The EMBO Journal (2009) 28, 112–122 | © 2009 European Molecular Biology Organization | All Rights Reserved 0261-4189/09
www.embojournal.org

THE EMBO JOURNAL

Functional specialization of transcription elongation factors

Georgiy A Belogurov[1], Rachel A Mooney[2], Vladimir Svetlov[1,3], Robert Landick[2] and Irina Artsimovitch[1,*]

[1]Department of Microbiology and The RNA Group, The Ohio State University, Columbus, OH, USA and [2]Department of Biochemistry, University of Wisconsin, Madison, WI, USA

Elongation factors NusG and RfaH evolved from a common ancestor and utilize the same binding site on RNA polymerase (RNAP) to modulate transcription. However, although NusG associates with RNAP transcribing most *Escherichia coli* genes, RfaH regulates just a few operons containing *ops*, a DNA sequence that mediates RfaH recruitment. Here, we describe the mechanism by which this specificity is maintained. We observe that RfaH action is indeed restricted to those several operons that are devoid of NusG *in vivo*. We also show that RfaH and NusG compete for their effects on transcript elongation and termination *in vitro*. Our data argue that RfaH recognizes its DNA target even in the presence of NusG. Once recruited, RfaH remains stably associated with RNAP, thereby precluding NusG binding. We envision a pathway by which a specialized regulator has evolved in the background of its ubiquitous paralogue. We propose that RfaH and NusG may have opposite regulatory functions: although NusG appears to function in concert with Rho, RfaH inhibits Rho action and activates the expression of poorly translated, frequently foreign genes.

The EMBO Journal (2009) **28,** 112–122. doi:10.1038/emboj.2008.268; Published online 18 December 2008
Subject Categories: chromatin & transcription; microbiology & pathogens
Keywords: NusG; paralogue; RfaH; Rho; RNA polymerase

Introduction

Throughout evolution, bacterial gene expression networks had to maintain basic housekeeping functions while inventing new regulatory circuits to allow access to new ecological niches or utilize new resources. One way to expand the repertoire of regulators is through duplication of already existing ones, with subsequent specialization. The hypothetical origin of virulence regulator RfaH is consistent with this model: *rfaH* apparently arose through duplication of the gene that encodes the widely conserved regulator NusG (Bailey *et al*, 1997). NusG is essential in wild-type *Escherichia coli* (Sullivan and Gottesman, 1992) and associates with RNA polymerase (RNAP) transcribing nearly all *E. coli* genes (Mooney *et al*, 2009). In contrast, RfaH is a non-essential protein, the action of which is restricted to a handful of operons containing an *ops* signal in their transcribed DNA. This *ops* element induces RNAP pausing and mediates RfaH recruitment (Artsimovitch and Landick, 2002).

Both RfaH and NusG increase the apparent transcript elongation rate of *E. coli* RNAP *in vitro*. However, they differ in their response to Rho-dependent terminators, recognition of nucleic acid components in the transcription elongation complex (TEC), and regulatory targets (Figure 1A). These similarities and differences are reflected in their structures: both proteins exhibit a two-domain architecture in which the N-terminal domains (NTDs) are quite similar but the C-terminal domains (CTDs) are strikingly different (Figure 1B). The interdomain interactions are also different—in NusG the two domains are separated by a linker, and in RfaH they are tightly associated to bury a large nonpolar surface on the NTD. We proposed that the buried surface constitutes the RNAP-binding site and showed that RfaH requires domain dissociation to become active (Belogurov *et al*, 2007). Dissociation is thought to be triggered by interactions with the *ops* element and is required for the stable binding of RfaH to its target site on RNAP, the β′ clamp helices (β′ CH).

This domain organization suggests that the common properties of NusG and RfaH (binding to, and acceleration of, the RNAP) are mediated by their structurally similar NTDs, whereas the CTDs may confer protein-specific functions. In NusG, the CTD most likely mediates interaction with Rho (Li *et al*, 1993) that promotes termination (Sullivan and Gottesman, 1992; Pasman and von Hippel, 2000), possibly the essential role of NusG in *E. coli* (Cardinale *et al*, 2008). In RfaH, the CTD indirectly confers sequence specificity (Belogurov *et al*, 2007) and may interact with the translation/secretion machineries (Bailey *et al*, 2000); however, RfaH does not bind to Rho directly (IA, unpublished data).

Molecular modelling indicates that the putative RNAP-binding site on the NTD is conserved between RfaH and NusG: thus, they should compete for binding to the TEC, yet both should be able to target their respective genes in the cell. In this study, we used *in vitro* transcription and ChIP-on-chip assays to study how *E. coli* RfaH and NusG maintain their separate regulatory niches, and phylogenetic analysis to trace transformation of a general transcription factor into a highly specialized, sequence-specific regulator.

Results

RfaH competes with NusG during Rho-dependent termination

NusG increases Rho-dependent termination at a subset of sites (Sullivan and Gottesman, 1992; Nehrke and Platt, 1994)

*Corresponding author. Department of Microbiology, The Ohio State University, 376 Biosciences Building, 484 W 12th Avenue, Columbus, OH 43210, USA. Tel.: +1 614 292 6777; Fax: +1 614 292 8120; E-mail: artsimovitch.1@osu.edu
[3]Present address: Department of Biochemistry, New York University School of Medicine, New York, NY 10016, USA

Received: 4 September 2008; accepted: 26 November 2008; published online: 18 December 2008

A

	RfaH	NusG
Pausing at hairpin sites	Reduces	Does not affect
Pausing at backtracked sites	Reduces	Reduces
Transcript elongation rate	Increases	Increases
Rho-independent termination	Reduces at some sites	Reduces at some sites
Rho-dependent termination	Reduces at some sites	Increases at most sites
Sequence-specific recruitment	At *ops* sites	None apparent
Targets in the transcription complex	β′ clamp helices non-template DNA strand	β′ clamp helices Rho, nascent RNA?
Regulatory targets	Virulence and *tra* operons LPS and capsules	*rrn* operons and many others

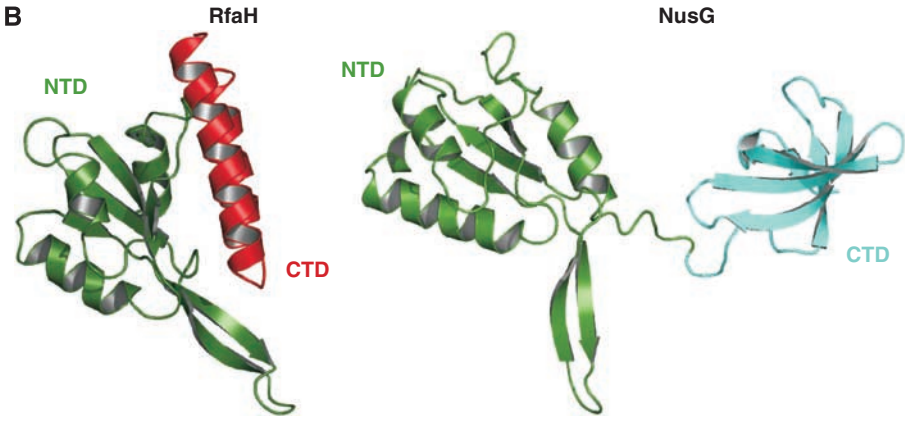

Figure 1 RfaH and NusG: similarities and differences. (**A**) Comparison of the functional properties. RfaH and NusG likely bind to the same site on RNAP (Belogurov *et al*, 2007 and unpublished data), the β′ CH and increase the transcript elongation rate. In contrast to RfaH, NusG *increases* Rho-dependent termination, presumably by stabilizing a quarternary Rho–TEC complex (Nehrke and Platt, 1994). In addition, NusG does not affect RNAP paused at the hairpin-dependent pause sites, whereas RfaH facilitates transcription similarly through both the hairpin-dependent and -independent signals (Artsimovitch and Landick, 2000, 2002). NusG also participates in the formation of multi-component transcription antitermination complexes (Mason *et al*, 1992; Squires *et al*, 1993) where it may make specific contacts to several transcription factors. Finally, NusG does not exhibit any DNA sequence specificity and does not bind to the *ops*-paused TECs *in vitro*. (**B**) Structures of the *E. coli* RfaH (Belogurov *et al*, 2007) and a model of the *E. coli* NusG (Steiner *et al*, 2002).

and it frequently shifts the window of Rho-released RNAs upstream; thus NusG appears to allow Rho to act earlier during transcription (Burns *et al*, 1999; Pasman and von Hippel, 2000). In contrast, RfaH modestly reduces Rho-dependent termination *in vivo* (Stevens *et al*, 1997) and *in vitro* (Artsimovitch and Landick, 2002). We reasoned that if (i) RfaH and NusG both bind to the β′ CH and (ii) RfaH remains bound to TEC downstream of the *ops* site, it should prevent the ability of NusG to enhance Rho-dependent termination.

To test this hypothesis, we examined Rho-dependent termination *in vitro* on a pIA267 template (Figure 2A) that encodes the *ops* site followed by a well-characterized phage λtR1 Rho-dependent termination signal (Lau and Roberts, 1985) that responds to both NusG (Sullivan and Gottesman, 1992) and RfaH (Artsimovitch and Landick, 2002). In the absence of accessory factors, 95% of the *E. coli* RNAP molecules reached the end of the linear DNA template, forming run-off transcripts (Figure 2B). Addition of Rho caused termination at multiple sites within the λ_{tR1} cluster, thereby reducing the run-off product to 8%. When NusG was also present, the overall efficiency of termination did not change dramatically, but the pattern of release products changed, with more RNAs released earlier. In particular, termination at one site, T*, was strongly enhanced by NusG (from 4 to 11%), but was not altered in the presence of RfaH. RfaH increased the run-off transcription modestly, less than two-fold (to 16% at high concentrations of RfaH), and when present in excess, eliminated NusG-dependent termination at the T* site. This result suggests that RfaH can compete with NusG for effects on Rho-dependent termination, and is consistent with the model that both proteins bind to the same site on the RNAP.

NusG may displace full-length RfaH from RNAP following recruitment

We next wanted to test whether RfaH and NusG compete for effects on transcriptional pausing. We used the pIA349

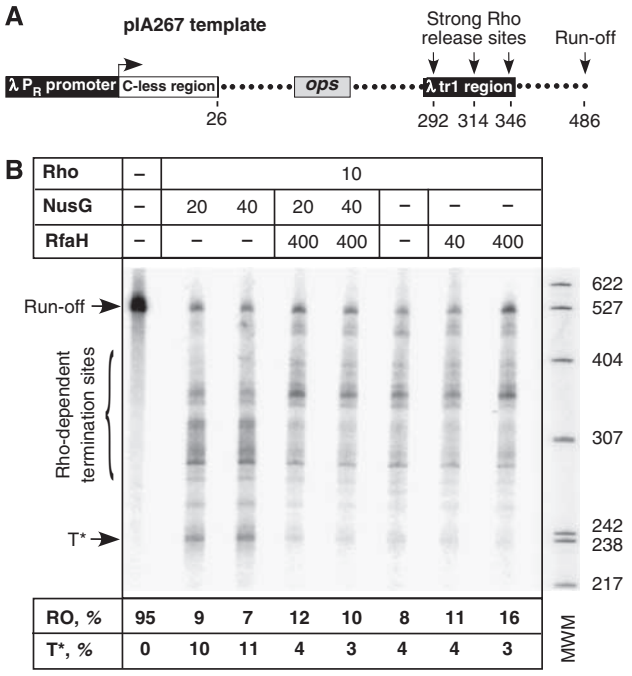

Figure 2 RfaH eliminates the NusG effect on Rho-dependent termination. (**A**) Transcript generated on a linear pIA267 DNA template; transcription start site (+1), *ops*, Rho-dependent RNA release sites, and transcript end are indicated. (**B**) Halted [α^{32}P]GMP TECs were formed at 40 nM with *E. coli* RNAP. Rho, RfaH, and NusG were added where indicated. Sizes of the [γ^{32}P]ATP pBR322 *Msp*I fragments used as molecular markers are indicated on the right. Positions of the run-off (RO) transcript and the NusG-enhanced termination site (T*) are shown by arrows; position of the Rho-dependent termination region is indicated by a bracket.

template (Artsimovitch and Landick, 2002) that encodes an *ops* pause (*opsP*) site followed by a hairpin-dependent *his* pause (*hisP*) site (Figure 3A). Previous analysis indicated that RfaH and NusG have distinct effects on pausing at these two sites: NusG decreases pausing at *opsP*, and does not affect RNAP transcribing through the *hisP* site (Artsimovitch and Landick, 2000), whereas RfaH decreases pausing at *hisP* site but delays RNAP at *opsP* through sequence-specific interactions with the non-template DNA strand (Artsimovitch and Landick, 2002). In a single-round elongation assay (Figure 3B) in the absence of factors, RNAP paused at the *opsP* site with a half-life of 15 s and at the *hisP* with a half-life of 56 s (Figure 3C); RNAP was also delayed at some sites between these two major pauses. Addition of equimolar full-length RfaH (50 nM) delayed TEC at the *opsP* site increasing the pause half-life to 200 s; this characteristic delay is a consequence of the RfaH binding to the TEC (but not a requirement for RfaH function, see Discussion). As observed earlier, addition of RfaH reduced RNAP pausing at the downstream *hisP* ($t_{1/2}=20$ s) and other pause sites, facilitating RNAP arrival at the end of the template. This overall increase in elongation rate is indicative of RfaH's ability to modify RNAP into a pause-resistant state (Svetlov *et al*, 2007).

NusG altered transcription in a distinct way: at 50 nM, it reduced pausing at the *opsP* site two-fold ($t_{1/2}=7$ s), but did not dramatically affect the recognition of the *hisP* site ($t_{1/2}=47$ s). When added together with the full-length RfaH, even a 10-fold excess of NusG reduced but did not eliminate the characteristic delay of RNAP by RfaH at the *opsP* site ($t_{1/2}=160$ s), but did counteract the RfaH effect at the *hisP* site ($t_{1/2}=27$ s); addition of NusG at 1:1 ratio produced an intermediate effect (Figure 3C).

To become recruited to the TEC in the cell, RfaH should load onto *ops*-containing DNA even in the presence of NusG. The results shown in Figure 2 demonstrate that this is the case: RfaH is apparently recruited to RNAP even in the presence of a 10-fold excess of NusG. However, as NusG recruitment to RNAP may be sensitive to the length of the nascent RNA (Mooney *et al*, 2009), we wanted to test whether moving the *ops* element 50 bp downstream would hinder RfaH recruitment. We found that RfaH was efficiently recruited to an *ops*-containing TEC with a 95-nt-long RNA whether or not NusG was present (Supplementary Figure 1). We also tested the effect of NusG on RfaH-mediated antitermination at the intrinsic (factor-independent) T_{hly} terminator (Carter *et al*, 2004). During single-round *in vitro* transcription in the absence of either factor, ~40% of transcripts were terminated at T_{hly} (Figure 3D), whereas addition of RfaH decreased termination efficiency more than two-fold, to 17%. In contrast, addition of 10-fold molar excess of NusG alone reduced termination efficiency at T_{hly} to 28% (similar to the efficiency of NusG antitermination at the transcriptional attenuator preceding the *rpoB* gene; Linn and Greenblatt, 1992) and also abrogated any additional antitermination by RfaH.

These results suggest that NusG does not block the RfaH recruitment at the *opsP* site but may interfere with RfaH or displace RfaH from the TEC at a later point. This observation is particularly notable because RfaH regulates the expression of very long operons (such as 8-kb *hly* or 30-kb *tra*) in the presence of competing NusG *in vivo*, and thus would be expected to maintain its contacts with RNAP throughout elongation. Although other models are possible, persistent association with the TEC is the most straightforward mechanism by which RfaH may switch RNAP into a fast state, and is used by other antiterminators (Rees *et al*, 1996; Roberts *et al*, 1998).

The RfaH NTD is stably bound to the TEC

The ability of RfaH to work *in vivo* despite the ability of NusG to compete for effect *in vitro* may be explained if RfaH is bound to the TEC more tightly *in vivo*. In fact, although binding to the *ops* element is apparently necessary for RNAP modification, it is not sufficient: *Vibrio cholerae* RfaH delays *E. coli* RNAP at the *opsP* site as efficiently as *E. coli* RfaH, but fails to function at downstream sites (Carter *et al*, 2004), suggesting that suboptimal interactions between heterologous RfaH and RNAP lead to dissociation of the TEC/RfaH complex. On release from the TEC, RfaH domains would re-establish their tight interdomain contacts, masking the RNAP-binding site and thereby precluding re-binding to the TEC at downstream positions (additional *ops* sites are not present at promoter distal sites in transcription units).

To test this idea, we used the isolated RfaH NTD in competition experiments *in vitro* (Figure 4). The NTD mediates all RfaH activities *in vitro* yet it is *ops*-independent—the physical removal of the CTD exposes the RNAP-binding site on RfaH, thereby obviating the requirement for *ops*-dependent isomerization (Belogurov *et al*, 2007). Thus, the NTD may rebind at any site even if it dissociates from the TEC spontaneously; consistently, it is more effective than the full-length RfaH in reducing factor-independent (Belogurov *et al*,

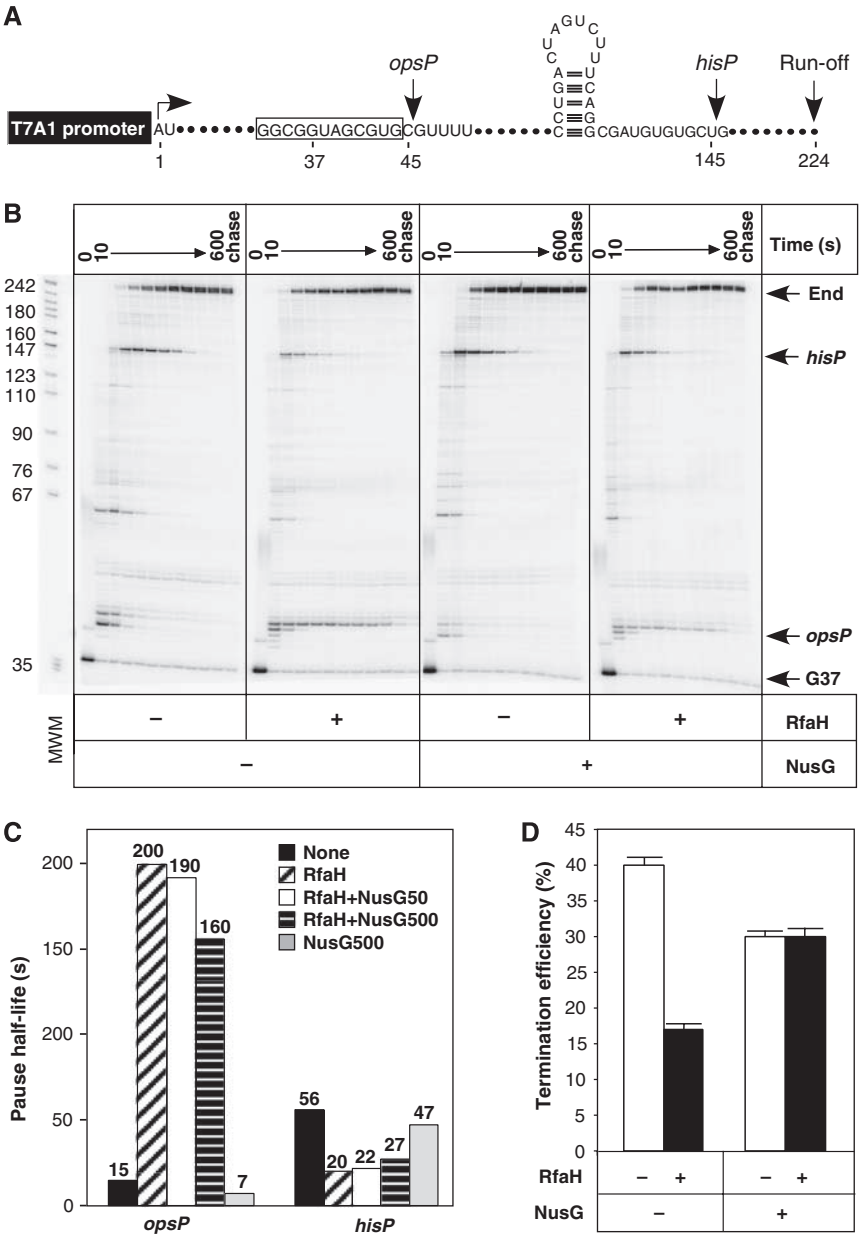

Figure 3 Interplay between RfaH and NusG during elongation *in vitro*. (**A**) Transcript generated from the T7A1 promoter on a linear pIA349 DNA template; start site (+1), *ops* and *his* pause sites, and RO are indicated. (**B**) Single-round pause assay. Halted [α^{32}P]CMP G37 TECs formed with *E. coli* RNAP were incubated with RfaH (50 nM) and/or NusG (500 nM), as indicated and transcription was restarted (see Materials and methods). Positions of the G37, *ops* and *his* paused RNAs and RO transcripts are indicated with arrows. Sizes of the pBR322 *Msp*I fragments are indicated on the left. (**C**) Quantification of the data shown in (B) and assays performed with 50 nM each RfaH and NusG; assays were repeated three or more times for each protein combination. (**D**) Termination assay was performed as described in Carter *et al* (2004); the error bars are the standard deviations from four independent measurements.

2007) and σ-dependent pausing (Sevostyanova *et al*, 2008). We found that the RfaH NTD was also a much better NusG competitor (Figure 4A and B): NusG failed to interfere with the RfaH NTD activity (at either *opsP* or *hisP* site) even when present at 10-fold excess over RfaH and TEC.

We hypothesize that the RfaH CTD may be sequestered *in vivo* through binding to an unknown target(s) after its dissociation from the RfaH NTD is triggered by RfaH loading onto the *ops*-paused TEC. This would inhibit rebinding of the RfaH CTD to the NTD and thus indirectly stabilize RfaH binding to the TEC. Among the most feasible CTD interactors are the components of the translational machinery (Figure 4C): such interactions would favour transcription–transcription coupling and ensure that processive association of RfaH with the TEC is conditional on engagement of translational machinery. Indeed, using a pull-down assay with immobilized RfaH, we have identified several ribosomal proteins as potential RfaH targets (data not shown). Association of RfaH with the translational apparatus *in vivo* and the role of the CTD therein are currently under investigation.

RfaH excludes NusG in vivo

A chromatin immunoprecipitation followed by microarray hybridization (ChIP-on-chip) study of *E. coli* elongation

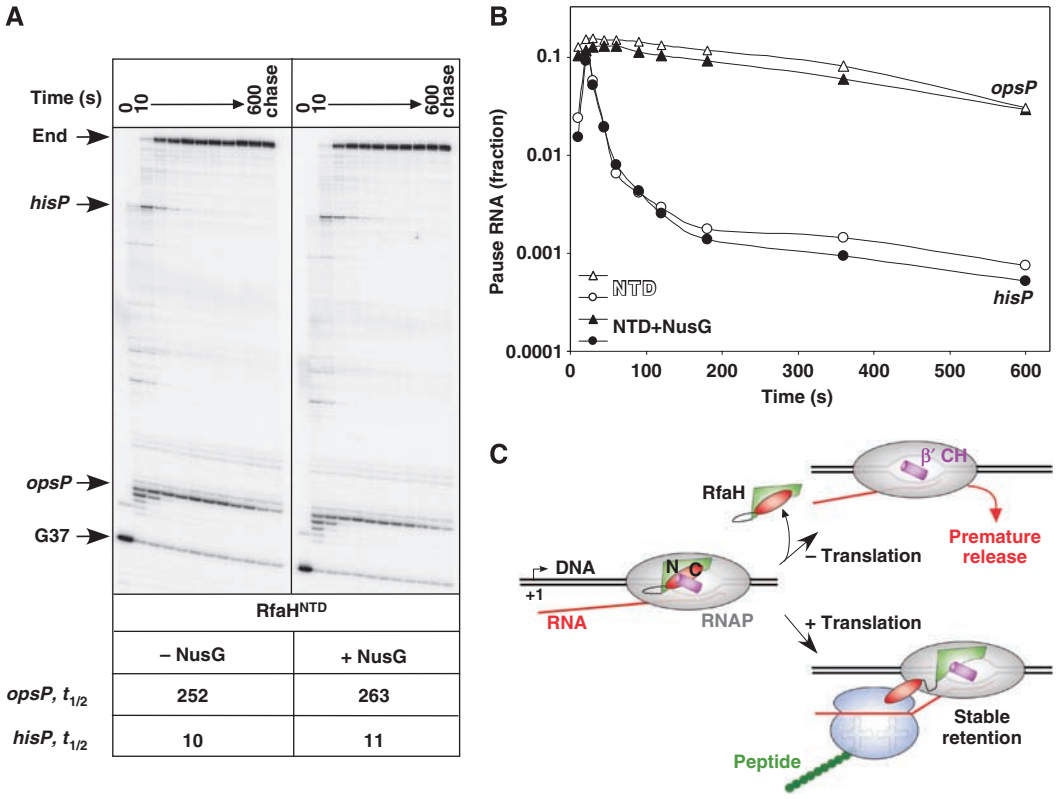

Figure 4 The RfaH NTD resists NusG competition. (**A**, **B**) The pause assay on template shown in Figure 3A. NusG was present at 500 nM, $RfaH^{NTD}$ was present at 50 nM. (B) Quantification of the *opsP* (triangles) and *hisP* (circles) pause RNAs from data shown in (A). Open symbols, NTD alone; filled symbols, NTD and NusG together. Pause half-lives are indicated below the gel panel. (**C**) Two possible fates of the RfaH-modified TEC. In the absence of translation *in vivo* (or *in vitro*), RfaH dissociates from the complex and re-establishes the interdomain interface. With concurrent translation, the CTD is sequestered, allowing the NTD to stay bound.

regulators (Mooney *et al*, 2009) demonstrated that NusG functions as a general transcription factor *in vivo*: it associates with RNAP transcribing most *E. coli* genes. However, if our *in vitro* observations are indicative of its *in vivo* activities, RfaH should be able to bind to RNAP at *ops* sites and remain bound throughout elongation. To determine whether this occurs *in vivo*, we performed ChIP-on-chip analysis of the genome-wide distribution of RfaH in *E. coli* strain MG1655 using polyclonal antiserum raised against *E. coli* RfaH. We observed a weak RfaH ChIP signal on most genes that increased on highly transcribed genes but that exhibited a continuous trend of RfaH/RNAP ratio (plotted versus $\log_2$ RNAP ChIP signal; Figure 5A); this weak RfaH signal could be due to (i) cross-reactivity of the polyclonal antibody, perhaps with NusG, or (ii) low-level binding of RfaH to high levels of RNAP in highly transcribed regions. However, a small subset of protein-coding genes exhibited significant enrichment of RfaH signal above this low-level background (Figure 5A). To ask whether this increased RfaH occupancy correlated with a decreased occupancy of NusG on these genes, we calculated the ratio of NusG/RNAP ChIP signals (Figure 5B; same *x* axis as in Figure 5A). The 37 protein-coding genes that exhibited elevated RfaH/RNAP ratios also exhibited exceptionally low NusG/RNAP ratios (average of 0.79 versus 0.99 for all genes with $\log_2$ RNAP signal below 0.5) and a distribution that was distinct from genes not enriched for RfaH (Mann–Whitney–Wilcoxon rank-sum test; $P<0.001$). Thus, genes enriched for RfaH exhibit significantly less NusG, confirming that RfaH excludes NusG *in vivo*.

The majority of RfaH-enriched genes clustered in two regions: *rfa* near min 82 of the *E. coli* chromosome and *rfb* near min 45. The *rfa* and *rfb* transcription units have been previously recognized as members of the RfaH regulon, defined as operons containing *ops* sequences that promote RfaH recruitment to TECs (Artsimovitch and Landick, 2002). To determine whether all transcription units containing this sequence exhibited increased RfaH signal, we compiled a list of *ops* sequences in the *E. coli* genome (MG1655; Supplementary Table 1). We then looked at these locations to determine whether they corresponded to sites with an increase in RfaH/RNAP ChIP signals. The average RfaH/RNAP for genes on which the RNAP signal is 0 or greater is 0.72 ± 0.25. Of the 17 genes, 5 exhibited a significantly higher-than-average RfaH/RNAP ratio: *yeaU*, *rfbB*, *wza*, *nirC*, and *rfaQ* (>two s.d. above the mean; $P<0.005$). An additional *ops*-containing gene, *ushA*, is right at the edge of significance with a ratio of 1.2, so it is possible that this gene is also a target of RfaH. However, we noted that many of these high ratios were due to lower than average RNAP signals. To determine which of these genes are actively transcribed under the growth conditions used (and thus likely to represent genes regulated by RfaH-bound TECs), we considered the reported expression levels of these genes (Allen *et al*, 2003). Only two of the five genes with a statistically significant high RfaH/RNAP ratio were transcribed (more than one transcript per cell; Supplementary Table 1). These two genes are *rfaQ* (in the *rfa* transcription unit; Figure 5D) and *rfbB* (in the *rfb* transcription unit; Figure 5E). It is also notable that of

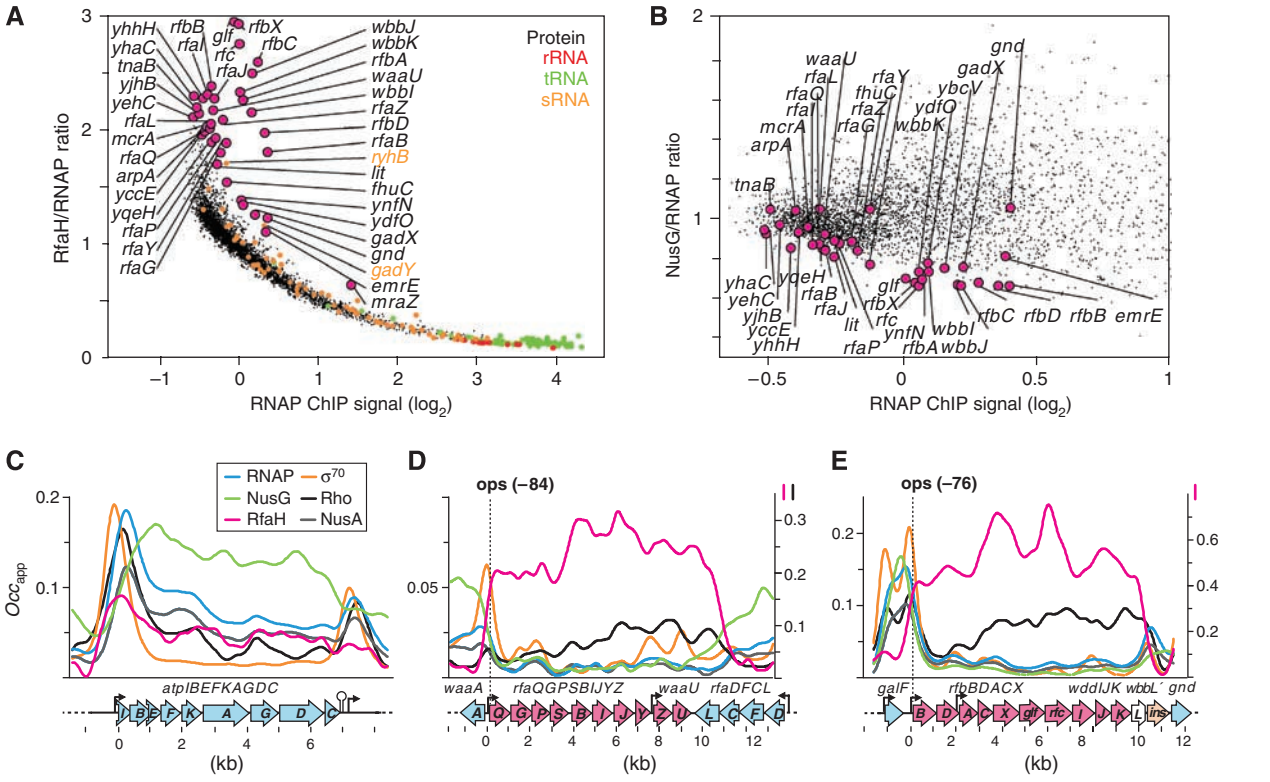

Figure 5 RfaH and NusG are targeted to different operons. (**A**) Gene-averaged RfaH ChIP-on-chip signal as a ratio to RNAP signal plotted as a function of $\log_2$ RNAP ChIP-on-chip signal (Materials and methods). Small black circles, protein-coding genes; red circles, rRNA genes; green circles, tRNA genes; orange circles, small RNA genes. The 37 protein-coding genes for which the $\log_2$ RfaH signal was greater than 0.5 and the $\log_2$ RNAP signal was less than 2 are shown as magenta circles. (**B**) Gene-averaged NusG/RNAP signals plotted as a function of $\log_2$ RNAP signal. Only genes with $\log_2$ RNAP signal below 1 are shown. Magenta circles as in (A). (**C**) Occ_{app} for *atp* (blue, RNAP; orange, σ^{70}; grey, NusA; green, NusG; black, Rho; magenta, RfaH) calculated as described in Mooney *et al* (2009) using two rounds of sliding-window smoothing (500 bp window). Genes are depicted as labelled open arrows; promoters, as vertical lines capped with arrows; and known intrinsic terminators, as hairpins. Vertical dotted lines indicate the positions of *ops* elements; number in parentheses corresponds to the distance between *opsP* and the start codon. (**D**, **E**) Occ_{app} for *rfa* and *rfb* regions depicted as for (C). RfaH is plotted using the secondary *y* axis in both panels, Rho– in (D) only. Note that the scales of Occ_{app} and TU length (in kilobase, denoted by hatchmarks) differ in (C–E).

the *ops*-containing genes, *rfaQ* and *rfbB* exhibit the highest RfaH signals. Another gene with RfaH/RNAP signal close to the cutoff for significance, *ushA*, is also transcribed, and thus it is possible that this gene exhibits weak RfaH binding, although it is not as significant as the *rfa* and *rfb* transcription units. It is possible that the low number of *ops*-containing genes with significant RfaH signal is due to our growth conditions: these experiments were carried out at early log phase in defined media supplemented with glucose. Alternate growth conditions, such as stress conditions or stationary phase, may result in an increased number of genes with RfaH signal.

To ask whether RfaH was retained across the *rfa* and *rfb* transcription units as expected, we calculated the apparent occupancy (Occ_{app}) of RfaH (Figure 5D and E; see Materials and methods) and compared it with RNAP and other transcription regulators as well as to a prototypical *E. coli* transcription unit, *atp* (Figure 5C). Occ_{app} for RNAP, NusA, and NusG were all near background on the *rfa* and *rfb* loci (though detectably above background for RNAP and NusA). RfaH exhibited high Occ_{app} beginning near the *ops* site within *rfa* and *rfb* (positioned less than 100 nt upstream of the first ORF in each operon) and extending across the transcription units. This pattern confirms retention of RfaH *in vivo* (and exclusion of NusG) after recruitment at an *ops* site. On *rfa* (Figure 5D), RfaH extends into the opposing *rfaDFC* transcription unit, suggesting that the site of termination may lie within *rfaL*. On *rfb* (Figure 5E), RfaH is lost when RNAP enters the insertion element that interrupts *wbbL*. Taken together, the ChIP-on-chip results strongly support the model of RfaH-directed persistent RNAP modification.

Some RfaH-enriched genes exhibited NusG/RNAP signal ratios close to the genome-wide average of ~1 (e.g. *rfaL* and *gnd*; Figure 5B) and were likely included in the RfaH-enriched set because they are adjacent to RfaH-regulated transcription units. Others (e.g. *arpA*, *yccE*, and *mcrA*) exhibit signals consistent with single-transcription units modified by RfaH. Three genes (*tnaB*, *fhuC*, and *mraZ*) were adjacent to significant RfaH peaks of unknown origin. Two sRNA genes (*ryhB* and *gadY*) may be anomalies attributable to the small number of probes present for each gene (4).

Interestingly, Rho appeared to also be enriched on the *rfa* and *rfb* transcription units. This pattern is strikingly different from that observed on most *E. coli* transcription units (e.g. *atp*, Figure 5C; see Mooney *et al*, 2009), but is consistent with the proposed role of Rho in blocking transcription of genes of foreign origin such as those acquired by horizontal transfer (Cardinale *et al*, 2008), a characteristic of several RfaH-controlled operons (Lawrence and Roth, 1996).

Phylogenetic analysis of the NusG family

To address how RfaH and NusG may have evolved by gene duplication and subsequent specialization, we performed phylogenetic analysis of NusG homologues from 511 prokaryotic genomes (Supplementary Table 2). Protein sequences were aligned using Muscle version 3.6 (Edgar, 2004). Regions of ambiguous alignment and indels were manually removed (Sjölander, 2004), followed by redundancy filtering. The resulting 119-residue alignment contained 243 taxa. Maximum likelihood analysis using this data set revealed three phylogenetically distinct families (Figure 6A). The first family includes all characterized eubacterial NusGs; the members of this clade are highly conserved, ubiquitous, and share their genomic location in the conserved gene cluster including *secE*, genes encoding ribosomal proteins L11, L1, L10, L7/L12, and (in most eubacterial genomes) RNAP β- and β′-subunits (*rpoB* and *rpoC*). The second family corresponds to highly conserved and ubiquitous archaeal NusGs. The third family includes all known paralogues of eubacterial NusG including RfaHs; we define this family as NusG^{SP} (for specialized paralogue of NusG). NusG^{SP} proteins display more sequence divergence, are widely distributed in eubacterial genomes (Supplementary Table 2 and Supplementary Figure 2), but are absent in Archaea. Many *nusG*^{SP} genes are not constrained to any particular gene cluster but are often found as the first ORF in operons encoding secreted proteins and LPS biosynthesis genes. A particularly prominent case is *Bacteroides fragilis*, which produces distinct capsular polysaccharides from eight operons, every one of which having an RfaH-like (UpxY) 5′ ORF thought to control its cognate operon expression (Krinos *et al*, 2001). Another example is TaA, which is required for synthesis of a polyketide antibiotic TA in *Myxococcus xanthus* (Paitan *et al*, 1999). In contrast, in proteobacterial genomes, *nusG*^{SP} genes form monocistronic transcription units or are positioned as the terminal ORFs (Supplementary Figure 3).

NusGs tolerate large insertions in the loop regions of the NTD, whereas NusG^{SP} NTDs and the CTDs within both

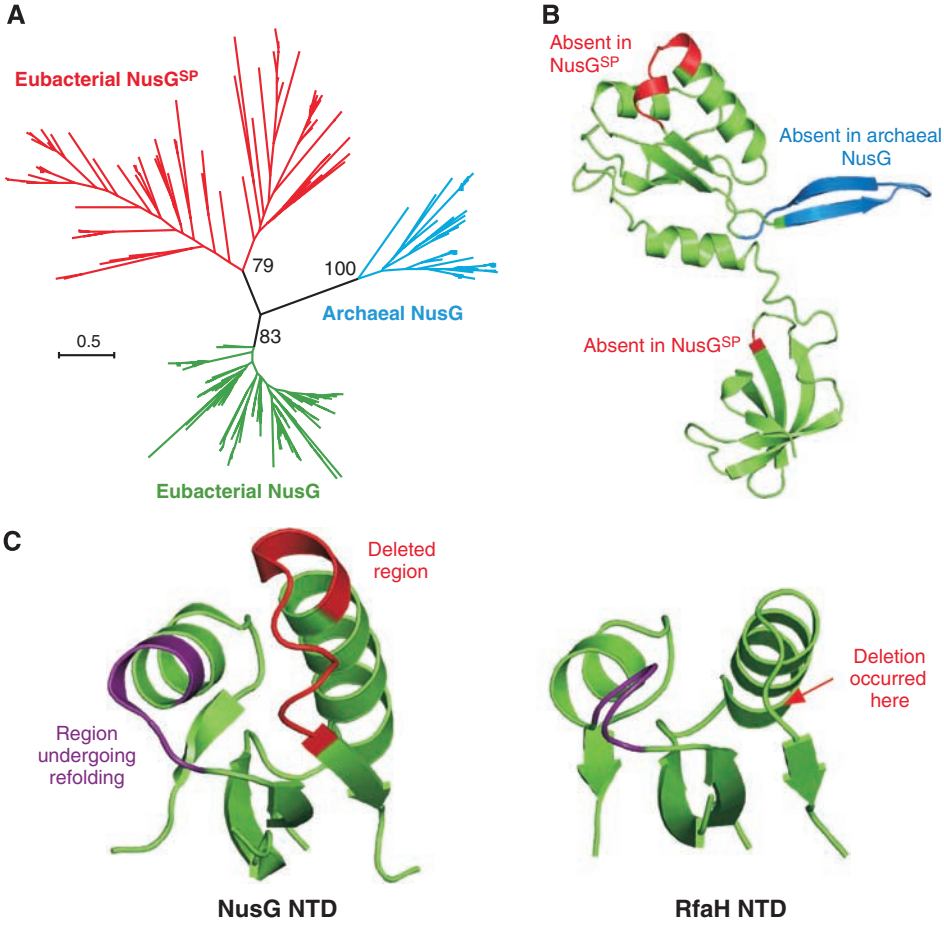

Figure 6 RfaH/NusG family. (**A**) Phylogenetic relationships of NusG homologues estimated with PHYML program (Guindon and Gascuel, 2003). Species names are omitted for clarity. Grouping into three separate clusters is well supported by bootstrap analysis (bootstrap percentages are indicated for the three nodes). In contrast, our data set of 119 reliably aligned amino-acid positions was too short for credible determination of the tree topology within clusters, as suggested by low bootstrap support for internal nodes (not shown). (**B**) The characteristic features of proteins from each cluster mapped onto *E. coli* NusG model. Archaeal NusG lacks the β-hairpin loop that is invariably present in eubacterial NusG homologues (blue). NusG^{SP} proteins possess a nine-residue deletion in NTD relative to eubacterial NusG (red). All but three deeply rooted NusG^{SP}s also contain one residue deletion in CTD (red). (**C**). Consequences of RfaH-specific deletion. The deletion (red) and the adjacent region undergoing refolding (purple) are mapped on the structures of *Aquifex aeolicus* NusG (PDB 1M1H, left) and *E. coli* RfaH (PDB 2OUG, right). The deletion shortens the flanking α-helix and β-strand by two residues, and changes the tilt of this helix, thereby reshaping the hydrophobic core and reducing its volume. The deletion also causes the N terminus of a neighbouring helix (purple) to unwind and fold into the protein interior, thereby partially compensating for the distortions in the hydrophobic core.

families are more constrained in size. Archaeal NusG lacks a 20-residue-long β-hairpin loop that inserts between the two β-strands of the NTD (Figure 6B) of the eubacterial proteins. The NusG^{SP} sequences invariably lack nine residues that adopt a helix-like structure positioned between the N-terminal α-helix and the second β-strand of NusG NTD. This deletion results in the shortening of the adjacent secondary structure elements and a significant reduction in the volume of the hydrophobic core of the NTD in RfaH relative to that of NusG (Figure 6C). Most NusG^{SP} sequences also possess a characteristic one-residue deletion in the CTD; this deletion maps within a loop region in both the all-β (NusG) and the all-α (RfaH) CTDs. Finally, both classes display significant variability of surface-exposed residues, but conserve structurally important prolines and glycines and the hydrophobicity of residues forming the protein core. Strikingly, in all NusG^{SP} sequences the hydrophobic residues forming the core of the all-β CTD are conserved (Supplementary Figure 4), even though only a fraction of these residues is involved in interdomain interactions in RfaH, leaving the remaining solvent exposed.

Discussion

In this study, we sought to determine the mechanism by which two homologous transcription elongation factors RfaH and NusG, which share a binding site on RNAP, are directed to control their respective genes. We show that RfaH and NusG compete for their effects on elongation *in vitro* and associate with distinct sets of *E. coli* operons *in vivo*. In contrast to NusG, which functions as a general transcription factor (Mooney *et al*, 2009), RfaH targets those few operons that are NusG-free and contain an *ops* sequence, which serves as the RfaH loading site. RfaH recognizes its target DNA even in the presence of NusG. Once recruited, RfaH remains stably bound to RNAP thereby preventing NusG loading onto the TEC. Finally, we discuss evolution of the NusG family that gave rise to proteins with different structures and distinct, perhaps even opposite, regulatory functions in the cell.

RfaH and NusG differ structurally and functionally

The drastic differences between the CTD structures most likely explain the distinct effects of RfaH and NusG on Rho-dependent termination and the differences in their mode of recruitment to the TEC (Figure 1). In NusG, the β-barrel CTD makes no contacts with the NTD (Reay *et al*, 2004) and may bind to Rho. In contrast, in RfaH the CTD is folded as an α-helical hairpin, makes extensive interactions with the NTD to shield the RNAP-binding site in the absence of *ops* (Belogurov *et al*, 2007), and does not bind to Rho. On the other hand, two distinguishing features of RfaH, specific binding to the *ops* DNA and acceleration of escape from hairpin-dependent pauses, are conferred by the isolated NTD (Belogurov *et al*, 2007), and are thus most likely attributed to variations in the NTDs (Figure 6C). The DNA-binding motif in RfaH has diverged significantly from the corresponding surface on NusG. Among 12 RfaH residues, the substitutions of which confer defects in *ops* binding (GB, unpublished data) only one (Arg16 in RfaH) is present in *E. coli* NusG. In contrast, the opposite, putative RNAP-binding side of the NTD is much more conserved (Supplementary Figure 5). The exact mechanistic explanation for the RfaH effects on pause escape rate awaits more detailed knowledge on the conformations of RNAP and associated factors in the paused complexes. Importantly, as hairpin-stabilized pause sites have not been characterized in RfaH-regulated operons, it is unclear whether this effect has any regulatory significance.

RfaH and NusG may have opposite functions in the cell

A recent study (Cardinale *et al*, 2008) suggests that one of the key roles of NusG is to maintain the operon borders and to limit the expression of horizontally acquired DNA by acting in concert with Rho to terminate the transcription of poorly translated messages. The apparent RfaH–NusG competition, the lack of Rho binding by RfaH, and the observation that many RfaH-controlled operons are horizontally transferred (Lawrence and Roth, 1996) or plasmid-encoded (Rehemtulla *et al*, 1986; Bailey *et al*, 1997) suggest that in *E. coli*, RfaH may have a function opposite to that of NusG in that it specifically increases the expression of foreign genes. Interestingly, many of these operons (absent in MG1655 used in our ChIP-on-chip assays) encode virulence functions; RfaH is necessary for virulence in animal models (Nagy *et al*, 2004). RfaH action may rely entirely on antitermination—indeed, RfaH defects are suppressed by mutations in *rho* (Farewell *et al*, 1991). Although RfaH does have a modest inhibitory effect on Rho-dependent termination *in vitro* (Figure 2), it is not clear whether RfaH directly interferes with Rho action *in vivo*. Alternatively, RfaH's major role could be to exclude NusG from the TEC, thereby blocking NusG-assisted Rho-dependent RNA release. RfaH may also promote translation through its interactions with the ribosome (see below).

Conversion of a NusG-like protein into RfaH

We envisage that the NusG^{SP} family of non-essential operon-specific transcription regulators originated from the duplication of ancestral NusG, which most likely occurred after the split between Bacteria and Archaea but before the differentiation of the major bacterial phyla (see Supplementary Figure 2). NusG^{SP} was subsequently lost in some species but has evolved into a highly specialized RfaH in Enterobacteria.

Conversion of a NusG duplicate into RfaH is characterized by four (not necessarily sequential) events: (i) a loss of binding to Rho, (ii) acquisition of a deletion in NTD with concomitant decrease in solubility and imposition of *cis* specificity, (iii) refolding of the CTD with concomitant regaining of solubility, and (iv) evolving of a mechanism of *ops*-specific recruitment allowing for *trans* action. In combination, these changes restrict RfaH action to a few, *ops*-containing operons while allowing NusG to regulate transcription of the remaining genome (Figure 5).

Given the importance of NusG–Rho cooperation in maintaining operon borders, the first event alone would be sufficient to make NusG^{SP} different. We propose that the NusG duplicate first lost its ability to bind Rho, most likely by acquiring mutations in the CTD. Extreme refolding of the CTD (Figure 1) would not be necessary at this point, as simply altering the Rho contact residues should be enough to eliminate its binding.

Next, a deletion that occurred in the NTD (Figure 6C) destabilized the domain structure by reshaping the hydrophobic core and reducing its volume. Although isolated NusG

NTD can be obtained at high concentrations, solubility of the isolated RfaH NTD and the NusG variant with a deletion in NTD (corresponding to that of RfaH) is below 20 μM (GAB, unpublished observations). Thus, the NTD deletion most likely generated a functional protein with very low solubility yet possessing the intact RNAP-binding site; indeed, RfaH NTD works well *in vitro* (Figure 4). Such a protein may favour elongation simply by tightly associating with RNAP to preclude NusG binding, thereby decreasing Rho-dependent termination. The low solubility would also restrict NusG^{SP} action to regions nearby its production site, thus establishing *cis* specificity (see below).

To gain control over several operons, NusG^{SP} had to become soluble. This was most likely achieved when CTD evolved the ability to refold into a helical conformation, the only state that can stably mask the hydrophobic surface on the NTD as the hydrophobic residues are exposed on the α-helix but hidden inside the β-barrel (Supplementary Figure 4). The characteristic one-residue deletion in NusG^{SP} CTD most likely has an important function in this process, because RfaH variants with one-residue insertions in CTD are insoluble, as is the isolated RfaH NTD (GAB, unpublished observations). However, as all the residues that constitute the hydrophobic core of the β-barrel are conserved in RfaH, it is possible that the CTD can adopt alternative folds. In full-length RfaH, interactions with the NTD would favour the α-hairpin. On domain dissociation, the CTD may be able to refold into a β-barrel, particularly if the latter is stabilized by interactions with an unknown cellular target. The β-fold may execute a function conserved in all NusG homologues (e.g. binding to ribosome). Refolding on this scale is not entirely unprecedented—indeed, rapid inter-conversion between two drastically different folds of lymphotactin, which mediate its interactions with different ligands, has been reported (Tuinstra *et al*, 2008).

Operon-specific control by NusG^{SP}

The most important and complex event in NusG^{SP} evolution is the development of the sequence-specific *trans* recruitment mechanism. Refolded CTD not only makes the protein soluble but also masks the RNAP-binding site on the NTD, thereby preventing RfaH recruitment in the absence of the *ops* site (Belogurov *et al*, 2007). At present, the events leading to the emergence of the *ops* site and evolution of the *ops*–NusG^{SP} interactions, as well as the mechanism of *ops*-induced domain dissociation *per se*, are poorly understood. The complex and multi-step nature of this transformation raise the question whether an alternative, simpler mechanism existed throughout evolution, allowing for NusG^{SP} functioning and thus fixation prior to development of *trans* recruitment.

Our inference of the low-solubility stage in NusG^{SP} evolution suggested that the factor's action would be restricted to the vicinity of its production site. Indeed, survey of bacterial genomes revealed that, in contrast to enterobacterial *rfaH* genes that form monocistronic units, other *nusG^{SP}* genes are commonly positioned as the first ORF in operons they potentially control. This observation suggests that many contemporary NusG^{SP}s, as well as their ancestor, may control their target operons in *cis*. We hypothesize that these proteins attach to RNAP through an RNA tether, a common mechanism for the recruitment of antitermination proteins.

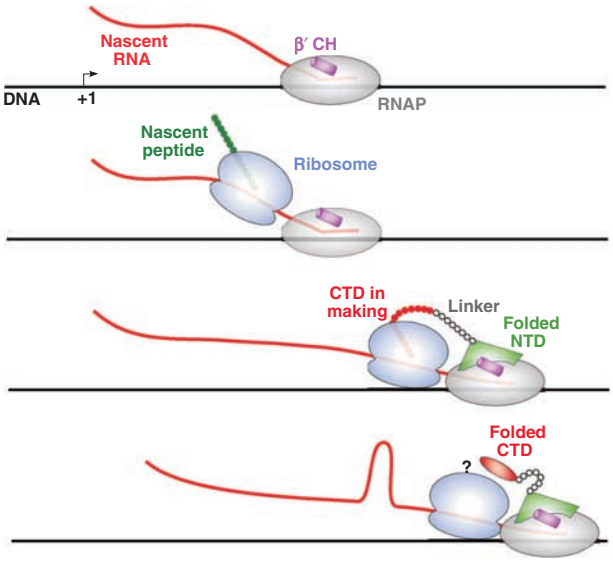

Figure 7 A *cis* action model of NusG^{SP}. See text for details.

In contrast to λN and the components of *rrn* antitermination complex, which recognize a specific site in mRNA (Mason *et al*, 1992; Squires *et al*, 1993), NusG^{SP} would bind to RNAP co-translationally, that is, through mRNA–ribosome–nascent protein tether (Figure 7) as soon as its NTD is folded after synthesis on the ribosome. If the NusG^{SP} CTD has also preserved the ability to bind to the ribosome (that we hypothesize exists in NusG), it could tie the latter to the transcribing RNAP (Figure 4C), thereby providing an additional line of defence against Rho-dependent termination. In the above mechanism, the problems of low NTD solubility and its targeting to a particular operon are solved by the *cis* action.

Following the development of *trans* recruitment mechanism, the NusG^{SP} gene became isolated as a monocistronic unit, thus abolishing the *cis* recruitment. This organization enables dual control of the NusG^{SP} regulon: global, by altering NusG^{SP} expression, and local, by modulating the promoter activity of each operon. The *trans* acting NusG^{SP} can also function as a true catalyst of transcription of the target genes as opposed to a *cis* acting factor that has to be produced in stoichiometric amounts.

The proposed stepwise pathway by which a general transcription factor NusG has evolved into an operon (or regulon)-specific regulator RfaH is certainly hypothetical, but it is broadly consistent with the available functional and phylogenetic data. We are currently testing the implication of this model.

Materials and methods

Proteins and reagents

Oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA), NTPs and [α^{32}P]NTPs were from GE Healthcare (Piscataway, NJ), restriction and modification enzymes from NEB (Ipswich, MA), PCR reagents from Roche (Indianapolis, IN), and other chemicals from Sigma (St Louis, MO) and Fisher (Pittsburgh, PA). Plasmid DNAs and PCR products were purified using spin kits from Qiagen (Valencia, CA). Rho protein was a gift from Dr J Richardson. *E. coli* RNAP and RfaH were purified as described in

Belogurov *et al* (2007). NusG was purified as described in Artsimovitch and Landick (2000).

Halted complex formation

Linear templates for *in vitro* transcription were generated by PCR amplification. TECs were formed with 40 nM of linear DNA template and 50 nM of RNAP holoenzyme in 20–100 μl of transcription buffer (20 mM Tris-chloride, 20 mM NaCl, 2 mM $MgCl_2$, 14 mM 2-mercaptoethanol, 0.1 mM EDTA, 5% glycerol, pH 7.9). To make the elongation complexes halted after the addition of G37 on pIA349 and pIA416 (or A38 at pVS55) templates, transcription was initiated in the absence of UTP, with ApU at 150 μM, ATP and GTP at 2.5 μM, CTP at 1 μM, with ^{32}P derived from [α^{32}P]CTP (3000 Ci/mmol). Halted complexes were formed for 15 min at 37°C and stored on ice prior to use.

Single-round pause assays

Halted [^{32}P]CTP elongation complexes were prepared in 50 μl of transcription buffer. Elongation factors were added followed by 3-min incubation at 37°C. Transcription was restarted by the addition of GTP to 15 μM, CTP, ATP and UTP to 150 μM, and rifapentin to 25 μg/ml. Samples were removed at 10, 20, 40, 60, 90, 120, 180, 300, and 600 s and after an additional 5-min incubation with 200 μM GTP (chase), and quenched by the addition of an equal volume of STOP buffer (10 M urea, 50 mM EDTA, 45 mM Tris-borate; pH 8.3, 0.1% bromophenol blue, 0.1% xylene cyanol).

Intrinsic termination assays

Halted complexes were prepared in 20 μl of transcription buffer with 25 nM of linear DNA pIA416 template and 40 nM of RNAP holoenzyme. Elongation factors were added followed by 3-min incubation at 37°C. Elongation was restarted by the addition of NTPs (10 μM UTP, 200 μM ATP, CTP, and GTP) and rifapentin. Reactions were incubated at 37°C for 15 min and quenched as above.

Rho-dependent termination assays

Halted after the addition of A26 on pIA267 template, elongation complexes were prepared in the absence of CTP in 30 μl of Rho buffer (40 mM Tris–HCl, 50 mM KCl, 5 mM $MgCl_2$, 0.1 mM dithiothreitol, 3% glycerol, pH 7.9) supplemented with ApU at 150 μM, ATP and UTP at 2.5 μM, GTP at 1 μM, and 5 μCi of [α^{32}P]GTP (3000 Ci/mmol) during 15-min incubation at 37°C. Elongation factors were added followed by 3-min incubation at 37°C. Transcription was restarted by the addition of GTP to 15 μM, CTP, ATP, and UTP to 150 μM, and rifapentin. Reactions were incubated at 37°C for 15 min and stopped as above.

Sample analysis

Samples were heated for 2 min at 90°C and separated by electrophoresis in denaturing acrylamide (19:1) gels (7 M Urea, 0.5 × TBE) of various concentrations (5–10%). RNA products were visualized and quantified using a PhosphorImager Storm 820 System (GE Healthcare), ImageQuant Software, and Microsoft Excel.

ChIP-on-chip assays

ChIP-on-chip analysis of RfaH was performed on MG1655 *HA_3::nusG* cells grown in MOPS minimal glucose medium to mid-log phase as described by Mooney *et al* (2009). RfaH polyclonal rabbit antibodies were obtained from ProSci Inc.; 12CA5 anti-HA monoclonal antibody targeting HA_3::NusG (used to avoid possible cross-reaction with RfaH) was from Roche. Chip-on-chip data were normalized, gene-averaged, and used to calculate apparent occupancy (Occ_{app}) as described by Mooney *et al* (2009). Occ_{app} is a function of true occupancy, relative 'crosslinkability', and possibly epitope-masking of the target. Thus, Occ_{app} is useful for comparing occupancy of a given target across or among transcription units, but has no reliable physical interpretation and cannot be used to compare different targets. Gene-averaged signals were computed from the ChIP signal on all probes, the midpoints of which were within the boundaries of genes. Genes for which the array design contained fewer than four probes were excluded from analysis, yielding 37 genes with an average $\log_2$ RfaH signal above 0.5 and an average $\log_2$ RNAP signal below 2 that were identified as enriched for RfaH (magenta circles in Figure 5A; genes with average $\log_2$ RfaH signal above 0.5 and an average $\log_2$ RNAP signal above 2 did not differ from the trend line of RfaH/RNAP signal ratio).

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Yuri Wolf and Michael Ludwig for discussions, John Richardson for the gift of Rho protein, and the University of Oslo Bioportal for providing CPU time. This study was supported by the National Institutes of Health grants GM67153 (to IA) and GM38660 (to RL).

References

Allen TE, Herrgard MJ, Liu M, Qiu Y, Glasner JD, Blattner FR, Palsson BO (2003) Genome-scale analysis of the uses of the *Escherichia coli* genome: model-driven analysis of heterogeneous data sets. *J Bacteriol* **185:** 6392–6399

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

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

Bailey M, Hughes C, Koronakis V (1997) RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. *Mol Microbiol* **26:** 845–851

Bailey MJ, Hughes C, Koronakis V (2000) *In vitro* recruitment of the RfaH regulatory protein into a specialised transcription complex, directed by the nucleic acid ops element. *Mol Gen Genet* **262:** 1052–1059

Belogurov GA, Vassylyeva MN, Svetlov V, Klyuyev S, Grishin NV, Vassylyev DG, Artsimovitch I (2007) Structural basis for converting a general transcription factor into an operon-specific virulence regulator. *Mol Cell* **26:** 117–129

Burns C, Nowatzke W, Richardson J (1999) Activation of Rho-dependent transcription termination by NusG. Dependence on terminator location and acceleration of RNA release. *J Biol Chem* **274:** 5245–5251

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

Carter HD, Svetlov V, Artsimovitch I (2004) Highly divergent RfaH orthologs from pathogenic proteobacteria can substitute for *Escherichia coli* RfaH both *in vivo* and *in vitro*. *J Bacteriol* **186:** 2829–2840

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797

Farewell A, Brazas R, Davie E, Mason J, Rothfield L (1991) Suppression of the abnormal phenotype of *Salmonella typhimurium* rfaH mutants by mutations in the gene for transcription termination factor Rho. *J Bacteriol* **173:** 5188–5193

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704

Krinos CM, Coyne MJ, Weinacht KG, Tzianabos AO, Kasper DL, Comstock LE (2001) Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature* **414:** 555–558

Lau LF, Roberts JW (1985) Rho-dependent transcription termination at lambda R1 requires upstream sequences. *J Biol Chem* **260:** 574–584

Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143:** 1843–1860

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

Linn T, Greenblatt J (1992) The NusA and NusG proteins of *Escherichia coli* increase the *in vitro* readthrough frequency of a transcriptional attenuator preceding the gene for the beta subunit of RNA polymerase. *J Biol Chem* **267:** 1449–1454

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

Mooney RA, Davis SE, Peters J, Rowland J, Ansari AZ, Landick R (2009) Regulator trafficking on bacterial transcription units *in vivo*. *Mol Cell* (in press)

Nagy G, Dobrindt U, Hacker J, Emody L (2004) Oral immunization with an rfaH mutant elicits protection against salmonellosis in mice. *Infect Immun* **72:** 4297–4301

Nehrke KW, Platt T (1994) A quaternary transcription termination complex. Reciprocal stabilization by Rho factor and NusG protein. *J Mol Biol* **243:** 830–839

Paitan Y, Orr E, Ron E, Rosenberg E (1999) A NusG-like transcription anti-terminator is involved in the biosynthesis of the polyketide antibiotic TA of Myxococcus xanthus. *FEMS Microbiol Lett* **170:** 221–227

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

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

Rees W, Weitzel S, Yager T, Das A, von Hippel P (1996) Bacteriophage lambda N protein alone can induce transcription antitermination *in vitro*. *Proc Natl Acad Sci USA* **93:** 342–346

Rehemtulla A, Kadam SK, Sanderson KE (1986) Cloning and analysis of the sfrB (sex factor repression) gene of *Escherichia coli* K-12. *J Bacteriol* **166:** 651–657

Roberts JW, Yarnell W, Bartlett E, Guo J, Marr M, Ko DC, Sun H, Roberts CW (1998) Antitermination by bacteriophage lambda Q protein. *Cold Spring Harb Symp Quant Biol* **63:** 319–325

Sevostyanova A, Svetlov V, Vassylyev DG, Artsimovitch I (2008) The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. *Proc Natl Acad Sci USA* **105:** 865–870

Sjölander K (2004) Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics* **20:** 170–179

Squires CL, Greenblatt J, Li J, Condon C (1993) Ribosomal RNA antitermination *in vitro*: requirement for Nus factors and one or more unidentified cellular components. *Proc Natl Acad Sci USA* **90:** 970–974

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

Stevens M, Clarke B, Roberts I (1997) Regulation of the *Escherichia coli* K5 capsule gene cluster by transcription antitermination. *Mol Microbiol* **24:** 1001–1012

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

Svetlov V, Belogurov GA, Shabrova E, Vassylyev DG, Artsimovitch I (2007) Allosteric control of the RNA polymerase by the elongation factor RfaH. *Nucleic Acids Res* **35:** 5694–5705

Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, Volkman BF (2008) Interconversion between two unrelated protein folds in the lymphotactin native state. *Proc Natl Acad Sci USA* **105:** 5057–5062

