁶ that trigger TEC isomerization into an unactivated state. This so-called 'elemental' pause intermediate⁵⁷ can slowly escape to the elongation pathway upon nucleotide addition, isomerize into long-lived paused states upon formation of a nascent secondary structure (a pause RNA hairpin⁵⁸) or lateral sliding (backtracking⁵⁹), or give rise to a termination complex (Figure 2). Formation of the elemental pause is accompanied by structural rearrangements near the active site that may include opening of the clamp^{58,60}. Factors that prevent opening of the clamp would block formation of the elemental pause and favor rapid elongation⁵¹.

Pausing plays numerous regulatory roles, is an obligatory step in termination pathways, and likely controls the overall rate of RNA chain elongation⁴¹. A relatively slow rate may be necessary for timely recruitment of regulatory factors, attenuation control, co-transcriptional folding of the nascent RNA, and efficient coupling of transcription and translation in *Bacteria* and *Archaea* (see⁴¹ for review). Most pauses are not mediated by specific interactions between the RNAP and the nucleic acids. However, core RNAP can recognize certain promoter elements during initiation⁶¹ and pausing at some sites could be explained

by specific RNAP/DNA interactions^{62–63}. Pausing may also be induced by proteins that recognize specific sequences. Factors that bind to double-stranded DNA in the path of elongating RNAP would hinder its progression, acting as roadblocks. Other factors, such as *E. coli* RfaH and σ^{70} , may specifically interact with the bases in the non-template DNA strand exposed on the TEC surface to induce RNAP pausing^{17,64}; interestingly, RfaH and σ^{70} bind to the same site on RNAP, the β' CH domain, despite the absence of any homology⁶⁵.

2.4 Termination

To enable differential regulation of adjacent transcription units, RNA chain synthesis has to be stopped at defined sites called terminators. This is particularly important in *Bacteria* and *Archaea* which have compact genomes with short intergenic sequences; indeed, RNA release frequently occurs at just one or two nucleotide positions. How does the TEC, which can add tens of thousands of nucleotides without fault and has a half-life of days *in vitro*⁶⁶, become abruptly destabilized?

The decision between elongation and termination pathways is kinetically controlled⁶⁷ and a dramatic destabilization is required to bring the dwell time of the TEC ($>10^5$ sec) into the characteristic range for nucleotide addition ($<10^{-1}$ sec); at most template positions the barrier to termination exceeds the one to elongation by ~ 18 kcal/mol⁶⁸. In the TEC, the energetically costly maintenance of the ~ 17 nt bubble is balanced by two favorable contributions, the formation of the RNA:DNA hybrid, the major stability determinant⁶⁹, and RNAP interactions with the nascent RNA and with the downstream DNA⁴².

Among prokaryotic systems, the mechanism of termination is best studied in *E. coli*, where termination signals can be grouped into two classes, intrinsic and factor-dependent (Figure 2; 42). At intrinsic sites, an RNA signal composed of a GC-rich RNA hairpin followed by a run of U residues triggers TEC dissociation in the absence of any accessory protein. Intrinsic terminators located at ends of operons serve as punctuation signals, whereas those in untranslated leader regions control expression of downstream genes by attenuation²⁰. These signals can be easily identified by computational approaches and give rise to $\sim 80\%$ of RNA ends in *E. coli*⁴². By contrast, RNA release at Rho-dependent sites is dependent on the action of Rho, a hexameric ATP-dependent translocase⁷⁰. These signals are composed of a 30+ nt pyrimidine-rich RNA element that serves as a Rho-loading site, *rut*, and a site of RNA release, which can be located hundreds of nucleotides downstream and frequently corresponds to a site at which RNAP pauses in the absence of Rho⁵⁵. This type of mechanism generates 3' ends of the remaining 20% of messages which include both mRNAs and stable small RNAs⁷¹. In addition, Rho carries out several essential quality-control tasks. It mediates polarity by interrupting transcription of damaged messages that can't be translated⁷² and suppresses the expression of horizontally acquired genes²². Rho may help to resolve R-loops⁷³ and traffic jams created by slow, unmodified RNAPs in heavily transcribed *rrn* operons⁷⁴. Most importantly, Rho blocks anti-sense transcription,²³ which is likely to be its essential function in *E. coli*.

Both types of termination signals likely disrupt the same interactions that hold the TEC together. Three models of RNA release have been put forth⁴². In the hybrid shearing

model ⁶⁹, the upstream portion of the RNA:DNA hybrid is disrupted by formation of the nascent hairpin or the helicase activity of Rho. In the allosteric model ^{75–76}, changes in the active site cleft induced by RNAP contacts to the hairpin or Rho weaken interactions with the RNA:DNA hybrid, and perhaps other nucleic acid chains. In the hypertranslocation model ⁷⁷, RNAP is pushed forward in the absence of nucleotide addition, thereby enlarging the bubble.

2.5 Antitermination

Termination efficiency, and thus the expression of downstream genes, can be modulated by accessory proteins, RNAs, and small molecules ⁷⁸. RNAP can ignore one, or many, termination signals in response to an antiterminator. In the first case, the action of the termination signal is compromised, *e.g.*, by preventing the formation of a terminator hairpin ⁷⁹, but the TEC properties are unaltered. In the second case, RNAP is modified into a processive antitermination state by a bound protein or RNA, and can ignore pause signals and many consecutive terminators located a few kilobases from the modification site ⁸⁰.

Processive broad-specificity antitermination modification is commonly employed by bacteriophages; three mechanisms, utilized by the phage λ N⁸¹ and Q proteins and HK202 phage nascent *put* RNA structure, have been characterized in molecular detail. These regulators allow RNAP to ignore all classes of pause and termination signals, allowing efficient expression of long phage operons during the lytic cycle. However, such run-away transcript elongation would render independent regulation of neighboring operons, insulated by hairpin-dependent terminators, impossible. Thus, one would expect that cellular antiterminators would increase RNAP processivity by reducing pausing and, consequently, Rho-mediated RNA release, but would not stabilize the TEC against dissociation. Indeed, S4 ⁸² and RfaH ⁶⁴ reduce RNAP pausing but do not have strong effects at intrinsic termination sites *in vitro*. RfaH dramatically reduces Rho-mediated polarity ²⁹ but does not prevent RNAP dissociation at the end of an operon *in vivo* ³⁸.

What is the molecular mechanism of antitermination? The branched mechanism of termination (Figure 3) suggests that an antitermination factor could inhibit the TEC isomerization into the elemental pause conformation or act at a later step, by inhibiting the binding/action of Rho or formation of the terminator hairpin. Antiterminators commonly target both steps. Bacteriophage λ N and Q proteins not only reduce pausing but also actively stabilize the TEC against dissociation at intrinsic terminators ^{83–85}. This activity is perhaps mediated via their additional contacts to the nascent RNA or the RNA exit channel ^{85–86}. However, RfaH inhibits Rho-mediated termination by excluding the Rho cofactor NusG from the TEC ^{29,38} and increasing translation³¹.

3. NUSG FAMILY OF REGULATORS

Recent studies of the *E. coli* NusG and its operon-specific paralog RfaH illustrated a universally conserved mechanism of RNAP processivity and highlighted the regulatory diversity of the proteins that constitute the NusG family, as well as their key roles in coordination of RNA synthesis with other cellular processes. Here, we will briefly review the history, structure and activities of these proteins.

3.1 NusG, a general transcription factor

NusG has been extensively studied in *E. coli*, whereas only fragmentary data on NusG homologs from other *Bacteria* are available. *E. coli* NusG is an abundant (~10,000 copies/cell⁸⁷) and essential⁸⁸ protein 181 residues in length. NusG was identified genetically⁸⁹ and biochemically^{90–91} as a factor required for efficient λ N-mediated antitermination. Subsequent studies demonstrated that NusG directly interacts with RNAP, Rho, and ribosomal protein S10³² (Figure 4), and possibly with a termination factor Nun⁹². These contacts enable NusG to mediate its diverse, and sometimes opposite, effects on gene expression: NusG increases Rho-dependent termination⁸⁹ and RNA chain elongation rates^{93–94}, is essential for efficient transcription of rRNA genes⁹⁵ and Nun-dependent termination⁹⁶, and may couple transcription to translation³².

3.1.1. λ N antitermination complex—The first antitermination mechanism, mediated by the N protein, was discovered in bacteriophage λ . Roberts proposed that λ N instructs RNAP to ignore Rho-dependent terminators in the early genes⁹⁷, enabling transcription of the entire phage genome. Genetic screens for host proteins involved in λ N antitermination identified several N-utilization substances, NusA, NusB, NusE (the ribosomal protein S10) and NusG, that, together with N, assemble into a complex modifying RNAP to a termination-resistant form⁹⁸. Assembly of this complex also requires two elements in the nascent RNA, boxA and boxB, jointly called the *nut* site^{99–100}. Interactions of N with the boxB RNA hairpin and NusA are sufficient for antitermination in a minimal *in vitro* system⁸¹, but long-range modification requires NusB/S10 contacts to boxA RNA¹⁰¹. The highly processive antitermination complex, which is able to act far downstream from the *nut* site⁹¹, is likely additionally stabilized by other interactions, such as those of S10 with RNAP¹⁰² and NusG³².

3.1.2. rRNA antitermination—RNAP transcribing the rRNA (*rrn*) genes moves twice as fast as the enzyme transcribing mRNAs and is resistant to the action of Rho¹⁰³. These properties are conferred by the assembly of a ribonucleoprotein complex that shares many features with the N-antitermination complex: *rrn* operons encode box A and box B elements and all the Nus factors are required for efficient synthesis of rRNA¹⁰⁴. NusG and NusB are necessary for high rate of *rrn* transcription⁹⁴, implying that an essential S10, which interacts both with NusG and NusB and is required for NusB binding to box A¹⁰¹, is also an integral part of the *rrn* antitermination complex. In addition, ribosomal protein S4, and perhaps other cellular factors, are involved⁸², likely defining the key feature of the *rrn* complex - in contrast to the λ N system, RNAP is efficiently stopped by intrinsic terminators¹⁰⁵, a property that is essential to prevent read-through beyond the end of the *rrn* operon.

3.1.3. Pausing and elongation rate—Antitermination properties of NusG could be explained by its effects on RNA chain elongation. NusG increases the rate of elongation *in vitro*^{93,106} and *in vivo*⁹⁴, and its effects are most pronounced at sites where RNAP is prone to backtracking⁵¹. An apparent increase in the rate of RNA chain elongation could be due to suppression of RNAP isomerization into off-pathway states or an acceleration of nucleotide addition along the main pathway or in the elemental pause intermediate (Figure 2). In ensemble experiments, NusG reduces the half-life of backtracked paused complexes,

supporting a model that NusG induces forward translocation of RNAP, thereby facilitating binding of the substrate NTP to the $i+1$ site^{107–109}. Recent single-molecule analysis¹¹⁰ lent further support to this model, demonstrating that NusG modestly increases the pause-free elongation velocity and reduces the RNAP propensity to transiently pause during elongation; stabilization of the pause-resistant, post-translocated state of the TEC by NusG is sufficient to explain this effect. In addition to disfavoring short-lived, frequent pauses, NusG had a more pronounced effect on long pauses, likely accompanied by backtracking¹¹⁰. Some of the NusG effects were position specific, raising a possibility of some sequence-specific contacts with the nontemplate DNA¹¹¹; specific interactions may underlie the delay of RNAP by *B. subtilis* NusG¹¹².

3.1.4. Rho- and Nun-dependent termination—Paradoxically, despite its elongation-enhancing properties, NusG is also required for efficient factor-dependent termination by bacteriophage HK022 Nun and *E. coli* Rho proteins¹¹³. The latter activity may represent the key cellular function of NusG which, together with Rho, limits expression of horizontally acquired genes²² and antisense RNAs²³. NusG becomes largely dispensable in strains lacking toxic prophage genes²².

NusG increases Rho-dependent termination at some sites *in vivo*¹¹³ and *in vitro*¹¹⁴, but does not increase Rho affinity or ATPase activity. In the presence of NusG, Rho releases shorter RNAs, acting at more promoter-proximal positions^{87,114–115}. The mechanism by which NusG increases Rho termination is yet unclear. A recent study shows that NusG increases Rho-termination only at weak *rut* elements characterized by a low C/G ratio²³, suggesting that NusG may favor Rho interactions with RNA; indeed, NusG slows the off-rate of Rho from the nascent transcript¹¹⁶. However, this simple model is inconsistent with an earlier observation that NusG does not change the Rho concentration required for termination¹¹⁴ and recent studies which concluded that NusG facilitates EC dissociation by Rho^{117–118}.

3.1.5. NusG and translation—A role for NusG in translation has been suggested by the presence of the KOW motif, a signature of several ribosomal proteins, in NusG¹¹⁹ and by an observation that the translation rate decreased after the depletion of NusG⁹⁴, which also leads to deceleration of transcription. Recent studies by Nudler and coauthors demonstrated that the rates of translation and transcription are highly correlated: the slowdown of translation induced by the addition of antibiotics or rare codons in the mRNA leads to a concomitant slowdown of transcription⁹⁹. A model derived from this study suggests that a trailing ribosome could push a backtracked RNAP forward to prevent pausing and maintain rapid and processive RNA synthesis⁹⁹.

Coincident work by Rösch and collaborators revealed that, in addition to its role in preventing backtracking, NusG may serve as a direct physical link between transcription and translation³². These authors reported an NMR structure of the NusG-CTD/S10/NusB complex that could explain how NusG could bridge the RNAP and the ribosome on one hand, and participate in the assembly of antitermination complexes on the other. S10 is not stable in isolation and can exist in the cell either as a heterodimer with NusB or as a component of the 30S ribosomal subunit. NusG-CTD and NusB interact with the opposite

faces of S10 (Figure 5), and S10 contacts with the ribosome and NusB are mutually exclusive. S10 is well suited to serve as a tether - it is deeply anchored in the ribosome via a long loop ¹²⁰, making it one of the most strongly bound ribosomal proteins. This study established that a link between transcription and translation is feasible, and suggested that other NusG paralogs, which share the conserved S10-binding interface, could couple the two processes. Future studies will address the details, timing, and regulatory functions of these contacts.

3.2 RfaH, a regulator of virulence and fertility genes

3.2.1. RfaH discovery—RfaH was discovered among the collection of spontaneous rough mutants of *Salmonella typhimurium* deficient in lipopolysaccharide (LPS) core biosynthesis. Since *rfaH* mutants lacked galactosyltransferase activity, the product of the *rfaH* gene was tentatively categorized as an enzyme that adds a galactose unit to glucose-modified LPS ¹²¹. Subsequent work has demonstrated, however, that *rfaH* mutations caused defects in activities of several LPS glycosyltransferases, suggesting that RfaH is a positive regulator of synthesis or activity of these enzymes, rather than a transferase ¹²². Mutations in *rfaH* have been isolated independently in *E. coli* as *sfrB* ¹²³ and *hlyT* ¹²⁴ alleles, functions required for the expression of the F plasmid *tra* operon and the *hly* operon for synthesis and secretion of α -hemolysin. Mutations in *sfrB* conferred pleiotropic phenotypes, from altered sensitivity to bacteriophages and antibiotics to synthesis of nonfunctional flagella, suggesting defects in the cell envelope ¹²³.

Analysis of *tra* operon expression revealed that SfrB was required to block premature transcription at Rho-dependent termination sites ¹²⁵. The same study showed that *sfrB* mutations reduced synthesis of full-length LPS, several outer membrane proteins, and functional flagella, identifying RfaH as a broad specificity antitermination factor regulating both plasmid- and chromosomally-encoded genes ¹²⁵. Later studies demonstrated that RfaH activates expression of the CNF-1 toxin ¹²⁶, group II and III polysaccharide capsules ^{127–128}, O-antigen ^{129–130}, and hemin receptor ¹³¹, but represses Ag43, thereby inhibiting biofilm formation. ¹³²

3.2.2. RfaH target genes—All known RfaH targets share several common features. First, they encode proteins involved in the synthesis, transport, or assembly of membrane and extracellular components ³⁹. Second, they share a common motif, called *ops*, in their upstream segments. Third, they comprise long horizontally-transferred operons; many of these genes are located in pathogenicity islands or on plasmids. Consistently, inactivation of *rfaH* dramatically reduces virulence of *E. coli* ¹³³ and *Salmonella typhimurium* ¹³⁴ while increasing invasiveness into antigen-presenting cells and intracellular killing, thereby making the *rfaH* mutant into a promising live attenuated vaccine ¹³⁴.

ChIP-chip analysis of RfaH association with *E. coli* genes ³⁸ revealed that RfaH is recruited to RNAP at the *ops* site, at the same position as observed *in vitro* (within the resolution of this analysis), and remains associated with the TEC throughout transcription of the entire operon (Figure 6).

Similarities in localization and functions of proteins encoded by RfaH-controlled genes suggest that RfaH expression could be induced by some environmental cues, for example, upon interactions with host cells. Very little is currently known about control of RfaH expression. In *Salmonella*, *rfaH* expression increases at the onset of the stationary phase and is induced by alternative σ factors RpoS and RpoN¹³⁵. Interestingly, expression of *rfaH* homolog *anfA1* from a cryptic pathogenic plasmid pADAP in *Serratia entomophila* is also dependent on RpoS¹³⁶

3.2.3 The *ops* element and its role in RfaH recruitment—Comparison of several RfaH-controlled operons identified an 8-bp sequence motif, GGCGGTAG¹³⁷. Deletion of this motif from its position upstream of the *hlyCABD* operon strongly increased polarity, mimicking the phenotype of the *rfaH* deletion¹²⁴. Conversely, placing this element downstream from a heterologous promoter increased expression far downstream¹³⁶. Hence, this element was termed *ops* (for operon polarity suppressor) and proposed to mediate RfaH recruitment^{39,137}.

The *ops* element comprises a part of the JUMPstart sequence identified earlier in operons encoding biosynthesis of polysaccharides¹³⁸; *ops* is conserved in many *Bacteria* (from *Yersinia enterocolitica* and *Vibrio cholerae* to *Halomonas maura*¹³⁹), and *rfaH* genes from several species are interchangeable *in vivo* and *in vitro*¹⁴⁰.

The presence of *ops* reduces the concentration of RfaH required to suppress transcription polarity in a partially purified *in vitro* system,¹³⁷ suggesting that *ops* mediates RfaH recruitment to RNAP. Given that *ops* must be positioned downstream from a promoter to elicit anti-polar effects, it could function as an RNA or DNA; however, RfaH does not bind to either DNA or RNA and does not have any recognizable nucleic acid-binding motifs. These observations, and the identification of RfaH as a homolog of NusG³⁹, which participates in assembly of transcription antitermination complexes, led Koronakis and colleagues to propose that *ops* interacts with an unknown cellular factor¹⁴¹ to tether RfaH to RNAP, which is consequently modified into a processive state³⁹.

Interestingly, however, RNAP was shown to pause *in vivo*¹⁴² and *in vitro*⁵¹ three nucleotides downstream from the *ops* element, positioning the conserved sequence inside the transcription bubble. If the *ops*-paused RNAP serves as a target for RfaH recruitment, as observed during the assembly of other antitermination complexes, the *ops* could act either indirectly, by altering RNAP conformation, or directly, by establishing base-specific contacts with RfaH or another factor. In the second case, only the nontemplate DNA exposed on the RNAP surface would be available for interactions (Figure 1). Indeed, RfaH directly crosslinks to the nontemplate DNA and binds to any heteroduplex TECs in which the nontemplate strand (but not RNA or the template strand) carries the *ops* sequence⁶⁴. This mechanism of recruitment may be utilized by other proteins targeted to the nontemplate strand transiently exposed on the TEC surface, for example, during activation-induced cytidine deamination¹⁴³ required for somatic hypermutation of immunoglobulin genes.

The *ops* element may play several roles in the RfaH mechanism. First, *ops* induces efficient pausing by RNAP, possibly providing sufficient time for RfaH recruitment to the TEC.

RfaH must bind to a short fragment of the nontemplate strand, which would be exposed only for a fraction of a second on a rapidly moving RNAP; once RNAP moves past the *ops* site, it cannot be modified by RfaH. Second, *ops* likely makes base-specific contacts with RfaH. Although a structure of RfaH bound to the TEC is not yet available, dramatic effects of point mutations in *ops* and single substitutions in RfaH on RfaH/*ops* interactions³⁴, together with crosslinking data⁶⁴ and molecular modeling³⁵ support the existence of such interactions. Third, *ops* may induce conformational changes in the TEC that are required for RfaH binding or antitermination modification. For example, *ops*-induced scrunching of the nontemplate DNA could favor both RfaH binding to the TEC and escape of RfaH-bound RNAP from the *ops* site⁴³; similarly, scrunching facilitates disruption of σ /DNA interactions during RNAP escape from the promoter¹². Finally, *ops* restricts the RfaH action to a small set of operons, thereby avoiding interference with the essential NusG³⁸. RfaH and NusG share the same binding site on the TEC but only RfaH requires a specific site for recruitment³⁵.

3.3 Structural features of the NusG regulators

3.3.1. Domain organization—Proteins from the NusG family follow a common theme of domain organization. In general, all homologs contain a single NGN (NusG-like N-terminal) domain¹⁴⁴, followed by one or more CTDs containing a KOW motif¹¹⁹ (Figure 7a). This makes NusG a perfect module to interact with both nucleic acid and proteins¹⁴⁵. The N-terminal domain (NTD) binds to RNAP and nucleic acids to mediate effects on RNA synthesis, while the KOW domain functions as a tether bridging the transcription apparatus and other cellular machineries, such as Rho or ribosome⁹². *E. coli* NusG represents a minimal system that contains each of these domains⁹² whereas other bacterial proteins, for instance *Aquifex aeolicus* NusG, also carry an additional domain that may perform species-specific regulatory functions¹⁴⁵. By contrast, eukaryotic proteins contain multiple KOW domains (3–6, although the domain numbers and assignment remain debatable¹⁴⁶). Eukaryotic NusG homologs also acquired additional regulatory domains to satisfy essential regulatory requirements in higher organisms; these domains may be subjected to modifications (such as phosphorylation) and serve as docking sites for other proteins²⁷.

3.3.2. Structural conservation of NusG proteins—The sequence, core structure and the mechanism of RNAP modification are well conserved among all NusG homologs¹⁴⁷, consistent with strong evolutionary relationships among these proteins¹⁴⁴ (Figure 7b). Simpler monomeric proteins (present in extant *Bacteria*) gave rise to a more complex heterodimeric system in eukaryotes and *Archaea*. In these organisms, NusG homolog Spt5 forms a heterodimer with the small zinc-binding protein Spt4¹⁴⁸ or its homolog RpoE, respectively¹⁴⁹; RpoE was earlier considered to be a subunit of archaeal RNAP¹⁵⁰.

Structures of NusG homologs from all three domains of life are available^{35,92,145,150–152} (Figure 8). NGN has similarities to ribonucleoprotein domain motifs¹⁴⁵ and is composed of four stranded antiparallel β -sheets that are covered by two helices on one side and by another helix on the other side, forming a α - β - α sandwich. The single helix serves as a connecting region between the NGN and KOW domains and interacts with several conserved hydrophobic residues that could make interactions with RNA bases¹⁴⁵. In

archaeal NusG, NGN is shortened by a 20-residue deletion of a long β -hairpin loop between the two β -strands. The KOW fold¹¹⁹ is commonly observed making contacts to RNA in several RNA-binding proteins¹¹⁹. Despite suggestions that both NusG domains may interact with RNA, neither domain has been shown to have RNA-binding activity, and only one NusG homolog, from *Thermotoga maritima*, binds to RNA (and DNA) nonspecifically¹⁵³.

In most NusG-like proteins, the KOW domain is a highly bent, mostly stranded β -barrel fold which is freely connected to NGN by a flexible linker loop. By contrast, the structure of RfaH CTD is unique (Figure 8c) - in the free protein, it is folded as an α -helical hairpin that is stabilized by hydrophobic interactions with the NTD³⁵. The α -helical fold is unstable, and the CTD refolds into a characteristic β -barrel KOW state upon domain dissociation triggered by recruitment to RNAP or weakening of the interdomain contacts³¹.

3.3.3. The autoinhibited state of RfaH—These differences in domain organization underlie important regulatory differences in the recruitment of NusG-like proteins to the TEC. RfaH is a highly specialized regulator whose action is limited to a small set of *E. coli* operons³⁸. RfaH-like proteins are present in only some bacterial species³⁸ and are absent in *Archaea* and eukaryotes. In free RfaH, the two domains form an interface stabilized by conserved hydrophobic interactions; this interface masks the RNAP-binding site located on the NGN and needs to be destabilized to allow for RfaH binding to the TEC³⁵. RfaH activation via domain separation is thought to be triggered by binding to the *ops*-paused RNAP; consistently, the isolated NTD does not require *ops* for binding to, and modification of, the elongating RNAP³⁸. In contrast, NusG and Spt5 are essential and ubiquitous general elongation factors – they do not display sequence-specific effects on transcription and are associated with RNAP transcribing numerous genes^{37,154} because the RNAP-binding surface on the NGN domain of NusG is exposed¹⁴⁵, obviating the need for activation and allowing for binding to any TEC.

A similar closed (spring-loaded) state in which the NusG NTD and CTD interact to bury a hydrophobic RNAP-binding site on the NTD was proposed based on the crystal structure of *A. aeolicus* NusG¹⁵¹. However, this model is inconsistent with an independent crystal structure of *A. aeolicus* NusG¹⁴⁵, solution structures of *A. aeolicus*, *E. coli*, and *Thermus thermophilus* NusGs^{92,152,155}, and functional studies of *E. coli* NusG⁹².

3.3.4. Functional regions of RfaH and NusG—Homology between RfaH and NusG³⁹ and their common antip pausing effects *in vitro*^{64,93} implied that both factors may interact with the TEC similarly. Even though NusG was studied much more extensively, it was the analysis of RfaH that led to identification of the NTD determinants responsible for interactions with the TEC. Two properties of RfaH facilitated its mechanistic analysis and served as constraints in the building of a heterologous model of *E. coli* RfaH bound to the *T. thermophilus* TEC (Figure 9): (i) RfaH directly interacts with the *ops* element in the nontemplate DNA strand⁶⁴ and (ii) the isolated RfaH-NTD recognizes *ops* and mediates antip ausing³⁵. In the model, the NTD positioned to contact the exposed segment of the nontemplate stand was in an interacting distance from the tip of the β' CH domain, the major binding site for σ factors (Figure 9a). This model is supported by several observations: the RfaH NTD efficiently competes with σ ⁷⁰ for binding to the β' CH during elongation⁶⁵

substitutions of several basic residues predicted to interact with the DNA compromise *ops* binding³⁴, and substitutions of the hydrophobic residues in RfaH and in the β' CH inhibit RfaH association with the TEC^{35,65}.

All transcriptional activities of RfaH are mediated by the NTD. Despite its small size (90 residues), the NTD contains three separate regions that mediate DNA recognition, retention on the RNAP throughout transcription, and antipausing modification (Figure 9b).

DNA binding: RfaH NTD binds to a rather unusual target, the single-stranded DNA strand interacting with the RNAP surface and lacks recognizable DNA-binding motifs⁶⁴. Mutational analysis identified five polar and charged residues (Lys¹⁰, Arg¹⁶, His²⁰, Thr⁷² and Arg⁷³) that are required for RfaH-induced delay at *ops*, which is presumed to arise from persistent RfaH/DNA interactions. These residues are the least conserved between RfaH and NusG.

β' CH binding: The antipausing activity of RfaH depends on its stable contacts to RNAP. Molecular modeling and mutational analysis of the nearby β' residues suggested that a hydrophobic patch on the RfaH NTD interacts with the tip of the β' CH³⁵. Substitutions of several RfaH residues (Trp⁴, Tyr⁵⁴, and Phe⁵⁶) decreased its activity at low RfaH concentration, consistent with the reduced affinity of the altered proteins³⁴.

β gate loop (β GL) binding: In a simple scenario, RfaH interactions with the β' CH and the nontemplate DNA at the upstream part of the bubble would be sufficient to promote strand reannealing, favoring translocation and reducing pausing. However, mutational analysis revealed a third motif (HTT) that was required for antipausing but dispensable for RfaH binding to the TEC³⁴. In the RfaH/TEC model, the HTT motif makes a contact with the β gate loop (β GL, Figure 9a). We showed that the β GL, which belongs to the β pincer that likely moves in concert with the β' clamp during isomerization into a paused state⁵⁸, is essential for antipausing by RfaH and NusG²⁹.

In vitro analysis of NusG variants with substitutions of residues in the homologous patch⁹² and two-hybrid assays¹⁵⁶ support the assumption that the N-terminal domains of RfaH and NusG bind to the β' CH similarly³⁸. Defects conferred by substitutions of residues Phe⁶⁵ and Tyr⁶⁸ (which corresponds to Tyr⁵⁴ in RfaH) were consistent with a reduced affinity for the TEC⁹². Furthermore, NusG NTD functionally interacts with the β GL²⁹. The phenotypes of the NTD mutants, competition between RfaH and NusG³⁸ and structural modeling suggest that RfaH, NusG and Spt5 bind to the same sites on the TEC³⁶. This indicates that the RfaH and NusG NTDs play identical roles in RNAP modification. However, RfaH makes tighter contacts with RNAP owing to a larger hydrophobic area on the RfaH NTD; this allows RfaH to remain associated with RNAP throughout transcription³⁸.

By contrast, the CTDs determine the protein-specific properties of RfaH and NusG. Both CTDs make similar contacts to S10^{31–32} which are thought to mediate cross-talk with the translation apparatus as well as NusG assembly into the *rrn* antitermination complex. The remaining functions of the CTD are very different. In NusG, the CTD directly interacts with Rho and perhaps with Nun⁹², and may establish some additional contacts to other cellular

factors. In RfaH, the CTD prevents spurious recruitment to the TEC at non-*ops* sites³⁵ by establishing a closed autoinhibited state but is not known to make any other contacts, including Rho³¹.

Mutational studies of the *E. coli* NusG identified residues that are necessary for Rho- and Nun-dependent termination (Figure 10). Analysis of NusG variants defective in λ exclusion by Nun identified a cluster of CTD residues (Phe¹⁴⁴, Asn¹⁴⁵ and Phe¹⁶⁵) that could constitute a Nun-binding site^{92,157}; substitutions of other residues (Trp⁹, Trp⁸⁰, Ile¹⁰³, and Val¹⁷⁸) are more likely to exert indirect effects on Nun function. The pattern of Rho-defective NusG mutants is more complex. The residues at which substitutions displaying defects in Rho binding and function have been isolated (Gly¹⁴⁶, Val¹⁴⁸, Leu¹⁵⁸, Val¹⁶⁰, and Ile¹⁶⁴) are mostly buried in the CTD^{92,117}, which must undergo conformational changes to allow for a direct contact of these residues with Rho. The effects of these variants could be indirect, arising as a consequence of structural changes in the CTD, and additional studies will be required to dissect the molecular details of NusG/Rho interactions.

3.4. RfaH and NusG in other Bacteria

Many *Bacteria* contain several NusG homologs, among which one can be viewed as a general transcription factor that carries out housekeeping functions, while others perform some dedicated tasks; we termed the latter class specialized NusGs, or NusG^{SP}³⁸. This view was based on *E. coli*, where the essential NusG regulates expression of many genes whereas the dispensable RfaH is specifically targeted to a limited set of genes that are involved in extracytoplasmic functions, including virulence and fertility.

In the case of *E. coli* RfaH, the basis for this specificity is well understood - the *ops* element is necessary and sufficient for defining the RfaH regulon³⁸. However, even in *E. coli*, additional uncharacterized NusG paralogs, such as ActX encoded on a large conjugative multidrug resistance plasmid pOLA52¹⁵⁸, exist. ActX regulates the expression of the conjugative functions encoded by the F plasmid¹²³. Its homology to RfaH and *actX* location on an autonomous genetic element suggests that it could be involved in the transcriptional control; however, such a function remains to be demonstrated.

Studies of other NusG^{SP} homologs support a common mode of action, activation of expression of long operons by antitermination. Induction of *S. entomophila* AnfA1, which has a high similarity to RfaH¹⁵⁹, allowed expression of *afp*, an 18-ORF anti-feeding prophage gene cluster, which encodes a virus-like structure thought to mediate the transport of toxins that cause a cessation of feeding by the grass grub larvae¹⁶⁰. Strikingly, a perfect *ops* element is located 108 bp upstream of the first ORF, indicating that AnfA1 may share the fine details of TEC recruitment with RfaH. Similarly, distant RfaH orthologs from *Klebsiella pneumoniae*, *V. cholerae*, and *Y. enterocolitica* readily recognized *ops in vitro* and complemented the disruption of the *E. coli* *rfaH* gene¹⁴⁰. In *Myxococcus xanthus*, enzymes required for the production of the antibiotic TA (myxovirescin) are encoded within a 36-kb TA gene cluster. Expression of this operon requires the first gene in the cluster, *taa*¹⁶¹. TaA is homologous to, and has been proposed to act like, RfaH; TaA is essential for the antibiotic production but not for growth and development¹⁶¹. In *Bacteroides fragilis*, expression of eight different capsular polysaccharide operons is controlled by

antitermination; each operon is controlled separately by a protein encoded by the first gene in each operon¹⁶². Although the detailed mechanism of these UpxY proteins remains unknown, the available data suggest that they belong to the nusG^{SP} family of regulators¹⁶³.

NusG^{SP} proteins control expression of highly specialized targets, and could thus be expected to differ in fine details of their mechanisms. However, even though “general” NusGs are ubiquitous, their properties may also differ. For example, NusG is not essential in *B. subtilis*¹⁶⁴ or *Staphylococcus aureus*¹⁶⁵, and, unlike the *E. coli* NusG, *B. subtilis* NusG strongly increases pausing at some regulatory sites *in vivo* and *in vitro*¹⁶⁶. *T. thermophilus* NusG does not bind to Rho¹⁵² and slows down rather than facilitates elongation by its cognate RNAP, even though it binds to the nontemplate DNA strand, competes with σ^A , and induces forward translocation by RNAP¹¹¹, properties shared with *E. coli* RfaH and NusG. NusGs may also differ in their regulatory interactions: *T. maritima* NusG is the only member of this family known to bind nonspecifically and cooperatively to nucleic acids¹⁵³, and one of three NusG paralogs in *Thermoanaerobacter tengcongensis* interacts with DnaA¹⁶⁷.

Striking structural similarity among NGN domains from all kingdoms (Figure 8e) implies a similar molecular mechanism of RNAP control. However, NusG-like proteins can be very versatile and include factors controlling just one operon (AnfA1 and TaA), a small set of operons with related functions (RfaH), or many operons (NusG). The specificity of recruitment to the TEC is likely controlled by contacts to the nontemplate DNA but the details of recruitment remain to be determined for any protein from this family.

4. REGULATION OF RHO-MEDIATED POLARITY

In *Bacteria*, Rho surveys the mRNA quality - if translation is inefficient, Rho releases the nascent RNA prematurely⁷² establishing polarity along the operons. Polarity is characteristic of all bacterial operons but is particularly pronounced in poorly translated, horizontally-transferred operons. Regulation of Rho-mediated termination constitutes the main functions of RfaH and NusG which, despite their similar effects on RNA chain elongation *in vitro*, play opposite regulatory roles in the cell.

NusG directly binds to Rho⁹² and stimulates RNA release by Rho¹¹⁶ to repress horizontally acquired DNA²² and antisense transcription²³. Although NusG can also inhibit polarity, acting as part of *rrn* antitermination complex¹⁰⁴ or a bridge between RNAP and the leading ribosome³², its essential role is to silence foreign DNA²². However, some horizontally-transferred genes that play important cellular roles have to be protected from the joint action of Rho and NusG. In *E. coli*, RfaH has evolved to play this role. Most RfaH-controlled operons are horizontally transferred¹⁶⁸ or plasmid-encoded³⁹. Observations that *rfaH* defects are suppressed by mutations in *rho*¹⁶⁹ and that RfaH almost completely abolishes Rho-mediated polarity in the *rfb* operon²⁹ support a model in which the sole function of RfaH is to counteract Rho-mediated polarity in long, poorly translated operons.

Since RfaH and NusG bind to the same site on the TEC, a mechanism that ensures that both proteins are targeted to their respective genes must exist in the cell (Figure 11). Our studies demonstrated that, in contrast to NusG, which associates with RNAP transcribing many operons³⁷, RfaH targets only those few operons that contain an *ops* sequence³⁸. This

specificity is dictated by the need for RfaH activation by *ops* (see above). RfaH recognizes its target DNA even in the presence of a large excess of NusG, likely because *ops* sequences are located near promoters, whereas NusG is recruited to the TEC further downstream³⁷. Once bound, RfaH prevents NusG loading³⁸.

RfaH has only a modest inhibitory effect on Rho-dependent termination *in vitro* but has a dramatic (hundred fold) effect on Rho-mediated polarity within the *rfb* operon *in vivo*²⁹. RfaH could interfere with Rho action directly, but evidence in support of this model is lacking. Instead, our data suggest that RfaH uses three distinct mechanisms to inhibit Rho-dependent termination.

4.1. Antipausing modification of RNAP

Pausing is a prelude to termination, and Rho is known to preferentially target paused RNAPs¹⁷⁰. One mode of RfaH action is to modify RNAP into a pause-resistant state¹⁰⁹ by restricting the mobility of the clamp, thereby inhibiting pausing and, in turn, termination²⁹. This model was proposed based on the identification of the β' CH as the RfaH-binding site and then confirmed by identification of the β GL as the second contact site on the TEC²⁹. We proposed that RfaH binds simultaneously to the β' CH and β GL located on the opposite sides of the main channel, locking the clamp in a closed state and preventing structural rearrangements that occur during transcriptional pausing⁵⁸. By bridging the gap between the β' clamp and the β lobe domains, RfaH completes a proteinaceous ring around the DNA template, thereby increasing RNAP processivity (Figure 11).

Structural studies of archaeal Spt5 proteins bound to their targets on RNAP^{147,171} provided strong support for the hypothetical model of RfaH/TEC, suggesting that the clamping mechanism is shared by all NusG-like proteins^{7,28}. However, we demonstrated that this mode of action makes only a minor (two- to six-fold) contribution to the anti-polar effects of RfaH²⁹. The remaining two mechanisms (Figure 13) are specific to RfaH.

4.2. Exclusion of NusG

The second mode of RfaH action is to exclude NusG from the elongating RNAP through competition for binding to the β' CH (Figure 13a). We showed that RfaH efficiently outcompetes NusG *in vitro* and *in vivo* provided that the *ops* element is present in the transcribed region³⁸. Since NusG stimulates Rho-dependent termination, competition with NusG ensures that Rho-mediated RNA release is inhibited. Consistent with the post-recruitment role of NusG in Rho termination, Rho is associated with the *rfb* operon (Figure 5) even though NusG is nearly completely excluded³⁸, and Rho-dependent polarity is essentially abolished in the presence of RfaH²⁹. These observations suggest that NusG exclusion could be an important component of the anti-Rho activity of RfaH, but the relative contribution of this mechanism to the RfaH action *in vivo* remains to be elucidated.

4.3. Recruitment of the ribosome

All RfaH-controlled operons are horizontally transferred, suggesting inefficient translation. This view is supported by the strong polarity observed in the *rfb* operon *in vivo*²⁹ and suggests another mode of RfaH action – to increase translation of its target operons. This

hypothesis was put forth to explain differences in RfaH retention on the TEC *in vivo* and *in vitro*³⁸ and was consistent with RfaH interactions with the components of the translational machinery in the cellular context¹⁴¹ and with transcription-translation coupling⁹⁹ mediated by NusG/S10 contacts³².

During canonical translation initiation, recruitment of the 30S preinitiation complex is mediated by interactions between a Shine-Dalgarno (SD) element upstream of the AUG start codon and a complementary sequence in the 16S rRNA¹⁷². However, all known RfaH targets lack SD elements, raising the question of how the ribosome is recruited to these mRNAs. Our recent data indicate that RfaH can substitute for an SD element during initiation. We showed that removal of the SD element made the expression of a reporter *luxCDABE* operon dependent on the RfaH CTD³¹ and proposed a model in which RfaH/S10 contact increases the local 30S concentration, thereby recruiting the ribosome to the nascent mRNA synthesized by RfaH-bound RNAP.

This mechanism would facilitate the first round of translation, with the leading ribosome insulating the nascent RNA from Rho-induced release (Figure 13b). It is not known whether RfaH targets are translated many times but it is the pioneer round of translation that serves as a target for RfaH – once transcription is completed, the RNA is no longer in danger of Rho-mediated release.

5. LINKING TRANSCRIPTION TO CONCURRENT CELLULAR PROCESSES

5.1. The CTD as a universal communication module

Identification of NusG as the only ubiquitous transcription factor appears to imply that its key regulatory role is RNAP modification during elongation⁷; indeed, the underlying contacts between the NTD and the RNAP pincers are universally conserved^{28,36}. However, NusG-like proteins have also been implicated in other, equally important processes that must be coordinated with ongoing RNA synthesis in all domains of life. This coordination is provided by the CTDs, which tether the transcribing RNAP to diverse cellular partners, which include antitermination complexes and ribosome in prokaryotes and the RNA splicing, capping, and degradation machineries in eukaryotes.

Even in *Bacteria*, the CTD can interact with very different partners, whereas the presence of multiple KOW domains in eukaryotic and archaeal proteins provides additional diversity, which likely reflect the increased complexity of gene regulation²⁷. Studies from yeast demonstrated direct roles of Spt5 in chromatin remodeling and histone modifications. Chromatin remodelers Chd1¹⁷³ and Paf1¹⁷⁴ associate with Spt5. During transcript elongation, in addition to controlling RNAP processivity, Spt5 is also suggested to contribute to overcoming the nucleosome barrier¹⁷⁵. Proteins involved in splicing and capping are reported to associate with Spt5³³. *Drosophila* Spt5 interacts with exosome, suggesting a direct role in RNA degradation¹⁷⁶.

The absence of nuclear membrane allows co-transcriptional translation in *Bacteria* and *Archaea*¹⁷⁷. Coupling of these two processes offers their robust and economical regulation: Coupling is advantageous in maintaining the local concentration of the translated products in

the vicinity of transcription sites^{178–179} and prevents formation of R-loops¹⁸⁰. A translating ribosome could function as an antiterminator during attenuation¹⁸¹: an uncharged tRNA triggers ribosome stalling at the attenuator site, blocking the formation of the terminator hairpin and allowing expression of the downstream genes, for example, those encoding histidine or tryptophan biosynthesis. Uncoupled translation, *e.g.*, as a result of a premature stop codon, allows Rho-mediated polarity to rectify faulty transcripts⁷².

Although the importance of transcription-translation coupling is established in *Bacteria*, its molecular details are not completely known. In addition to mRNA, other factors may directly link RNAP to ribosome. Many proteins (reviewed in¹⁸²) are shared between the two processes, and several ribosomal proteins are involved in feedback control of their genes¹⁸³. NusG proteins likely play the key role in bridging RNAP and ribosome, with the NusG CTD providing an interacting surface for S10³². For a small protein like NusG, holding two large nucleoprotein complexes together appears to be an immense task, particularly since the speed of both RNAP and ribosome could be modified independently. Given that NusG/S10 interactions are weak in a “binary” NusG-CTD/S10/NusB complex³², other factors may strengthen the link between RNAP and ribosome, once the first contact is made by NusG.

The role of NusG in coupling is supported by many reports (see 3.1.5) but it is yet unclear what contribution it makes to the expression of cellular genome. In fact, it is the quality control function of NusG, mediated by its contacts with Rho that are incompatible with the CTD/S10 interactions, appears to underlie its essentiality in *E. coli*^{22–23}. In contrast, ribosome recruitment by RfaH could be critical for its dramatic antipolar effects. RfaH increases the expression of its target genes, which are poorly translated, by hundreds of fold²⁹; this effect could be achieved simply by removing a ribosome-binding site from a heterologous mRNA, implying that RfaH may load ribosome by an alternative route³¹. Strikingly, productive interactions between RfaH and ribosome require a complete transformation of the CTD (Figure 14).

5.2. Transformation as a novel regulatory paradigm

In RfaH, the CTD is folded as an α -helical hairpin, and the conserved residues implicated in NusG/S10 interactions are inaccessible³⁵. However, noting that the sequence of the RfaH CTD is readily compatible with the NusG-like β -barrel fold³⁵, we proposed that after domain dissociation, a prerequisite for the NTD binding to RNAP, the CTD undergoes an unprecedented refolding into a β barrel to interact with the ribosome during coupled translation elongation³⁸. Recently, we demonstrated that the RfaH CTD does indeed make a barrel when its interactions with the NTD are severed or destabilized³¹. This newly formed barrel is nearly identical to NusG CTD and binds to S10 in a similar manner³¹, apparently enabling ribosome loading on mRNA lacking canonical translation initiation signals. It remains to be seen whether this hypothetical contact is maintained throughout translation elongation, but RfaH is known to be stably associated with the transcribing RNAP until its dissociation at a terminator³⁸.

The RfaH CTD is truly unique in its structural and regulatory plasticity, prompting us to define a new class of transformer proteins⁴⁰, in which a structural change coincides with an

acquisition of a new function. Along with other proteins, such as lymphotactin¹⁸⁴, RfaH combines features of *moonlighting proteins*¹⁸⁵, which use the same fold to carry out different functions, and *metamorphic proteins*¹⁸⁶, which can adopt different folds in identical environments. However, RfaH CTD stands out from this list by completely changing both its secondary structure elements and its function. This observation is even more surprising in light of the ubiquity of SH3 fold that the β barrel CTD possesses.

We argue that transformation is not an isolated phenomenon unique to RfaH but may be utilized by other proteins from the NusG family to broaden their regulatory repertoire. It is possible that not all of the regulatory interactions reported for NusG-like proteins are mediated by the β -fold. It is also possible that other multi-domain proteins that operate at crossroads of complex regulatory networks could employ similar strategies in interactions with their cellular partners. Transformation opens up an entirely new paradigm for regulation and intermolecular communication and puts proteins at par with ancestral RNA enzymes in terms of structural flexibility.

Acknowledgments

We thank Dmitri Svetlov for comments on the manuscript. This work was supported by the National Institutes of Health grant GM67153.

References

1. Mooney RA, Artsimovitch I, Landick R. J Bacteriol. 1998; 180:3265. [PubMed: 9642176]
2. Steitz TA. Curr Opin Struct Biol. 2009; 19:683. [PubMed: 19811903]
3. Gleghorn ML, Davydova EK, Rothman-Denes LB, Murakami KS. Mol Cell. 2008; 32:707. [PubMed: 19061645]
4. Deshpande AP, Patel SS. Biochim Biophys Acta. 2012; 1819:930. [PubMed: 22353467]
5. Werner F, Grohmann D. Nat Rev Microbiol. 2011; 9:85. [PubMed: 21233849]
6. Svetlov V, Nudler E. Biochim Biophys Acta. 2012
7. Werner F. J Mol Biol. 2012; 417:13. [PubMed: 22306403]
8. Gruber TM, Gross CA. Annu Rev Microbiol. 2003; 57:441. [PubMed: 14527287]
9. Saecker RM, Record MT Jr, Dehaseth PL. J Mol Biol. 2011; 412:754. [PubMed: 21371479]
10. Treutlein B, Muschielok A, Andrecka J, Jawhari A, Buchen C, Kostrewa D, Hog F, Cramer P, Michaelis J. Mol Cell. 2012; 46:136. [PubMed: 22424775]
11. Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, Ebright RH. Science. 2012; 338:1076. [PubMed: 23086998]
12. Revyakin A, Liu C, Ebright RH, Strick TR. Science. 2006; 314:1139. [PubMed: 17110577]
13. Daube SS, von Hippel PH. Proc Natl Acad Sci U S A. 1999; 96:8390. [PubMed: 10411885]
14. Mooney RA, Darst SA, Landick R. Mol Cell. 2005; 20:335. [PubMed: 16285916]
15. Shimamoto N, Kamigochi T, Utiyama H. J Biol Chem. 1986; 261:11859. [PubMed: 2427513]
16. Feklistov A, Darst SA. Cell. 2011; 147:1257. [PubMed: 22136875]
17. Mooney RA, Landick R. Genes Dev. 2003; 17:2839. [PubMed: 14630944]
18. Li GW, Oh E, Weissman JS. Nature. 2012; 484:538. [PubMed: 22456704]
19. Nechaev S, Geiduschek EP. Embo J. 2006; 25:1700. [PubMed: 16601684]
20. Henkin TM, Yanofsky C. Bioessays. 2002; 24:700. [PubMed: 12210530]
21. Santangelo TJ, Reeve JN. J Mol Biol. 2006; 355:196. [PubMed: 16305799]
22. Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. Science. 2008; 320:935. [PubMed: 18487194]

23. Peters JM, Mooney RA, Grass JA, Jessen ED, Tran F, Landick R. *Genes Dev.* 2012; 26:2621. [PubMed: 23207917]
24. Selby CP, Sancar A. *Science.* 1993; 260:53. [PubMed: 8465200]
25. Brow DA. *Mol Cell.* 2011; 42:717. [PubMed: 21700218]
26. Gaillard H, Aguilera A. *Biochim Biophys Acta.* 2013; 1829:141. [PubMed: 23046879]
27. Hartzog GA, Fu J. *Biochim Biophys Acta.* 2012
28. Hartzog GA, Kaplan CD. *Mol Cell.* 2011; 43:161. [PubMed: 21777806]
29. Sevostyanova A, Belogurov GA, Mooney RA, Landick R, Artsimovitch I. *Mol Cell.* 2011; 43:253. [PubMed: 21777814]
30. Grohmann D, Nagy J, Chakraborty A, Klose D, Fielden D, Ebright RH, Michaelis J, Werner F. *Mol Cell.* 2011; 43:263. [PubMed: 21777815]
31. Burmann BM, Knauer SH, Sevostyanova A, Schweimer K, Mooney RA, Landick R, Artsimovitch I, Rosch P. *Cell.* 2012; 150:291. [PubMed: 22817892]
32. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rosch P. *Science.* 2010; 328:501. [PubMed: 20413501]
33. Lindstrom DL, Squazzo SL, Muster N, Burckin TA, Wachter KC, Emigh CA, McCleery JA, Yates JR 3rd, Hartzog GA. *Mol Cell Biol.* 2003; 23:1368. [PubMed: 12556496]
34. Belogurov GA, Sevostyanova A, Svetlov V, Artsimovitch I. *Mol Microbiol.* 2010; 76:286. [PubMed: 20132437]
35. Belogurov GA, Vassilyeva MN, Svetlov V, Klyuyev S, Grishin NV, Vassilyev DG, Artsimovitch I. *Mol Cell.* 2007; 26:117. [PubMed: 17434131]
36. Svetlov V, Nudler E. *Embo J.* 2011; 30:1190. [PubMed: 21468097]
37. Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, Landick R. *Mol Cell.* 2009; 33:97. [PubMed: 19150431]
38. Belogurov GA, Mooney RA, Svetlov V, Landick R, Artsimovitch I. *Embo J.* 2009; 28:112. [PubMed: 19096362]
39. Bailey MJ, Hughes C, Koronakis V. *Mol Microbiol.* 1997; 26:845. [PubMed: 9426123]
40. Knauer SH, Artsimovitch I, Rosch P. *Cell Cycle.* 2012; 11:4289. [PubMed: 23095672]
41. Landick R. *Biochem Soc Trans.* 2006; 34:1062. [PubMed: 17073751]
42. Peters JM, Vangeloff AD, Landick R. *J Mol Biol.* 2011; 412:793. [PubMed: 21439297]
43. Vassilyev DG. *Curr Opin Struct Biol.* 2009; 19:691. [PubMed: 19896365]
44. Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. *Cell.* 1999; 98:811. [PubMed: 10499798]
45. Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. *Science.* 2001; 292:1876. [PubMed: 11313499]
46. Tagami S, Sekine SI, Yokoyama S. *Transcription.* 2011; 2:162. [PubMed: 21922057]
47. Vassilyev DG, Vassilyeva MN, Perederina A, Tahirov TH, Artsimovitch I. *Nature.* 2007; 448:157. [PubMed: 17581590]
48. Landick R. *Cell.* 2001; 105:567. [PubMed: 11389826]
49. Tagami S, Sekine S, Kumarevel T, Hino N, Murayama Y, Kamegamori S, Yamamoto M, Sakamoto K, Yokoyama S. *Nature.* 2010; 468:978. [PubMed: 21124318]
50. Chakraborty A, Wang D, Ebright YW, Korlann Y, Kortkhonjia E, Kim T, Chowdhury S, Wigneshweraraj S, Irschik H, Jansen R, Nixon BT, Knight J, Weiss S, Ebright RH. *Science.* 2012; 337:591. [PubMed: 22859489]
51. Artsimovitch I, Landick R. *Proc Natl Acad Sci U S A.* 2000; 97:7090. [PubMed: 10860976]
52. Steitz TA. *Nature.* 1998; 391:231. [PubMed: 9440683]
53. Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM. *Nature.* 2005; 438:460. [PubMed: 16284617]
54. Vassilyev DG, Vassilyeva MN, Zhang J, Palangat M, Artsimovitch I, Landick R. *Nature.* 2007; 448:163. [PubMed: 17581591]

55. Richardson, J.; Greenblatt, J. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. Neidhardt, F.; Curtiss, R., III; Ingraham, J.; Lin, E.; Low, K.; Magasanik, B.; Reznikoff, W.; Riley, M.; Schaechter, M.; Umberger, H., editors. Vol. 1. ASM Press; Washington, D. C: 1996.
56. Neuman KC, Abbondanzieri EA, Landick R, Gelles J, Block SM. *Cell*. 2003; 115:437. [PubMed: 14622598]
57. Herbert KM, La Porta A, Wong BJ, Mooney RA, Neuman KC, Landick R, Block SM. *Cell*. 2006; 125:1083. [PubMed: 16777599]
58. Touloukhonov I, Zhang J, Palangat M, Landick R. *Mol Cell*. 2007; 27:406. [PubMed: 17679091]
59. Nudler E. *Cell*. 2012; 149:1438. [PubMed: 22726433]
60. Sydow JF, Brueckner F, Cheung AC, Damsma GE, Dengl S, Lehmann E, Vassylyev D, Cramer P. *Mol Cell*. 2009; 34:710. [PubMed: 19560423]
61. Yuzenkova Y, Tadigotla VR, Severinov K, Zenkin N. *Embo J*. 2011; 30:3766. [PubMed: 21792175]
62. Bochkareva A, Yuzenkova Y, Tadigotla VR, Zenkin N. *Embo J*. 2012; 31:630. [PubMed: 22124324]
63. Kireeva ML, Kashlev M. *Proc Natl Acad Sci U S A*. 2009; 106:8900. [PubMed: 19416863]
64. Artsimovitch I, Landick R. *Cell*. 2002; 109:193. [PubMed: 12007406]
65. Sevostyanova A, Svetlov V, Vassylyev DG, Artsimovitch I. *Proc Natl Acad Sci U S A*. 2008; 105:865. [PubMed: 18195372]
66. Arndt KM, Chamberlin MJ. *J Mol Biol*. 1990; 213:79. [PubMed: 1692594]
67. von Hippel PH, Yager TD. *Proc Natl Acad Sci U S A*. 1991; 88:2307. [PubMed: 1706521]
68. Greive SJ, von Hippel PH. *Nat Rev Mol Cell Biol*. 2005; 6:221. [PubMed: 15714199]
69. Sidorenkov I, Komissarova N, Kashlev M. *Mol Cell*. 1998; 2:55. [PubMed: 9702191]
70. Richardson J. *Biochim Biophys Acta*. 2002; 1577:251. [PubMed: 12213656]
71. Peters JM, Mooney RA, Kuan PF, Rowland JL, Keles S, Landick R. *Proc Natl Acad Sci U S A*. 2009; 106:15406. [PubMed: 19706412]
72. Richardson JP, Grimley C, Lowery C. *Proc Natl Acad Sci U S A*. 1975; 72:1725. [PubMed: 1098042]
73. Harinarayanan R, Gowrishankar J. *J Mol Biol*. 2003; 332:31. [PubMed: 12946345]
74. Klumpp S, Hwa T. *Proc Natl Acad Sci U S A*. 2008; 105:18159. [PubMed: 19017803]
75. Epshtein V, Cardinale CJ, Ruckenstein AE, Borukhov S, Nudler E. *Mol Cell*. 2007; 28:991. [PubMed: 18158897]
76. Epshtein V, Dutta D, Wade J, Nudler E. *Nature*. 2010; 463:245. [PubMed: 20075920]
77. Santangelo TJ, Roberts JW. *Mol Cell*. 2004; 14:117. [PubMed: 15068808]
78. Santangelo TJ, Artsimovitch I. *Nat Rev Microbiol*. 2011; 9:319. [PubMed: 21478900]
79. Grundy FJ, Winkler WC, Henkin TM. *Proc Natl Acad Sci U S A*. 2002; 99:11121. [PubMed: 12165569]
80. Weisberg RA, Gottesman ME. *J Bacteriol*. 1999; 181:359. [PubMed: 9882646]
81. Rees W, Weitzel S, Yager T, Das A, von Hippel P. *Proc Natl Acad Sci U S A*. 1996; 93:342. [PubMed: 8552635]
82. Torres M, Condon C, Balada JM, Squires C, Squires CL. *Embo J*. 2001; 20:3811. [PubMed: 11447122]
83. Gusarov I, Nudler E. *Mol Cell*. 1999; 3:495. [PubMed: 10230402]
84. Yarnell WS, Roberts JW. *Science*. 1999; 284:611. [PubMed: 10213678]
85. Gusarov I, Nudler E. *Cell*. 2001; 107:437. [PubMed: 11719185]
86. Shankar S, Hatoum A, Roberts JW. *Mol Cell*. 2007; 27:914. [PubMed: 17889665]
87. Li J, Mason SW, Greenblatt J. *Genes Dev*. 1993; 7:161. [PubMed: 8422985]
88. Downing WL, Sullivan SL, Gottesman ME, Dennis PP. *J Bacteriol*. 1990; 172:1621. [PubMed: 2137819]
89. Sullivan SL, Ward DF, Gottesman ME. *J Bacteriol*. 1992; 174:1339. [PubMed: 1531224]
90. Li J, Horwitz R, McCracken S, Greenblatt J. *J Biol Chem*. 1992; 267:6012. [PubMed: 1532577]

91. Mason SW, Li J, Greenblatt J. *J Biol Chem.* 1992; 267:19418. [PubMed: 1388170]
92. Mooney RA, Schweimer K, Rosch P, Gottesman M, Landick R. *J Mol Biol.* 2009; 391:341. [PubMed: 19500594]
93. Burova E, Hung SC, Sagitov V, Stitt BL, Gottesman ME. *J Bacteriol.* 1995; 177:1388. [PubMed: 7868616]
94. Zellars M, Squires CL. *Mol Microbiol.* 1999; 32:1296. [PubMed: 10383769]
95. Torres M, Balada JM, Zellars M, Squires C, Squires CL. *J Bacteriol.* 2004; 186:1304. [PubMed: 14973028]
96. Hung SC, Gottesman ME. *J Mol Biol.* 1995; 247:428. [PubMed: 7714899]
97. Krishna Leela J, Syeda AH, Anupama K, Gowrishankar J. *Proc Natl Acad Sci U S A.* 2013; 110:258. [PubMed: 23251031]
98. Friedman DI, Court DL. *Mol Microbiol.* 1995; 18:191. [PubMed: 8709839]
99. Proshkin S, Rahmouni AR, Mironov A, Nudler E. *Science.* 2010; 328:504. [PubMed: 20413502]
100. Nodwell JR, Greenblatt J. *Genes Dev.* 1991; 5:2141. [PubMed: 1834523]
101. Nodwell JR, Greenblatt J. *Cell.* 1993; 72:261. [PubMed: 7678781]
102. Mason SW, Greenblatt J. *Genes Dev.* 1991; 5:1504. [PubMed: 1831176]
103. Condon C, Squires C, Squires CL. *Microbiol Rev.* 1995; 59:623. [PubMed: 8531889]
104. Squires CL, Greenblatt J, Li J, Condon C. *Proc Natl Acad Sci U S A.* 1993; 90:970. [PubMed: 8430111]
105. Albrechtsen B, Ross BM, Squires C, Squires CL. *Nucleic Acids Res.* 1991; 19:1845. [PubMed: 1709493]
106. Burns CM, Richardson LV, Richardson JP. *J Mol Biol.* 1998; 278:307. [PubMed: 9571053]
107. Bar-Nahum G, Epshtein V, Ruckenstein AE, Rafikov R, Mustaev A, Nudler E. *Cell.* 2005; 120:183. [PubMed: 15680325]
108. Pasman Z, von Hippel PH. *Biochemistry.* 2000; 39:5573. [PubMed: 10820031]
109. Svetlov V, Belogurov GA, Shabrova E, Vassilyev DG, Artsimovitch I. *Nucleic Acids Res.* 2007; 35:5694. [PubMed: 17711918]
110. Herbert KM, Zhou J, Mooney RA, Porta AL, Landick R, Block SM. *J Mol Biol.* 2010; 399:17. [PubMed: 20381500]
111. Sevostyanova A, Artsimovitch I. *Nucleic Acids Res.* 2010; 38:7432. [PubMed: 20639538]
112. Yakhnin AV, Babitzke P. *Mol Microbiol.* 2010; 76:690. [PubMed: 20384694]
113. Sullivan SL, Gottesman ME. *Cell.* 1992; 68:989. [PubMed: 1547498]
114. Burns CM, Richardson JP. *Proc Natl Acad Sci U S A.* 1995; 92:4738. [PubMed: 7761393]
115. Nehrke KW, Zalatan F, Platt T. *Gene Expr.* 1993; 3:119. [PubMed: 7505669]
116. Nehrke KW, Platt T. *J Mol Biol.* 1994; 243:830. [PubMed: 7525972]
117. Chalissery J, Muteeb G, Kalarickal NC, Mohan S, Jisha V, Sen R. *J Mol Biol.* 2011; 405:49. [PubMed: 21040729]
118. Kalyani BS, Muteeb G, Qayyum MZ, Sen R. *J Mol Biol.* 2011; 413:548. [PubMed: 21920369]
119. Kyripides NC, Woese CR, Ouzounis CA. *Trends Biochem Sci.* 1996; 21:425. [PubMed: 8987397]
120. Ramakrishnan V. *Cell.* 2002; 108:557. [PubMed: 11909526]
121. Wilkinson RG, Stocker BA. *Nature.* 1968; 217:955. [PubMed: 4868316]
122. Lindberg AA, Hellerqvist CG. *J Gen Microbiol.* 1980; 116:25. [PubMed: 6988545]
123. Beutin L, Achtman M. *J Bacteriol.* 1979; 139:730. [PubMed: 383690]
124. Bailey MJ, Koronakis V, Schmoll T, Hughes C. *Mol Microbiol.* 1992; 6:1003. [PubMed: 1584020]
125. Beutin L, Manning PA, Achtman M, Willetts N. *J Bacteriol.* 1981; 145:840. [PubMed: 7007352]
126. Landraud L, Gibert M, Popoff MR, Boquet P, Gauthier M. *Mol Microbiol.* 2003; 47:1653. [PubMed: 12622819]
127. Clarke BR, Pearce R, Roberts IS. *J Bacteriol.* 1999; 181:2279. [PubMed: 10094710]
128. Stevens MP, Clarke BR, Roberts IS. *Mol Microbiol.* 1997; 24:1001. [PubMed: 9220007]

129. Marolda CL, Valvano MA. *J Bacteriol.* 1998; 180:3070. [PubMed: 9620955]
130. Wang L, Jensen S, Hallman R, Reeves PR. *FEMS Microbiol Lett.* 1998; 165:201. [PubMed: 9711858]
131. Nagy G, Dobrindt U, Kupfer M, Emody L, Karch H, Hacker J. *Infect Immun.* 2001; 69:1924. [PubMed: 11179376]
132. Beloin C, Michaelis K, Lindner K, Landini P, Hacker J, Ghigo JM, Dobrindt U. *J Bacteriol.* 2006; 188:1316. [PubMed: 16452414]
133. Nagy G, Dobrindt U, Schneider G, Khan AS, Hacker J, Emody L. *Infect Immun.* 2002; 70:4406. [PubMed: 12117951]
134. Nagy G, Danino V, Dobrindt U, Pallen M, Chaudhuri R, Emody L, Hinton JC, Hacker J. *Infect Immun.* 2006; 74:5914. [PubMed: 16988271]
135. Bittner M, Saldias S, Altamirano F, Valvano MA, Contreras I. *Microb Pathog.* 2004; 36:19. [PubMed: 14643636]
136. Nieto JM, Bailey MJ, Hughes C, Koronakis V. *Mol Microbiol.* 1996; 19:705. [PubMed: 8820641]
137. Bailey MJ, Hughes C, Koronakis V. *Mol Microbiol.* 1996; 22:729. [PubMed: 8951819]
138. Hobbs M, Reeves PR. *Mol Microbiol.* 1994; 12:855. [PubMed: 8052136]
139. Arco Y, Llamas I, Martinez-Checa F, Argandona M, Quesada E, del Moral A. *Microbiology.* 2005; 151:2841. [PubMed: 16151197]
140. Carter HD, Svetlov V, Artsimovitch I. *J Bacteriol.* 2004; 186:2829. [PubMed: 15090525]
141. Bailey MJ, Hughes C, Koronakis V. *Mol Gen Genet.* 2000; 262:1052. [PubMed: 10660066]
142. Leeds JA, Welch RA. *J Bacteriol.* 1997; 179:3519. [PubMed: 9171395]
143. Chaudhuri J, Khuong C, Alt FW. *Nature.* 2004; 430:992. [PubMed: 15273694]
144. Ponting CP. *Nucleic Acids Res.* 2002; 30:3643. [PubMed: 12202748]
145. Steiner T, Kaiser JT, Marinkovic S, Huber R, Wahl MC. *Embo J.* 2002; 21:4641. [PubMed: 12198166]
146. Viktorovskaya OV, Appling FD, Schneider DA. *J Biol Chem.* 2011; 286:18825. [PubMed: 21467036]
147. Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. *Embo J.* 2011; 30:1302. [PubMed: 21386817]
148. Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F, Buratowski S, Handa H. *Genes Dev.* 1998; 12:343. [PubMed: 9450929]
149. Zhou H, Liu Q, Gao Y, Teng M, Niu L. *Proteins.* 2009; 76:787. [PubMed: 19475703]
150. Guo M, Xu F, Yamada J, Egelhofer T, Gao Y, Hartzog GA, Teng M, Niu L. *Structure.* 2008; 16:1649. [PubMed: 19000817]
151. Knowlton JR, Bubunenko M, Andrykovitch M, Guo W, Routzahn KM, Waugh DS, Court DL, Ji X. *Biochemistry.* 2003; 42:2275. [PubMed: 12600194]
152. Reay P, Yamasaki K, Terada T, Kuramitsu S, Shirouzu M, Yokoyama S. *Proteins.* 2004; 56:40. [PubMed: 15162485]
153. Liao D, Lurz R, Dobrinski B, Dennis PP. *J Bacteriol.* 1996; 178:4089. [PubMed: 8763936]
154. Mayer A, Schreieck A, Lidschreiber M, Leike K, Martin DE, Cramer P. *Mol Cell Biol.* 2012; 32:1321. [PubMed: 22290438]
155. Burmann BM, Scheckenhofer U, Schweimer K, Rosch P. *Biochem J.* 2011; 435:783. [PubMed: 21345171]
156. Nickels BE. *Methods.* 2009; 47:53. [PubMed: 18952173]
157. Burova E, Hung SC, Chen J, Court DL, Zhou JG, Mogilnitskiy G, Gottesman ME. *Mol Microbiol.* 1999; 31:1783. [PubMed: 10209750]
158. Norman A, Hansen LH, She Q, Sorensen SJ. *Plasmid.* 2008; 60:59. [PubMed: 18440636]
159. Hurst MR, Glare TR, Jackson TA. *J Bacteriol.* 2004; 186:5116. [PubMed: 15262948]
160. Hurst MR, Beard SS, Jackson TA, Jones SM. *FEMS Microbiol Lett.* 2007; 270:42. [PubMed: 17263838]
161. Paitan Y, Orr E, Ron EZ, Rosenberg E. *FEMS Microbiol Lett.* 1999; 170:221. [PubMed: 9919671]

162. Chatzidaki-Livanis M, Weinacht KG, Comstock LE. *Proc Natl Acad Sci U S A*. 2010; 107:11976. [PubMed: 20547868]
163. Chatzidaki-Livanis M, Coyne MJ, Comstock LE. *J Bacteriol*. 2009; 191:7288. [PubMed: 19801412]
164. Ingham CJ, Dennis J, Furneaux PA. *Mol Microbiol*. 1999; 31:651. [PubMed: 10027981]
165. Xia M, Lunsford RD, McDevitt D, Iordanescu S. *Plasmid*. 1999; 42:144. [PubMed: 10489332]
166. Yakhnin AV, Yakhnin H, Babitzke P. *Proc Natl Acad Sci U S A*. 2008; 105:16131. [PubMed: 18852477]
167. Liu J, Pei H, Mei S, Li J, Zhou L, Xiang H. *Biochem Biophys Res Commun*. 2008; 371:573. [PubMed: 18457667]
168. Lawrence JG, Roth JR. *Genetics*. 1996; 143:1843. [PubMed: 8844169]
169. Farewell A, Brazas R, Davie E, Mason J, Rothfield LI. *J Bacteriol*. 1991; 173:5188. [PubMed: 1860828]
170. Morgan WD, Bear DG, von Hippel PH. *J Biol Chem*. 1984; 259:8664. [PubMed: 6330119]
171. Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS. *Proc Natl Acad Sci U S A*. 2011; 108:546. [PubMed: 21187417]
172. Milon P, Rodnina MV. *Crit Rev Biochem Mol Biol*. 2012; X
173. Simic R, Lindstrom DL, Tran HG, Roinick KL, Costa PJ, Johnson AD, Hartzog GA, Arndt KM. *Embo J*. 2003; 22:1846. [PubMed: 12682017]
174. Squazzo SL, Costa PJ, Lindstrom DL, Kumer KE, Simic R, Jennings JL, Link AJ, Arndt KM, Hartzog GA. *Embo J*. 2002; 21:1764. [PubMed: 11927560]
175. Kireeva ML, Hancock B, Cremona GH, Walter W, Studitsky VM, Kashlev M. *Mol Cell*. 2005; 18:97. [PubMed: 15808512]
176. Andrulis ED, Werner J, Nazarian A, Erdjument-Bromage H, Tempst P, Lis JT. *Nature*. 2002; 420:837. [PubMed: 12490954]
177. French SL, Santangelo TJ, Beyer AL, Reeve JN. *Molecular biology and evolution*. 2007; 24:893. [PubMed: 17237472]
178. Dahan O, Gingold H, Pilpel Y. *Trends in genetics: TIG*. 2011; 27:316. [PubMed: 21763027]
179. Nevo-Dinur K, Nussbaum-Shochat A, Ben-Yehuda S, Amster-Choder O. *Science*. 2011; 331:1081. [PubMed: 21350180]
180. Gowrishankar J, Harinarayanan R. *Mol Microbiol*. 2004; 54:598. [PubMed: 15491353]
181. Yanofsky C. *Nature*. 1981; 289:751. [PubMed: 7007895]
182. Squires CL, Zaporozets D. *Annu Rev Microbiol*. 2000; 54:775. [PubMed: 11018144]
183. Warner JR, McIntosh KB. *Mol Cell*. 2009; 34:3. [PubMed: 19362532]
184. Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, Volkman BF. *Proc Natl Acad Sci U S A*. 2008; 105:5057. [PubMed: 18364395]
185. Copley SD. *Bioessays*. 2012; 34:578. [PubMed: 22696112]
186. Murzin AG. *Science*. 2008; 320:1725. [PubMed: 18583598]
187. Saitou N, Nei M. *Molecular biology and evolution*. 1987; 4:406. [PubMed: 3447015]
188. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. *Molecular biology and evolution*. 2011; 28:2731. [PubMed: 21546353]

Biographies

Sushil Kumar Tomar studied pharmacy for his undergraduate studies from Rajiv Gandhi university of Health Sciences, Karnataka, India. He completed *Bachelor of pharmacy* in 2004. Being fascinated to basic science, later he moved to Indian Institute of Technology, Kanpur and received PhD in *Structural and Molecular Biology* in 2010. There, he worked with Dr. Balaji Prakash to understand the role of bacterial GTPase in the process of ribosome assembly. Since 2011, he is working with Dr. Irina Artsimovitch at The RNA Center, The Ohio State University, Columbus (Ohio), USA. His current research focus is transcription elongation factor, RfaH.

Irina Artsimovitch completed her M. Biochemistry (Hons) at Moscow State University (Russia) in 1989. She received her Ph.D. in Microbiology from the University of Tennessee-Memphis in 1996, working on regulation of transcription in bacteriophage Mu under the supervision of Martha M. Howe. Following postdoctoral work with Robert Landick at the University of Wisconsin-Madison (1996–2001) aimed to elucidate the molecular mechanism of transcriptional pausing, she joined the Department of Microbiology at the Ohio State University as an Assistant Professor in 2001, was promoted to an Associate Professor in 2006, and to a Professor in 2011. Her lab studies molecular mechanism and regulation of bacterial RNA polymerases as well as cross-talk between transcription, DNA repair, and protein synthesis in *Bacteria*.

Figure 1.

The main features of the TEC. Bacterial core RNAP (a pentasubunit $\alpha_2\beta\beta'\omega$ complex) clamps around the nucleic acid chains using pincers composed of the mobile domains of the β and β' subunits. The active site contains a catalytic Mg^{2+} ion and is composed of i and $i+1$ sub-sites. The RNA:DNA hybrid and the downstream DNA are buried in the RNAP, and only a short segment of the nontemplate DNA is exposed on the surface.

Figure 2.

Transcription cycle and isomerization into paused and termination states. The DNA template loads into the holoenzyme with the clamp open to form a closed promoter complex (not shown); following the strand separation, the clamp closes (in the open complex) and remains closed throughout elongation (in the TEC). At some sites, the clamp may partially open, facilitating rearrangements into the elemental pause state, which further isomerizes into pause and termination complexes.

Figure 3. The branched mechanism of RNA chain elongation suggests several points of antiterminator (AT) action.

Figure 4.
E. coli NusG interacts with RNAP via its NTD (in green) and with NusE (S10), Rho, and Nun via the CTD (in cyan) to control transcription and translation.

Figure 5.

The ribosomal protein S10 simultaneously interacts with NusB and NusG CTD. A structural superposition generated from NusE:NusB (PDB: 3D3B) and NusG CTD:S10 (PDB: 2KVQ) complexes.

Figure 6.

RfaH-controlled *rfa* and *rfb* operons; promoters are indicated by bent arrows, the length (in kb) is shown below. Relative occupancies of RNAP (gray), Rho (light blue) and RfaH (red) along the operon are plotted below. Vertical arrows indicate the positions of *ops* elements; the number in parentheses corresponds to the distance relative to the nearest ORF.

Figure 7.

Domain organization and phylogeny of NusG proteins. (a) Domain organization in bacterial, archaeal and eukaryotic NusG proteins is depicted. (b) A simple phylogenetic tree, inferred using Neighbor-Joining method¹⁸⁷ generated in MEGA5¹⁸⁸ program illustrates an evolutionary history of NusG family.

Figure 8.

Structures of (a) *E. coli* NusG (PDB: 2KO6, 2JVV); (b) *A. aeolicus* NusG (PDB: 2KO6, 2JVV); (c) *E. coli* RfaH (PDB: 2OUG) and (d) *Pyrococcus furiosus* Spt5-Spt4 (PDB: 3P8B). (e) A structural superposition of NGN domains including human DSIF (PDB: 3H7H) reveals high structural conservation among these proteins.

Figure 9.

RfaH interactions with the TEC. (a) A model of RfaH NTD (green) bound to the TEC. The RNAP is shown as a gray surface; the β' bridge helix, β' CH and β GL are shown as cartoons. The RNAP active site is marked by the Mg^{2+} ion (magenta sphere). The template DNA, the nontemplate DNA, and the RNA are colored in black, blue, and red, respectively. (b) A close-up view with the side chains making the key interactions with the nontemplate DNA, β' CH and β GL shown as surfaces and colored in blue, orange and magenta, respectively.

Figure 10. Functional interactions of the NusG CTD. Residues whose substitutions confer defects in Nun function are shown in red (F144, N145 and F165), those that confer defects in Rho function - in blue (V148, L158, V160 and I164).

Figure 11.

NusG and RfaH regulons. NusG exists in an open, active state and can bind to RNAP at all sites. Once bound to the β' CH domain, the NusG CTD interacts with Rho. In contrast, RfaH requires domain dissociation to convert a closed, autoinhibited state into an open state capable of interactions with RNAP.

Figure 12.

The universally conserved clamping action of RfaH NTD. This model was built based on the binary RNAP/NusG complex ¹⁴⁷, with *E. coli* RfaH NTD replacing NusG NTD. The NGN domain anchored by the β' CH domain reaches across the active site cleft towards the β lobe domain, effectively sealing the double-stranded DNA within the closed cleft.

Figure 13.

Anti-Rho activities of RfaH. (a) RfaH excludes NusG from the TEC, inhibiting Rho-mediated RNA release but not Rho binding. (b) An RfaH-tethered ribosome blocks Rho binding to the nascent RNA.

Figure 14.

RfaH transformation. RfaH recruitment to RNAP at *ops* leads to domain dissociation. This triggers refolding of the CTD. The refolded CTD is now capable of coupling transcription to translation by binding to S10.