A large insertion domain in the Rho factor from a low G+C, Gram-negative bacterium is critical for RNA binding and transcription termination activity

Isabelle Simon, Mildred Delaleau, Annie Schwartz, Marc Boudvillain

A large insertion domain in the Rho factor from a low G+C, Gram-negative bacterium is critical for RNA binding and transcription termination activity

Isabelle Simon[a,b], Mildred Delaleau[a], Annie Schwartz[a], and Marc Boudvillain[a,c]

[a]: Centre de Biophysique Moléculaire, CNRS UPR4301, rue Charles Sadron, 45071 Orléans cedex 2, France.

[b]: ED 549, Sciences Biologiques & Chimie du Vivant, Université d'Orléans, France.

[c]: Corresponding author (marc.boudvillain@cnrs.fr; +33 238 25 55 85)

Running title: Termination factor Rho from *Bacteroides fragilis*

ABSTRACT

Rho-dependent termination of transcription (RDTT) is a critical regulatory mechanism specific to bacteria. In a subset of species including most Actinobacteria and Bacteroidetes, the Rho factor contains a large, poorly conserved N-terminal insertion domain (NID) of cryptic function. To date, only two NID-bearing Rho factors from high G+C Actinobacteria have been thoroughly characterized. Both can trigger RDTT at promoter-proximal sites or with structurally constrained transcripts that are unsuitable for the archetypal, NID-less Rho factor of *Escherichia coli* ($_{Ec}$Rho). Here, we provide the first biochemical characterization of a NID-bearing Rho factor from a low G+C bacterium. We show that *Bacteroides fragilis* Rho ($_{Bf}$Rho) is a *bona fide* RNA-dependent NTPase motor able to unwind long RNA:DNA duplexes and to disrupt transcription complexes. The large NID (~40% of total mass) strongly increases $_{Bf}$Rho affinity for RNA, is strictly required for RDTT, but does not promote RDTT at promoter-proximal sites or with a structurally constrained transcript. Furthermore, the NID does not preclude modulation of RDTT by transcription factors NusA and NusG or by the Rho inhibitor bicyclomycin. Although the NID contains a prion-like Q/N-rich motif, it does not spontaneously trigger formation of β-amyloids. Thus, despite its unusually large RNA binding domain, $_{Bf}$Rho behaves more like the NID-less $_{Ec}$Rho than NID-bearing counterparts from high G+C Actinobacteria. Our data highlight the evolutionary plasticity of Rho's N-terminal region and illustrate how RDTT is adapted to distinct genomic contents.

Keywords: RNA synthesis; transcription termination; Rho factor; molecular motor; RNA helicase; RNA folding; Bacteroidetes

Abbreviations: NID (N-terminal insertion domain); PBS (Primary binding site); RDTT (Rho-dependent termination of transcription); RNAP (RNA polymerase); Rut (Rho utilization); SBS (Secondary binding site); TEC (transcription elongation complex);

INTRODUCTION

Bacteria are very diverse microorganisms able to grow in a wide variety of habitats and conditions. Both species-specific traits and general regulatory mechanisms help tune bacterial gene expression programs to specific demands and contexts. A relevant regulatory mechanism is Rho-dependent termination of transcription (RDTT), which monitors the coupling of transcription and translation during bacterial gene expression[1-5]. Uncoupling of the transcription and translation machineries, either because of a 'program bug' (e.g. nonsense mutation) or a contextual event (e.g. sRNA- or riboswitch-induced translation impediment), provides Rho access to intragenic RDTT signals in the nascent transcript (**Figure 1A**). Rho can then use its ATPase-driven activity to dissociate RNA polymerase (RNAP), thereby ending transcription of the corresponding gene or operon[2, 3, 6]. Beside surveillance of transcription-translation coupling, a major function of RDTT is the silencing of pervasive transcription, mostly in regions antisense to genes where transcription elongation complexes (TECs) are not protected from Rho by translating ribosomes[7-10]. Yet, a ribosome-free TEC is not the sole requirement for efficient RDTT as diffuse RNA structure/sequence determinants (e.g. C>G skewed sequence in *Escherichia coli*) or action of cofactors also play an important role. Remarkably, conditional RDTT signals controlled by riboswitches, RNA-binding proteins, or sRNAs have been found in 5'-untranslated and intergenic regions where they contribute to a variety of regulatory mechanisms[4, 11-15]. The broad regulatory function of Rho is also illustrated by its implication in the maintenance of bacterial genome integrity through the disruption of deleterious transcriptional R-loops and stalled TECs[16, 17].

RDTT has long been regarded as essential in Gram-negative bacteria but dispensable in Gram-positive species (reviewed in [18]). As a result, most studies aimed at elucidating the mechanism of RDTT have focused on the *Escherichia coli* paradigm. These studies support that *E. coli*'s Rho ($_{Ec}$Rho) preferentially binds RNA at C>G sequence-skewed and poorly

structured *Rut* (*R*ho *ut*ilization) sites. Once bound to a *Rut* site, the $_{Ec}$Rho hexamer switches to a closed ring conformation entrapping RNA within its central channel and becomes competent for ATP-dependent RNA translocation. The classical RDTT model posits that such directional ATPase-fueled movement from a *Rut* 'anchoring' site towards the transcript 3'-end allows Rho to catch up with RNAP and then disrupt the TEC (see: [19-22] and references within). However, a series of cryoEM structures from two recent studies[23, 24] supports an alternative 'allosteric' model whereby $_{Ec}$Rho first binds ribosome-free RNAP[25]. The Rho-RNAP interaction involves the N-terminal face of the $_{Ec}$Rho hexamer (**Figure 1A**) and is stabilized by cofactors NusA and NusG that both bind $_{Ec}$Rho and the RNAP[23, 24]. An additional role of NusA may be to help guide the RNA chain from the RNA exit channel of RNAP towards the central channel of the $_{Ec}$Rho hexamer[24], which remains open in the cryoEM structures. Recognition of a *Rut* site upon transcript scanning by $_{Ec}$Rho's primary binding site (PBS) would sequentially trigger allosteric inactivation of RNAP, closure of the $_{Ec}$Rho ring and activation of its ATPase-fueled motor capacity, and finally TEC dissociation by the $_{Ec}$Rho motor[23, 24]. In some instances, RDTT can also initiate at suboptimal *Rut* sites with the help of the transcription factor NusG[7, 26-28]. This could stem from the ability of *E. coli*'s NusG ($_{Ec}$NusG) to stimulate closure of the $_{Ec}$Rho ring around suboptimal RNA substrates[29, 30]. The direct action of $_{Ec}$NusG on RNAP, limiting termination-resistant backtracked states, may also be particularly relevant at suboptimal RDTT terminators ([24] and references within). Overall, these rather lax RNA recognition rules combined with high intracellular concentrations of $_{Ec}$Rho and $_{Ec}$NusA/G factors sustain RDTT at hundreds of sites throughout the *E. coli* genome[7, 10, 31-33].

Renewed interest in RDTT beyond the *E. coli* paradigm has spurred from the recent findings that *rho* inactivation also triggers genome-wide pervasive (mostly antisense) transcription in Gram-positive species such as *Bacillus subtilis*[34], *Staphylococcus aureus*[8], or *Mycobacterium tuberculosis*[9]. Contrary to early expectations, depletion of Rho in these species

can have significant consequences, from changes in cell differentiation or virulence programs in Firmicutes[34-36] to rapid death for *M. tuberculosis*[9]. The recent discovery of conditional regulatory mechanisms involving RDTT in the industrially important bacterium *Corynebacterium glutamicum*[37] also pinpoints to RDTT regulation being more widespread and decisive than previously envisioned. Yet, significant differences in genome compositions, metabolisms, or ecosystems are likely to affect how RDTT is carried out in phylogenetically divergent species. One central question is how RDTT is tuned to the specific features of individual transcriptomes, which can significantly vary in sequence skewness and degrees of secondary structure. Another key issue is whether the role of RDTT cofactors such as NusA/G is conserved throughout the bacterial kingdom.

Although the sequence of Rho is relatively well conserved among bacteria, a large amino acid (aa) insertion is often found in the N-terminal domain (NTD) of the protein (**Figure 1B,C** and **Figure S1**)[38]. This occurrence is particularly frequent in Actinobacteria and Bacteroidetes, implying that the NTD insertion domains (NIDs) have important roles in these phyla[38]. The NIDs are poorly conserved but are often R/K-rich in Actinobacteria and Q/N-rich in Bacteroidetes[38]. The role of the NID has been investigated for two GC-rich Actinobacteria (*Micrococcus luteus* and *M. tuberculosis*) where it apparently helps the Rho factor to deal with structured transcripts[39-42]. The NID also promotes promoter-proximal transcription termination *in vitro*[39-41], much like $_{Ec}$NusG does with $_{Ec}$Rho[43, 44], suggesting that the NID may substitute, at least in part, for NusG function in Actinobacteria[40, 45]. By contrast, the role(s) of the NIDs in Rho factors from low G+C bacteria, in particular from Bacteroidetes, has not been determined yet.

In this work, we have prepared and characterized the Rho factor from the opportunistic pathogen *Bacteroides fragilis* ($_{Bf}$Rho) to evaluate the role of the NID from a low G+C species (43.1% GC-richness vs. 50.8% for *E. coli* and 65.6% for *M. tuberculosis*). Growth sensitivity

to the Rho-specific inhibitor bicyclomycin (BCM)[46] and transposon inactivation[47] experiments point to an important role of Rho in *B. fragilis*. Consistent with these observations, we find that $_{Bf}$Rho is a functional NTP-dependent motor that is able to disrupt RNA-DNA duplexes and transcription elongation complexes *in vitro*. Much like *M. tuberculosis* Rho ($_{Mtb}$Rho), $_{Bf}$Rho needs its large NID to bind RNA and work effectively. Contrary to $_{Mtb}$Rho however, $_{Bf}$Rho cannot deal effectively with structured transcripts or trigger promoter-proximal transcription termination. In these respects, $_{Bf}$Rho performance is more akin to (but weaker than) that of $_{Ec}$Rho. Notably, $_{Bf}$Rho activity can be regulated by NusA and NusG factors and inhibited by BCM. Altogether, our data illustrate how Rho enzymatic features are tuned to the composition of the bacterium transcriptome. Furthermore, they demonstrate that NIDs can be necessary in low G+C species and, thus, can encompass functions beyond helping Rho dealing with highly structured transcripts.

RESULTS

Analysis of the $_{Bf}$Rho sequence and preparation of protein variants to probe $_{Bf}$Rho features

The archetypal $_{Ec}$Rho hexamer is composed of six identical protomers, each made of two distinct domains[1, 2]. The NTD carries the PBS motifs involved in Rho recognition of YC-rich (Y being a C or U residue) transcript *Rut* sites (**Figure 1A,C** and **Figure S1**)[48]. The NTD also carries the residues contacting RNAP in the recent cryoEM structures of the $_{Ec}$Rho:RNAP complex[23, 24]. The C-terminal domain (CTD) carries the motifs necessary for intersubunit cohesion and for ATP-dependent RNA translocation. The CTD notably includes the Walker A and B motifs forming ATPase pockets at subunit interfaces and the "catalytic Glu', 'Arg valve' and 'Arg finger' residues required for catalysis of ATP hydrolysis[49, 50]. The CTD also contains the secondary binding site (SBS) Q-loop and R-loop motifs that allosterically move RNA

through the hexamer central channel as a function of the chemical state of the ATPase pockets (**Figure 1A** and **Figure S1**)[49].

As already mentioned, Rho factors from *Bacteroidetes* frequently contain large NIDs[38]. This is the case of the 688 aa-long $_{Bf}$Rho factor, which carries an even longer NID than does $_{Mtb}$Rho (270 aa and ~30kDa versus 144 aa and ~15kDa) (**Figure 1B, C**). The sequence identity/similarity between both NIDs ($_{Bf}$Rho-NID and $_{Mtb}$Rho-NID) is actually very low (**Figure S1**). Their predicted isoelectric points are also significantly different (8.98 for $_{Bf}$Rho-NID vs. 5.20 for $_{Mtb}$Rho-NID) despite similar contents in basic residues (18.1 vs 17.3%). The higher pI of $_{Bf}$Rho-NID suggests that the domain might be an even better RNA interactor than $_{Mtb}$Rho-NID[42]. While both NIDs are predicted to be intrinsically disordered by XtalPred[51], an intriguing additional feature of $_{Bf}$Rho-NID is the presence of a 33 aa-long, asparagine/glutamine-rich sequence that displays characteristics of sequences able to form β-amyloids (**Figure S2**). Such prion-like Q/N-rich NID motifs appear to be frequent in *Bacteroidetes* Rho factors, except in the *Prevotella* genus (**Figure S3**). They are also found in other phyla, as in the Rho factor from Gram-positive *Clostridium botulinum*, which is indeed able to form prion-like aggregates compromising its RDTT activity[52, 53].

The remaining $_{Bf}$Rho sequence is well conserved, for instance displaying 77.1% similarity and 56.7% identity with the NID-less $_{Ec}$Rho sequence. All the residues and motifs critical for the RNA binding and catalytic activities of $_{Ec}$Rho are highly conserved in $_{Bf}$Rho (**Figure S1**). Notably and in contrast to $_{Mtb}$Rho[41], $_{Bf}$Rho does not contain indels in the CSD-like RBD region that carries the PBS residues anchoring $_{Ec}$Rho to *Rut* sites[48, 54]. The Q- and R-loops that translocate RNA in the hexamer central channel[49, 55, 56] are also highly conserved between $_{Bf}$Rho and $_{Ec}$Rho (**Figure S1**). Three aa substitutions are noticeable in the ATPase Walker motifs of $_{Bf}$Rho (P→Q and A→T in motif A and L→F in motif B; **Figure S1**) but correspond to side-chains that do not contact ATP directly in $_{Ec}$Rho[49, 50]. Finally, all the aa targeted by the

natural inhibitor BCM near the ATPase pocket[57, 58] are also conserved in $_{Bf}$Rho (**Figure S1**). Thus, the $_{Bf}$Rho factor does not display strikingly unusual sequence features, except for its very large and basic NID.

To probe the activity of the $_{Bf}$Rho factor and the properties of its NID, we have prepared wild-type (WT) $_{Bf}$Rho and derivatives lacking the full NID (Δ_{Bf}Rho) or only the putative prion-like Q/N-rich motif (aa 264-296; $_{Bf}$Rho$_{\beta del}$). For sake of comparison, we have also prepared WT $_{Mtb}$Rho and its NID-less variant (Δ_{Mtb}Rho)[41]. All proteins were purified to >90% homogeneity and yielded high-resolution mass spectrometry (HR-MS) peaks consistent with their calculated masses after proteolytic cleavage of the starting methionine (data not shown)[59]. Diffusion light scattering and gel filtration experiments support that all proteins form monodisperse species, most likely hexamers (**Figure S4**)[59]. The $_{Bf}$Rho and $_{Bf}$Rho$_{\beta del}$ proteins display anomalously low electrophoretic mobilities on SDS-PAGE gels when compared to protein standards or to Δ_{Bf}Rho (**Figure S4**). This band 'shift' effect cannot be assigned easily to a specific feature of the $_{Bf}$Rho-NID since many factors, including single amino acid substitutions, can strongly affect SDS-PAGE migration[60].

The $_{Bf}$Rho factor is a RNA-dependent NTP hydrolase

To assess the substrate and cofactor preferences of $_{Bf}$Rho, we determined its steady-state NTP hydrolysis activity in the presence of various rNTPs and polyribonucleotide homopolymers. We observed that the ATPase activity of $_{Bf}$Rho is strongly stimulated by the presence of poly[rC], reaching a plateau value of ~60 ATP/hexamer/s at saturating poly[rC] concentrations **(Figure 2A**). As for $_{Ec}$Rho[61], the ATPase rate displays a strong sigmoidal dependence on poly[rC] concentration (Hill coefficient ~5; $K_{d,app}$ ~1.5 µM), suggesting cooperativity among RNA binding sites **(Figure 2A**). By contrast, the ATPase activity of $_{Mtb}$Rho has an hyperbolic,

non-cooperative dependence on poly[rC] concentration up to a maximum value of ~35 ATP/hexamer/s[41]. Other polyribonucleotide cofactors are much less efficient than poly[rC] at stimulating $_{Bf}$Rho's ATPase **(Figure 2B)**, a feature also shared with $_{Ec}$Rho[39, 62]. Another similitude between $_{Bf}$Rho and $_{Ec}$Rho is their capacity to hydrolyze the four rNTP substrates with comparable efficiencies **(Figure 2C)**[39] whereas evolutionary distinct Rho factors from *M. tuberculosis*[41], *Micrococcus luteus*[39] and *Streptomyces lividans*[63] are more discriminatory, especially against CTP. Deletion of the full NID (Δ_{Bf}Rho) or that of the smaller Q/N-rich motif ($_{Bf}$Rho$_{\beta del}$) do not significantly affect ATP hydrolysis under saturating poly[rC] conditions (**Figure 2B**).

Taken together, these data show that $_{Bf}$Rho is a RNA-dependent NTPase that is more promiscuous and efficient than $_{Mtb}$Rho[41]. The large $_{Bf}$Rho-NID does not confer cofactor/substrate preference features that differ significantly from those of $_{Ec}$Rho. The lower steady-state ATPase rate of $_{Bf}$Rho as compared to $_{Ec}$Rho (**Figure 2B**) likely stems from suboptimal, NID-less core components, possibly including the three non-conservative aa substitutions in the Walker motifs (see section above).

$_{Bf}$Rho is a moderately efficient RNA-DNA helicase with 5'→3' directionality

To assess the molecular motor features of $_{Bf}$Rho, we performed helicase experiments with model RNA-DNA hybrid substrates[41, 64]. We observed that $_{Bf}$Rho is able to unwind a relatively long RNA-DNA hybrid (57 base pairs [bp]) substrate bearing a 5'single-stranded RNA (ssRNA) overhang (**Figure 2D**). This substrate is also unwound by $_{Ec}$Rho (at a ~5-fold higher rate) but not by $_{Mtb}$Rho (**Figure 2D**, graph and **Table S2**), which works poorly on long RNA-DNA duplex regions[41]. The $_{Bf}$Rho factor could not unwind a RNA-DNA substrate bearing a

3'ssRNA instead of a 5'ssRNA overhang (**Figure S5A**), supporting that the enzyme behaves as a 5'→3' RNA translocase, as do the $_{Ec}$Rho[64, 65] and $_{Mtb}$Rho[41] enzymes.

The NID of $_{Bf}$Rho is important for productive RNA binding

While $_{Bf}$Rho and $_{Bf}$Rho$_{\beta del}$ unwind the 57 bp duplex bearing a 5'ssRNA overhang with comparable efficiencies, the Δ_{Bf}Rho derivative is a significantly less efficient helicase under our standard experimental conditions (**Figure 2D** and **Table S2**). To determine the origin of this difference, we performed equilibrium-binding measurements with the *Rut*-containing RNA strand (**Figure 2D**, inset) as a substrate. The $_{Bf}$Rho and $_{Bf}$Rho$_{\beta del}$ factors display comparable affinities for this RNA substrate, which are only 2-4 fold lower than affinities determined for $_{Ec}$Rho and $_{Mtb}$Rho (**Table 1**). By contrast, the Δ_{Bf}Rho derivative displays a ~14 fold lower affinity than $_{Bf}$Rho (**Table 1**). Consistent with this binding defect, the duplex unwinding activity of Δ_{Bf}Rho can be strongly stimulated upon increasing the enzyme concentration, while there is not such a dramatic effect for the full-length $_{Bf}$Rho (**Figure S5B**). Taken together, these data demonstrate that the NID contributes significantly to $_{Bf}$Rho binding to RNA.

$_{Bf}$Rho can trigger termination of transcription with E. coli RNAP

To determine if $_{Bf}$Rho is capable to terminate transcription, we performed *in vitro* RDTT experiments with the RNAP of *E. coli* and a DNA template encoding the Rho-dependent terminator λtR1 (**Figure 3A**)[56, 66]. First, we compared the transcription patterns obtained under standard conditions (see methods) with the $_{Ec}$Rho, $_{Mtb}$Rho, Δ_{Mtb}Rho, and $_{Bf}$Rho factors. As described previously[67], transcription termination with $_{Ec}$Rho occurs mostly at three cluster sites (sites I, II and III) in the downstream part of the λtR1 region (**Figure 3B**, lane 2). By contrast,

the termination window of $_{Mtb}$Rho (i.e., the region encompassing all RNA release sites) is much larger and starts upstream from the λtR1 region (**Figure 3B**, lane 4) while the NID-less Δ_{Mtb}Rho factor displays an intermediate RDTT pattern that includes the three cluster sites of $_{Ec}$Rho (lane 6), as expected[41]. The $_{Bf}$Rho factor is also able to induce RDTT with our heterologous transcription system (**Figure 3B**, lane 8) but its RDTT pattern is more akin to the pattern of $_{Ec}$Rho (lane 2) than to that of $_{Mtb}$Rho (lane 4) despite the presence of the large and basic $_{Bf}$Rho-NID. RDTT is less efficient for $_{Bf}$Rho than for $_{Ec}$Rho but is similarly inhibited by BCM, with IC_{50} values of ~13 and ~42 μM for $_{Bf}$Rho and $_{Ec}$Rho, respectively (**Figure S6A**).

The RDTT patterns of $_{Bf}$Rho and $_{Ec}$Rho are also similar in the presence of the NusG factor from *E. coli* ($_{Ec}$NusG). In both cases, $_{Ec}$NusG stimulates RDTT and slightly expands the termination window towards the upstream λtR1 boundary (**Figure 3B**, lanes 3 and 9). The effect of $_{Ec}$NusG is more dramatic with Δ_{Mtb}Rho. In this case, the RDTT window is almost completely shifted upstream from the λtR1 region (**Figure 3B**, lane 7), with an upstream RDTT boundary even closer to the promoter than for the full-length $_{Mtb}$Rho (lane 4). This observation is consistent with the idea that $_{Mtb}$Rho-NID has a role similar to $_{Ec}$NusG[45]. However, we note that $_{Ec}$NusG also slightly changes the termination pattern of $_{Mtb}$Rho, which starts at an 'earlier' (closer to promoter) termination stop point (**Figure 3B**, lane 5). This may be due to indirect effects on the TEC since $_{Ec}$NusG is unable to bind $_{Mtb}$Rho directly[45]. Similar results were obtained with bead-immobilized TECs (**Figure S7**), confirming that the Rho-dependent changes are due to RNA release events and, thus, to RDTT (see also discussion in [33]).

Next, we assessed the transcription termination capabilities of the $_{Bf}$Rho$_{\beta del}$ and Δ_{Bf}Rho mutants. While $_{Bf}$Rho$_{\beta del}$ triggered RDTT in patterns highly similar to those observed for WT $_{Bf}$Rho (**Figure 3C**, compare lanes 1, 2 with lanes 5, 6), Δ_{Bf}Rho was unable to induce termination efficiently, even in the presence of $_{Ec}$NusG (lanes 3, 4). This RDTT deficiency could not be

alleviated upon increasing Δ_{Bf}Rho concentration (**Figure S6B**), suggesting that $_{Bf}$Rho-NID is also important for interactions with the TEC.

We also assayed RDTT with the various Rho variants in the presence of the NusA factor from *E. coli* ($_{Ec}$NusA). The factor delays the RDTT process with $_{Ec}$Rho[24, 68], resulting in a downstream shift of the λtR1 termination window (**Figure 3D**, lane 3). A NusA-dependent delay is also observed with $_{Mtb}$Rho but RDTT still starts closer to the promoter than with $_{Ec}$Rho (**Figure 3D**, lane 5). The delaying effect of $_{Ec}$NusA is more dramatic with Δ_{Mtb}Rho and $_{Bf}$Rho, resulting in nearly complete losses of RDTT signals (**Figure 3D**, lanes 7 and 9). A similar result was obtained with $_{Bf}$Rho$_{\beta del}$ while Δ_{Bf}Rho remained completely inactive even in the presence of both $_{Ec}$NusG and $_{Ec}$NusA (**Figure S6C** and data not shown). The NIDs thus do not shield the $_{Mtb}$Rho and $_{Bf}$Rho factors from the action of $_{Ec}$NusA. Their roles are distinct and not interchangeable with that of $_{Ec}$NusA.

Overall, these data demonstrate that $_{Bf}$Rho is a *bona fide* transcription termination factor. They also show that $_{Bf}$Rho-NID is strictly required for productive interaction with the TEC, which contrasts with the less critical role of $_{Mtb}$Rho-NID in the conditions of our RDTT assay. The two NIDs also differ in their capacities to direct their cognate Rho factor towards early (promoter-proximal) RDTT sites while neither NID appears to work as a NusA surrogate.

$_{Bf}$Rho cannot trigger RDTT at a structurally constrained terminator

To better comprehend the role of $_{Bf}$Rho-NID, we also performed transcription experiments with a DNA template encoding the structurally constrained *pgaA* terminator from the 5'leader region of the *pgaABCD* operon of *E. coli* (**Figure 4A**)[69]. Part of the terminator *Rut* site can be sequestered in a hairpin-like RNA structure that prevents interaction with $_{Ec}$Rho and inhibits

RDTT (**Figure 4B** and **4C**, lane 3)[69]. Binding of the regulatory protein CsrA to the upstream arm of the hairpin prevents formation of the RNA structure, thereby allowing $_{Ec}$Rho to access to the full *Rut* site and to trigger RDTT (**Figure 4B** and **4C**, lane 5)[69]. $_{Ec}$NusG can also switch on the *pgaA* terminator with $_{Ec}$Rho, albeit with a lower efficiency than CsrA (**Figure 4C**, lane 4)[69].

The $_{Bf}$Rho factor is unable to induce efficient RDTT at the *pgaA* terminator, even in the presence of CsrA or $_{Ec}$NusG (**Figure 4C**, lanes 7 and 8). Similar results were obtained with the $_{Bf}$Rho$_{\beta del}$ and Δ_{Bf}Rho mutants (data not shown). By contrast, $_{Mtb}$Rho can trigger highly efficient RDTT, even in absence of CsrA or $_{Ec}$NusG (**Figure 4D**, lanes 7-9). As with the λtR1 terminator (see above), the termination window is larger and starts closer to the promoter than for $_{Ec}$Rho. RDTT with Δ_{Mtb}Rho displays intermediate features, requiring CsrA or NusG for activation but starting at earlier positions than with $_{Ec}$Rho (**Figure 4D**, lanes 10-12). These observations provide the first direct proof for a prominent role of $_{Mtb}$Rho-NID in the handling of structured transcripts by $_{Mtb}$Rho[41, 42]. They also prove that the $_{Bf}$Rho-NID lacks this capacity and may in fact prevent RDTT at the structurally constrained *pgaA* terminator. The last conjuncture is inferred from the lack of RDTT with $_{Bf}$Rho in the presence of CsrA (**Figure 4C**, lane 7). Binding of CsrA to the *pgaA* leader prevents formation of the hairpin-like structure but also occludes part of the mRNA (**Figure 4B**). This steric constraint next to the *Rut* site could be particularly unfavorable to the $_{Bf}$Rho factor, which bears six bulky NIDs around its ring structure (**Figure 1C**, inset). Alternatively, other features of the *pgaA* terminator may not match $_{Bf}$Rho requirements as well as does the λtR1 terminator.

The predicted prion-like NID motif does not significantly affect $_{Bf}$Rho behavior

Although the PLAAC server[70] predicted the presence of a prion-like sequence in $_{Bf}$Rho-NID (**Figure S2**), we did not detect significant differences in RNA binding or enzymatic activities between $_{Bf}$Rho and the $_{Bf}$Rho$_{\beta del}$ mutant, even after several months of storage as 50% glycerol solutions at -20°C. This was surprising given that the amyloidogenic motif in *C. botulinum* Rho ($_{Cb}$Rho) readily promotes the formation of inactive $_{Cb}$Rho aggregates[52, 53]. However, the prion-like sequence is longer in $_{Cb}$Rho (63 vs 33 aa) and yields higher PLAAC prion-likelihood scores (43 vs 23). We thus analyzed the $_{Cb}$Rho and $_{Bf}$Rho sequences with the conceptually distinct PrionW algorithm[71]. PrionW identified a prion-like motif in $_{Cb}$Rho but did not replicate the PLAAC findings for $_{Bf}$Rho (data not shown). Consistent with the PrionW rather than PLAAC prediction, we did not detect formation of prion-like $_{Bf}$Rho aggregates with the amyloid dye Thioflavin-T. Slight decreases of basal Thioflavin-T fluorescence were observed after incubation of $_{Bf}$Rho or $_{Bf}$Rho$_{\beta del}$ for several days at 25°C (F**igure 5A**) rather than the characteristic enhancement of Thioflavin-T fluorescence upon amyloid formation[53]. Using circular dichroism (CD), we also did not observe significant changes in the β-sheet content of $_{Bf}$Rho or $_{Bf}$Rho$_{\beta del}$ after several days of incubation at 25°C (F**igure 5B**). These data argue against a self-sufficient ability of the Q/N-rich NID motif to drive dramatic changes in $_{Bf}$Rho structure, in particular formation of cross-β amyloids. To assess if more subtle changes could affect $_{Bf}$Rho, we performed helicase activity measurements after 2h of pre-incubation of the $_{Bf}$Rho and $_{Bf}$Rho$_{\beta del}$ proteins at 37°C. We assumed that this delay would be sufficient to detect physiologically relevant inactivation effects since it matches the generation time of *B. fragilis* in minimal media[72]. However, the pre-incubation period did not affect the helicase activities of $_{Bf}$Rho and $_{Bf}$Rho$_{\beta del}$ (**Figure S5C**), which are thus comparably stable enzymes. Taken together, these data support that the Q/N-rich NID motif of $_{Bf}$Rho does not readily promote formation of inactive, prion-like aggregates.

DISCUSSION

Although RDTT appears to be widespread in bacteria[38], most studies have focused on $_{Ec}$Rho or on RDTT in *E. coli* (or in close species such as *Salmonella*). However, even in Gram-negative bacteria, RDTT may not always be as critical as in *E. coli* or may take significantly different forms. For instance, a number of Gram-negative bacteria such as *Pseudomonas aeruginosa, Proteus vulgaris*, or *Neisseria meningitides* are highly resistant to Rho's inhibitor BCM[73] or contain BCM biosynthesis gene clusters[74] without carrying Rho mutations known to confer BCM resistance[38]. Rho sequences also often contain large deletions or insertions[38] that could highlight phylogenetically distinct RDTT mechanisms or that could be detrimental and indicative of a lesser importance of RDTT in the corresponding species. Of particular notice is the prevalence of large N-terminal domain insertions (NIDs) in Rho factors from Bacteroidetes[38], a phylum of Gram-negative bacteria containing many low G+C species such as *B. fragilis*. In addition to their unusual sizes, the Bacteroidetes Rho-NIDs also frequently contain putative prion-like sequences (**Figures S2** and **S3**) that may promote functional inactivation, as observed for the phylogenetically distinct $_{Cb}$Rho[52]. Despite these unusual features, the representative $_{Bf}$Rho factor from *B. fragilis* displays a rather orthodox enzymatic behavior when benchmarked against $_{Ec}$Rho. Indeed, $_{Bf}$Rho hydrolyzes NTPs in an RNA-dependent manner, with poly[rC] as preferred cofactor (**Figure 2A-C**); it unwinds RNA-DNA duplexes in an ATP-dependent and 5'→3' directional manner (**Figure 2D** and **Figure S5A**); it can dissociate *E. coli*'s TECs at the same sites than $_{Ec}$Rho and, in doing so, is similarly stimulated by NusG and delayed by NusA (**Figure 3**). In all these biochemical activities, however, $_{Bf}$Rho is less efficient than $_{Ec}$Rho. This may reflect experimental conditions that are more adequate for $_{Ec}$Rho than for $_{Bf}$Rho. For instance, $_{Bf}$Rho may prefer RNA substrates with distinct (e.g. AT-richer) sequence contents or may form termination-prone contacts with *B. fragilis* RNAP that are not fully recreated with *E. coli* RNAP. We note that the side-chain

contacts between $_{Ec}$Rho and $_{Ec}$RNAP that are best resolved in the recent cryoEM structures[23, 24] are poorly conserved in *B. fragilis* and *M. tuberculosis* (**Table S3** and **Figure S8**). Alternatively, the lower biochemical activity of $_{Bf}$Rho may stem from intrinsic features such as, possibly, the non-conservative mutations in its Walker motifs (**Figure S1**) or the steric bulk imposed by the NIDs around the hexamer (~40% of protein mass; **Figure 1C**). A lower Rho activity may be an adaptive trait in a bacterium having a slower metabolism than *E. coli*[72]. In any case, our data support that $_{Bf}$Rho is a *bona fide* transcription termination factor that is also capable to disrupt RNA-DNA duplexes efficiently and may thus contribute to protect the *B. fragilis* genome against transcriptional R-loops, as is the case of $_{Ec}$Rho in *E. coli*[75, 76]. The presence of a large NID is thus not a marker of Rho inactivation in *B. fragilis*. On the contrary, $_{Bf}$Rho-NID stimulates Rho activity, at least in the conditions of our helicase (**Figure 2D**) and transcription termination (**Figure 3**) assays. In line with this hereby-demonstrated enzymatic proficiency, the $_{Bf}$Rho factor seems to be essential in *B. fragilis*. This is inferred from transposon inactivation experiments that failed to detect viable transpositions in the *rho* gene[47] and from the BCM sensitivity of tens of *B. fragilis* strains[46]. Our data support that this BCM sensitivity stems from direct inhibition of the $_{Bf}$Rho factor (**Figure S6**) and, thus, that RDTT is a valid pharmacological target in the opportunistic *B. fragilis* pathogen.

Until now, only NID-bearing Rho factors from the high G+C *Mic. luteus* ($_{Mic}$Rho) and *M. tuberculosis* species had been characterized[39-42, 45]. In both cases, the NID increases Rho affinity for RNA, favors promoter-proximal RDTT in a manner similar to NusG, and stimulates RDTT with structurally constrained transcripts (**Figure 4D**)[40, 41]. In the case of $_{Bf}$Rho, the NID also significantly increases Rho affinity for RNA (**Table 1**) but does not promote promoter-proximal RDTT (**Figure 3B**) or RDTT with a structurally constrained transcript (**Figure 4C**). Moreover, $_{Bf}$Rho-NID appears to be strictly required for termination activity (**Figure 3C**) whereas both $_{Mic}$Rho and $_{Mtb}$Rho remain able to trigger promoter-distal RDTT when deprived

of their NID (**Figure 3B**)[40, 41]. This may be due to the comparably lower affinity of Δ_{Bf}Rho for RNA (**Table 1**)[40], a feature that is by itself surprising given that the primary RNA binding site (PBS) domain contains disruptive indels in $_{Mic}$Rho and $_{Mtb}$Rho but not in $_{Bf}$Rho (**Figure S1**)[39]. One possible explanation is that the poorly conserved, N-terminal NHB domain (**Figure 1C** and **Figure S1**) is able to mitigate the lack of NID in $_{Mic}$Rho and $_{Mtb}$Rho but not in $_{Bf}$Rho. The NHB domain has been proposed to electrostatically help Rho's PBS to catch the RNA chain[77], a role that could be strictly devoted to the basic NID in $_{Bf}$Rho. Alternatively, the NID may form contacts with RNAP that are especially critical for $_{Bf}$Rho. In support of this proposal is the observation that Δ_{Bf}Rho retains ATPase and helicase activities (**Figure 2**) but has no significant transcription termination activity even at high concentration (**Figure 3C** and **Figure S6B**).

It is worth considering the hypothetical roles of $_{Bf}$Rho-NID in light of the recent cryoEM structures of the $_{Ec}$Rho:$_{Ec}$RNAP complex[23, 24]. The structures support a model for RDTT whereby Rho hitchhikes on ribosome-free RNAP while scanning the emerging transcript for a *Rut* site (**Figure 1A**). The orientation of the $_{Ec}$Rho hexamer in the structures suggests a rather convoluted and sterically constrained path of the transcript at the Rho:RNAP interface and around Rho (**Figure 1A**)[23, 24]. It is envisioned that *Rut* recognition by Rho's PBS at the Rho:RNAP interface triggers allosteric destabilization of the TEC by the bound Rho hexamer[23, 24]. Importantly, the positions of NID insertions (corresponding to residues 43-45 in $_{Ec}$Rho) are located on the outer edge of the Rho ring facing RNAP (**Figure 6A**, residues in purple). Assuming that this Rho:RNAP configuration is conserved in *B. fragilis*, the NIDs may be in ideal positions to make peripheral contacts with RNAP, thereby extending the Rho:RNAP interface (**Figure 6A**, cartoon). This could be critical to compensate for otherwise poorly conserved Rho:RNAP interactions (**Table S3** and **Figure S8**)[23]. The NIDs may also be adequately located on the ring periphery to help guide the transcript from the RNA exit channel of RNAP to the inner Rho channel and, possibly, to extrude the RNA 5'-end from the

Rho:RNAP complex (**Figure 6A**). A similar 'guiding' role has been ascribed to $_{Ec}$NusA[24] but our RDTT data (**Figure 3D** and **Figure S6C**) argue against interchangeability between NIDs and $_{Ec}$NusA. The cooperative capture of *Rut* sequences by both the NIDs and $_{Bf}$Rho's PBS is also possible; however, the physical separation between 'external' NIDs and a 'buried' PBS (**Figure 6A**) makes it less easily envisioned than for the classical RDTT model of 'tethered tracking' (**Figure 6B**)[19]. Further structural characterization of the Rho:RNAP complex, including from species that are phylogenetically distinct from *E. coli*, will be needed to clarify these important aspects.

The *B. fragilis rho* gene is significantly GC-richer than the whole genome (48.2% vs 43.1%), a feature that usually stresses the importance of the corresponding gene and its belonging to the core genome[78]. Intriguingly, the upstream section of the gene (until the end of the NID-encoding region) displays a strong sequence bias when compared to the downstream section (**Figure 6C**). Similar sequence biases were found in other Bacteroidetes *rho* genes but not in the *rho* gene of *M. tuberculosis* (**Figure 6D** and data not shown). These observations suggest that the unusual NTD of $_{Bf}$Rho is the product of relatively recent insertion(s) of foreign DNA into the *rho* gene. This may have helped tune $_{Bf}$Rho activity to the AT-rich transcriptome of *B. fragilis* (e.g. by increasing binding to less C-rich *Rut* sites) or to specifics of its transcription machinery (**Table S3** and **Figure S8**). Alternatively, the acquisition of an unusually long NTD-encoding sequence (**Figure 1B**) could have been a source of new, evolutionary advantageous properties. Notably, the presence of a putative prion-like Q/N-rich sequence in $_{Bf}$Rho-NID (**Figure S2**) suggested that $_{Bf}$Rho could form amyloid structures, as was observed previously for the $_{Cb}$Rho factor of *C. botulinum*[52]. Functional amyloids support a diversity of advantageous functions in microorganisms, from the structural scaffolding of biofilms to the regulation of the life cycle or response to stresses[79]. However, we were unable to detect formation of $_{Bf}$Rho amyloids or significant structural/activity differences between

$_{Bf}$Rho and the $_{Bf}$Rho$_{\beta del}$ variant (**Figures 5 and S5C**). In retrospect, this is consistent with prion prediction scores that are lower for $_{Bf}$Rho than for $_{Cb}$Rho (see results). Notwithstanding, we cannot exclude that $_{Bf}$Rho promptly self-assembles only under specific conditions. Many 'facultative' functional amyloids have been described, which adopt the amyloid-forming conformation only upon a specific trigger such as a pH change, post-translational modification, cleavage of the peptide chain, or presence of a metal ion, nucleic acid, or protein 'nucleator'[79]. A minimalistic amyloidogenic core sequence such as the Q/N-rich sequence found in $_{Bf}$Rho-NID (**Figure S2**) might help keep such structural transition 'conditional'. Further work will be needed to explore this intriguing possibility.

EXPERIMENTAL PROCEDURES

Materials. Chemicals and enzymes were purchased from Sigma-Aldrich and New England Biolabs, respectively. BCM was purchased from Santa Cruz Biotechnology. Nucleoside triphosphates were purchased from GE-Healthcare while radionucleotides were from PerkinElmer. Synthetic oligonucleotides were obtained from Eurogentec. Polynucleotide fragment (> 300 nt) stocks were prepared as described previously [59] from polynucleotide batches obtained from Santa Cruz Biotechnology and Midland Certified Reagent Company (Midland, Texas). The DNA templates containing the λtR1 and *pgaA* terminators used in transcription termination experiments with *E. coli* RNAP were prepared as described previously[33, 69]. RNA substrates were obtained by *in vitro* transcription of PCR amplicons with T7 RNAP and purified by denaturing polyacrylamide gel electrophoresis (PAGE), as described[59]. Genomic DNA from *B. fragilis* type strain NCTC 9343 was obtained from the DSMZ repository (Germany). Plasmid for over-expression of WT $_{Bf}$Rho (pET28b-$_{Bf}$Rho) was prepared by subcloning the $_{Bf}$Rho coding sequence within plasmid pET28b (Novagen). To this

end, the $_{Bf}$Rho coding sequence was PCR amplified from genomic DNA using primers BfRho-F and BfRho-R (**Table S1**), digested with the KpnI and XhoI enzymes, and then subcloned within the KpnI and XhoI restriction sites of the pET28b plasmid. Plasmids for overexpression of the Δ_{Bf}Rho and $_{Bf}$Rho$_{\beta del}$ variants were engineered from the pET28b-$_{Bf}$Rho plasmid using the NEBuilder HiFi DNA Assembly kit (NEB) and, respectively, primers ΔBfRho-F and ΔBfRho-R or βdelBfRho-F and βdelBfRho-R (**Table S1**), following manufacturer's instructions. The sequences of DNA templates and plasmids were verified by capillary DNA sequencing (Genoscreen, France). Plasmid pET28b-NusA for overexpression of the *E.coli* NusA protein with a C-terminal His6 tag was custom synthesized by Genscript.

Preparation of proteins. The $_{Ec}$NusG, $_{Ec}$Rho, $_{Mtb}$Rho and Δ_{Mtb}Rho proteins were prepared and purified as described previously[41, 59]. The $_{Bf}$Rho, Δ_{Bf}Rho, and $_{Bf}$Rho$_{\beta del}$ proteins were prepared following similar procedures. Briefly, they were over-expressed as N-terminal His6 tag fusions in Rosetta 2(DE3) cells (Merck-Millipore) harboring the appropriate pET28b plasmid derivative. They were purified by affinity chromatography on a HisTrap HP column (GE-Healthcare) with a 10-500 mM imidazole gradient and by ion exchange chromatography on a SP- ($_{Bf}$Rho and $_{Bf}$Rho$_{\beta del}$) or Q- (Δ_{Bf}Rho) sepharose FF column (GE-Healthcare) with a 0.15-1M NaCl gradient. The NusA protein was overexpressed in standard BL21(DE3) cells harboring the pET28b-NusA plasmid and purified by HisTrap chromatography, as described[80]. Protein identities and purities were checked by liquid chromatography coupled to HR-MS at the CBM mass spectrometry facility, Orléans, France. All proteins were stored at -20°C as micromolar solutions in 100 mM KCl, 10 mM Tris-HCl, pH 7.9, 0.1 mM EDTA, 0.1 mM DTT, and 50% (v/v) glycerol. Concentrations of Rho factors are expressed in hexamers throughout the manuscript.

Preparation of duplex substrates. Duplexes were assembled by mixing 10 pmoles of ^{32}P-end labeled RNA transcript with 12 pmoles of complementary oligonucleotide in annealing buffer (150 mM potassium acetate, 20 mM HEPES, pH 7.5, 1 mM EDTA). Mixtures were heated at 95°C and slowly cooled to 20°C. Duplexes were then purified by native 7 % PAGE and stored at -20°C in helicase buffer (20 mM HEPES, pH 7.5, 0.1 mM EDTA, 0.5 mM DTT, and 150 mM potassium acetate), as described[59].

NTP hydrolysis assay. The NTP hydrolysis activities of the Rho enzymes were determined as described previously with a thin layer chromatography-based ^{32}P radiometric assay[64] or with the EnzCheck Phosphate Assay kit (Molecular Probes)[56, 81]. Control experiments confirmed that both assays yield comparable results. Reaction mixtures contained 20 nM Rho, 1 mM NTP and the indicated concentration of RNA cofactor (e.g. poly[rC] at a standard concentration of 10 µM in rC residues) in NTPase buffer (1 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5, 0.1 mM sodium azide). Experiments were performed at 37°C.

Equilibrium binding assays. Equilibrium Rho-RNA dissociation constants were determined using a filter-binding assay, as described previously[64, 82]. Briefly, ~10 fmoles of ^{32}P-labeled RNA substrate were mixed with various amounts of Rho in 100 µl of binding buffer (20 mM HEPES, pH 7.5, 0.1 mM EDTA, 0.5 mM DTT, 150 mM potassium acetate, and 20 µg/ml BSA). After incubation for 10 min at 30°C, the samples were filtered through stacked [top] nitrocellulose (Amersham Protran) and [bottom] cationic nylon (Pall Biodyne B) membranes using a Bio-dot SF apparatus (Biorad). The fractions of free and Rho-bound RNA (retained on the nylon and nitrocellulose membranes, respectively) as a function of Rho concentration were then determined by phosphorimaging of the membranes using a Typhoon FLA9500 imager and dedicated ImageQuant TL v8.1 software (GE healthcare).

Duplex unwinding assay. Helicase reactions were performed as described previously[83] with ^{32}P-labeled duplex substrates. Briefly, 5 nM substrate was mixed with 20 nM Rho in helicase buffer and incubated for 3 min at 30°C. Then, 1 mM $MgCl_2$, 1 mM ATP, and 400 nM oligo trap (unlabeled oligonucleotide having the same sequence than the released 'reporter' strand) were added to the helicase mixture before further incubation at 30°C. Reaction aliquots were taken at various times and mixed with two volumes of quench buffer (30 mM EDTA, 0.75% SDS, 150 mM sodium acetate, 6 % Ficoll-400) before being loaded on 8% polyacrylamide gels that contained 1X TBE and 0.5% SDS. Detection and quantification of gel bands were performed by phosphorimaging with a Typhoon FLA9500 imager as described[81].

Transcription termination experiments. Standard transcription termination experiments were performed as described previously[33]. Briefly, DNA template (0.1 pmol), *E. coli* RNAP (0.45 pmol), Rho (0 or 1.4 pmol), and Superase-In (0.5U/µL; Ambion) were mixed in 18 µL of transcription buffer (40 mM Tris-HCl, pH 8, 0, 5 mM $MgCl_2$, 1,5 mM DTT, and 100 or 150 mM KCl for experiments with the λtR1 or *pgaA* terminator, respectively) and incubated for 10 min at 37°C. Whenever necessary (see text and figure legends), NusG (2.8 pmol), NusA (2.8 pmol), CsrA (2.8 pmol), and/or BCM (0-3 nmol) were also added to the premixes. Then, 2 µL of initiation solution (2 mM ATP, GTP, and CTP, 0.2 mM UTP, 2.5 µCi/µL of ^{32}P-αUTP, and 250 µg/mL of rifampicin in transcription buffer) were added to the reaction mixtures before further incubation for 20 min at 37°C. Transcription reactions were stopped with 4 µL of EDTA (0.5 M), 6 µL of tRNA (0.25 mg/mL), and 80µL of sodium acetate (0.42 M) before precipitation at -20°C with 330 µL of ethanol. Reaction pellets were dissolved in denaturing loading buffer (95% formamide, 5 mM EDTA), and analyzed by denaturing 7% polyacrylamide gel electrophoresis (PAGE) and Typhoon FLA9500 phosphorimaging (GE-Healthcare). Gamma

settings of gel images have been optimized to facilitate the comparison of termination signals[41]. Transcription termination efficiencies with the λtR1 terminator were estimated from gel band intensities after normalization for distinct ^{32}P-αU incorporation levels in the various transcript species, as described previously[56].

Circular Dicroism (CD). Experiments were performed with 10µM (in monomers) solutions of the $_{Bf}$Rho and $_{Bf}$Rho$_{βdel}$ proteins in CD buffer (100 mM NaCl, 10 mM sodium phosphate, pH 7.5). The solutions were incubated in a dry bath at 25°C in the dark. At various times, CD spectra of the protein solutions were recorded at 20°C on a Jasco J-810 spectropolarimeter from 260 to 200 nm using a 1-mm quartz cell, a scan rate of 50 nm/min, a response time of 4 s, a bandwidth of 2 nm, and a resolution of 1 nm. Each CD spectrum represents the average of five scans, with the buffer subtracted.

Amyloid detection assay. The assay was performed with the Thioflavin-T dye as described previously[53], with minor modifications. Briefly, the $_{Bf}$Rho and $_{Bf}$Rho$_{βdel}$ proteins at a concentration of 20µM (monomers) in CD buffer were incubated at 25°C in the dark. Aliquots were withdrawn at various times and mixed gently with and equal volume of a 50 µM aqueous solution of Thioflavin-T dye. Fluorescence emission spectra were recorded on a CLARIOstar Plus microplate reader (BMG labtech) set for excitation at 440 nm. Background fluorescence of the Thioflavin-T dye in CD buffer was substracted from the spectra.

CONFLICT OF INTEREST STATEMENT

The authors declare not conflict of interest.

AKNOWLEDGMENTS

We warmly thank Franck Coste for his help with circular dichroism and for insightful discussions along the project. We also thank François D'Heygère for his participation in the early stages of the project. This work was supported by the French Agence Nationale de la Recherche (ANR-15-CE11-0024-02) and by CNRS core funding. I.S. is a recipient of a doctoral fellowship from Région Centre-Val de Loire.

FIGURE LEGENDS

Figure 1: Rho-dependent termination of transcription (RDTT). **(A)** Schematic depiction of transcription-translation coupling in *E. coli* (top) and how the lack of ribosome allows Rho binding to RNAP[23] and, eventually, to a *Rut* site within the nascent transcript (bottom). In 5'untranslated regions (UTRs), Rho may first bind RNAP and trigger RDTT. If the 5'UTR is too short or does not encode a *Rut* site promoting RDTT, a ribosome may displace Rho from RNAP once the ribosome-binding site (RBS) has been transcribed. For sake of clarity, cofactors NusA and NusG are not shown and only three subunits of the Rho hexamer are depicted with the PBS sites facing RNAP in orange (diagram not to scale). Once bound to the *Rut* site (blue segment), the Rho hexamer presumably triggers allosteric destabilization of the TEC[23-25]. Full disruption of the TEC likely requires closure of the Rho ring around the transcript and ATP-dependent RNA translocation as depicted (small back arrows)[23]. **(B)** Distribution of NID lengths from 1259 phylogenetically diverse Rho factors[38]. NIDs < 20 aa are not shown. **(C)** Diagram showing the arrangement of key motifs within the primary sequence of Rho subunits (numbering is for $_{Bf}$Rho; role of specific residues is based on information obtained with $_{Ec}$Rho). *Rut* recruitment is ensured primarily by PBS residues (orange arrows)[84] within the RNA binding domain (RBD) of Rho; the N-terminal helix bundle (NHB) domain may help[77]. The large NID of $_{Bf}$Rho (sequence in **Figure S1**) is depicted by a purple triangle. SBS residues translocating RNA though the $_{Ec}$Rho hexamer channel[49] are depicted by green arrows. A model of the $_{Bf}$Rho Rho hexamer based on the $_{Ec}$Rho structure (PDB #3ICE) with purple spheres representing the NIDs (one per subunit, not to scale) is shown inset. Note that the NIDs are probably unstructured in absence of cofactors (see main text).

Figure 2: Standard 'molecular motor' properties of the $_{Bf}$Rho factor. **(A)** The steady-state ATP hydrolysis of $_{Bf}$Rho is stimulated by poly[rC]. **(B)** ATPase activity as a function of RNA ligand (with $_{Bf}$Rho; left graph) or Rho factor (with poly[rC]; right graph). **(C)** The identity of the NTP hardly affects the NTP hydrolysis activity of $_{Bf}$Rho (with poly[rC]). Unless specified otherwise, the concentrations of NTP and RNA polymer in the NTPase assays were 1 mM and 10 µM (in nucleotide residues), respectively. Graphs show mean ± error values from at least two independent experiments. **(D)** Duplex unwinding activity. The $_{Bf}$Rho helicase is able to unwind a 57 bp-long RNA-DNA duplex bearing a 5'-single stranded RNA tail containing a synthetic *Rut* site optimized for $_{Ec}$Rho[64], as depicted. Unwinding reactions were performed under standard conditions with 20 nM Rho and 5 nM of the RNA-DNA substrate. The graph shows unwinding time courses (fraction means from three independent experiments) for the various factors identified on the right. A representative 9% PAGE gel illustrates the helicase activity of $_{Bf}$Rho.

Figure 3: The $_{Bf}$Rho factor triggers RDTT with *E. coli* RNAP and the prototypical λtR1 terminator. **(A)** Schematic of the DNA template used. **(B)** The termination windows of $_{Bf}$Rho and $_{Ec}$Rho are similar. **(C)** The NID of $_{Bf}$Rho is critical for RDDT activity. **(D)** RDTT with either the $_{Mtb}$Rho or $_{Bf}$Rho factor is sensitive to the delaying effect of $_{Ec}$NusA. Representative denaturing PAGE gels show the products of transcription termination experiments performed under standard conditions with *E. coli* RNAP and the λtR1 template in the presence of various Rho factors and with or without $_{Ec}$NusA or $_{Ec}$NusG (indicated above gel lanes).

Figure 4: RDTT with the structurally constrained *pgaA* terminator of *E. coli*. **(A)** Schematic of the DNA template used. The template encodes the full *pgaA* 5'leader, which includes the CsrA-

regulated Rho-dependent *pgaA* terminator[69]. **(B)** A hairpin-like structure shields part of the *pgaA Rut* site. Binding of the CsrA protein prevents formation of the hairpin structure and allows $_{Ec}$Rho to interact productively with the *pgaA* transcript[69]. **(C)** The $_{Bf}$Rho factor cannot terminate transcription at the *pgaA* terminator. **(D)** $_{Mtb}$Rho triggers early, promoter proximal RDTT with the *pgaA* terminator even in absence of CsrA or NusG.

Figure 5: The $_{Bf}$Rho factor does not spontaneously form β-amyloids. Thioflavin-T fluorescence emission **(A)** and circular dichroism **(B)** spectra of the $_{Bf}$Rho and $_{Bf}$Rho$_{\beta del}$ proteins after 0 to 7 days of incubation at 25°C in 100 mM NaCl, 10 mM Phosphate, pH 7.5 buffer. The β-sheet content of the proteins was predicted from the CD spectra by using the BETSEL server[85].

Figure 6: Role of the $_{Bf}$Rho-NID. **(A)** Model based on the recent cryoEM structures of the $_{Ec}$Rho:RNAP complex[23, 24]. Residues 43-45 of $_{Ec}$Rho are shown in magenta on the filled-sphere structure of the complex (PDB #6XAS) to appreciate where NIDs would be protruding from the hexameric $_{Bf}$Rho ring. In the model, the NIDs would contribute to stabilize the Rho:RNAP complex and, possibly, help keep the RNA chain into the path required for transcript scanning by the Rho factor until a *Rut* site is detected. **(B)** In the classical RDTT model, The NIDs would cooperate with the PBS to anchor Rho to the *Rut* site and to stabilize the *Rut*-Rho interaction during ATP-dependent 'tethered tracking' of the hexamer motor along RNA towards the RNAP. At present, it cannot be totally excluded that the CryoEM-like and 'tethered tracking' models represent distinct, physiologically relevant pathways[23, 24]. **(C & D)** The graphs show the distributions of A and C residues versus G and T residues along the sequences of the genes encoding $_{Bf}$Rho **(C)** and $_{Mtb}$Rho **(D)**. Percentages were calculated with a 20 nt sliding window.

Note that the AC>GT sequence bias in the upstream portion of the *Bf**rho* gene is not detected with a standard %GC vs. %AT plot (**Figure S9**).

REFERENCES

[1] Boudvillain M, Figueroa-Bossi N, Bossi L. Terminator still moving forward: expanding roles for Rho factor. Curr Opin Microbiol. 2013;16:118-24.

[2] Ray-Soni A, Bellecourt MJ, Landick R. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. Annual review of biochemistry. 2016;85:319-47.

[3] Grylak-Mielnicka A, Bidnenko V, Bardowski J, Bidnenko E. Transcription termination factor Rho: a hub linking diverse physiological processes in bacteria. Microbiology. 2016;162:433-47.

[4] Turnbough CL, Jr. Regulation of Bacterial Gene Expression by Transcription Attenuation. Microbiol Mol Biol Rev. 2019;83.

[5] Bastet L, Turcotte P, Wade JT, Lafontaine DA. Maestro of regulation: Riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay. RNA biology. 2018;15:679-82.

[6] Roberts JW. Mechanisms of Bacterial Transcription Termination. J Mol Biol. 2019;431:4030-9.

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

[8] Mader U, Nicolas P, Depke M, Pane-Farre J, Debarbouille M, van der Kooi-Pol MM, et al. Staphylococcus aureus Transcriptome Architecture: From Laboratory to Infection-Mimicking Conditions. PLoS Genet. 2016;12:e1005962.

[9] Botella L, Vaubourgeix J, Livny J, Schnappinger D. Depleting Mycobacterium tuberculosis of the transcription termination factor Rho causes pervasive transcription and rapid death. Nature communications. 2017;8:14731.

[10] Magan A, Amman F, El-Isa F, Hartl N, Shamovsky I, Nudler E, et al. iRAPs curb antisense transcription in E. coli. Nucleic Acids Res. 2019;47:10894-905.

[11] Kriner MA, Sevostyanova A, Groisman EA. Learning from the Leaders: Gene Regulation by the Transcription Termination Factor Rho. Trends Biochem Sci. 2016;41:690-9.

[12] Chen J, Morita T, Gottesman S. Regulation of Transcription Termination of Small RNAs and by Small RNAs: Molecular Mechanisms and Biological Functions. Front Cell Infect Microbiol. 2019;9:201.

[13] Sedlyarova N, Shamovsky I, Bharati BK, Epshtein V, Chen J, Gottesman S, et al. sRNA-Mediated Control of Transcription Termination in E. coli. Cell. 2016;167:111-21 e13.

[14] Ojha S, Jain C. Dual-level autoregulation of the E. coli DeaD RNA helicase via mRNA stability and Rho-dependent transcription termination. RNA. 2020;26:1160-9.

[15] Bossi L, Figueroa-Bossi N, Bouloc P, Boudvillain M. Regulatory interplay between small RNAs and transcription termination factor Rho. Biochim Biophys Acta Gene Regul Mech. 2020;1863:194546.

[16] Gowrishankar J, Harinarayanan R. Why is transcription coupled to translation in bacteria? Mol Microbiol. 2004;54:598-603.

[17] Washburn RS, Gottesman ME. Transcription termination maintains chromosome integrity. Proc Natl Acad Sci U S A. 2010:doi: 10.1073/pnas.1009564108

[18] Rabhi M, Rahmouni AR, Boudvillain M. Transcription termination factor Rho: a ring-shaped RNA helicase from bacteria. In: Jankowsky E, editor. RNA helicases. Cambridge (UK): RSC Publishing; 2010. p. 243-71.

[19] Steinmetz EJ, Platt T. Evidence supporting a tethered tracking model for helicase activity of Escherichia coli Rho factor. Proc Natl Acad Sci U S A. 1994;91:1401-5.

[20] Koslover DJ, Fazal FM, Mooney RA, Landick R, Block SM. Binding and translocation of termination factor rho studied at the single-molecule level. J Mol Biol. 2012;423:664-76.

[21] Gocheva V, Le Gall A, Boudvillain M, Margeat E, Nollmann M. Direct observation of the translocation mechanism of transcription termination factor Rho. Nucleic Acids Res. 2015;43:2367-77.

[22] Thomsen ND, Lawson MR, Witkowsky LB, Qu S, Berger JM. Molecular mechanisms of substrate-controlled ring dynamics and substepping in a nucleic acid-dependent hexameric motor. Proc Natl Acad Sci U S A. 2016;113:E7691-E700.

[23] Said N, Hilal T, Sunday ND, Khatri A, Burger J, Mielke T, et al. Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase rho. Science. 2021;371.

[24] Hao Z, Epshtein V, Kim KH, Proshkin S, Svetlov V, Kamarthapu V, et al. Pre-termination Transcription Complex: Structure and Function. Mol Cell. 2021;81:281-92 e8.

[25] Epshtein V, Dutta D, Wade J, Nudler E. An allosteric mechanism of Rho-dependent transcription termination. Nature. 2010;463:245-9.

[26] Sullivan SL, Gottesman ME. Requirement for E. coli NusG protein in factor-dependent transcription termination. Cell. 1992;68:989-94.

[27] Shashni R, Qayyum MZ, Vishalini V, Dey D, Sen R. Redundancy of primary RNA-binding functions of the bacterial transcription terminator Rho. Nucleic Acids Res. 2014;42:9677-90.

[28] Bossi L, Ratel M, Laurent C, Kerboriou P, Camilli A, Eveno E, et al. NusG prevents transcriptional invasion of H-NS-silenced genes. PLoS Genet. 2019;15:e1008425.

[29] Valabhoju V, Agrawal S, Sen R. Molecular Basis of NusG-mediated Regulation of Rho-dependent Transcription Termination in Bacteria. J Biol Chem. 2016;291:22386-403.

[30] Lawson MR, Ma W, Bellecourt MJ, Artsimovitch I, Martin A, Landick R, et al. Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. Mol Cell. 2018;71:911-22 e4.

[31] Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, Nudler E. Termination Factor Rho and Its Cofactors NusA and NusG Silence Foreign DNA in E. coli. Science. 2008;320:935-8.

[32] Peters JM, Mooney RA, Kuan PF, Rowland JL, Keles S, Landick R. Rho directs widespread termination of intragenic and stable RNA transcription. Proc Natl Acad Sci U S A. 2009;106:15406-11.

[33] Nadiras C, Eveno E, Schwartz A, Figueroa-Bossi N, Boudvillain M. A multivariate prediction model for Rho-dependent termination of transcription. Nucleic Acids Res. 2018;46:8245-60.

[34] Bidnenko V, Nicolas P, Grylak-Mielnicka A, Delumeau O, Auger S, Aucouturier A, et al. Termination factor Rho: From the control of pervasive transcription to cell fate determination in Bacillus subtilis. PLoS Genet. 2017;13:e1006909.

[35] Nagel A, Michalik S, Debarbouille M, Hertlein T, Gesell Salazar M, Rath H, et al. Inhibition of Rho Activity Increases Expression of SaeRS-Dependent Virulence Factor Genes in Staphylococcus aureus, Showing a Link between Transcription Termination, Antibiotic Action, and Virulence. mBio. 2018;9.

[36] Trzilova D, Anjuwon-Foster BR, Torres Rivera D, Tamayo R. Rho factor mediates flagellum and toxin phase variation and impacts virulence in Clostridioides difficile. PLoS pathogens. 2020;16:e1008708.

[37] Takemoto N, Tanaka Y, Inui M. Rho and RNase play a central role in FMN riboswitch regulation in Corynebacterium glutamicum. Nucleic Acids Res. 2015;43:520-9.

[38] D'Heygere F, Rabhi M, Boudvillain M. Phyletic distribution and conservation of the bacterial transcription termination factor Rho. Microbiology. 2013;159:1423-36.

[39] Nowatzke WL, Richardson JP. Characterization of an unusual Rho factor from the high G + C gram-positive bacterium Micrococcus luteus. J Biol Chem. 1996;271:742-7.

[40] Nowatzke WL, Burns CM, Richardson JP. Function of the novel subdomain in the RNA binding domain of transcription termination factor Rho from Micrococcus luteus. J Biol Chem. 1997;272:2207-11.

[41] D'Heygere F, Schwartz A, Coste F, Castaing B, Boudvillain M. ATP-dependent motor activity of the transcription termination factor Rho from Mycobacterium tuberculosis. Nucleic Acids Res. 2015;43:6099-111.

[42] Mitra A, Misquitta R, Nagaraja V. Mycobacterium tuberculosis Rho Is an NTPase with Distinct Kinetic Properties and a Novel RNA-Binding Subdomain. PLoS One. 2014;9:e107474.

[43] Burns CM, Nowatzke WL, Richardson JP. Activation of Rho-dependent transcription termination by NusG. Dependence on terminator location and acceleration of RNA release. J Biol Chem. 1999;274:5245-51.

[44] Nehrke KW, Zalatan F, Platt T. NusG alters rho-dependent termination of transcription in vitro independent of kinetic coupling. Gene Expr. 1993;3:119-33.

[45] Kalarickal NC, Ranjan A, Kalyani BS, Wal M, Sen R. A bacterial transcription terminator with inefficient molecular motor action but with a robust transcription termination function. J Mol Biol. 2010;395:966-82.

[46] Watt B, Brown FV. The in-vitro activity of bicozamycin against anaerobic bacteria of clinical interest. The Journal of antimicrobial chemotherapy. 1983;12:549-53.

[47] Veeranagouda Y, Husain F, Tenorio EL, Wexler HM. Identification of genes required for the survival of B. fragilis using massive parallel sequencing of a saturated transposon mutant library. BMC Genomics. 2014;15:429.

[48] Bogden CE, Fass D, Bergman N, Nichols MD, Berger JM. The structural basis for terminator recognition by the Rho transcription termination factor. Mol Cell. 1999;3:487-93.

[49] Thomsen ND, Berger JM. Running in reverse: the structural basis for translocation polarity in hexameric helicases. Cell. 2009;139:523-34.

[50] Balasubramanian K, Stitt BL. Evidence for amino acid roles in the chemistry of ATP hydrolysis in Escherichia coli Rho. J Mol Biol. 2010;404:587-99.

[51] Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. XtalPred: a web server for prediction of protein crystallizability. Bioinformatics. 2007;23:3403-5.

[52] Yuan AH, Hochschild A. A bacterial global regulator forms a prion. Science. 2017;355:198-201.

[53] Pallares I, Iglesias V, Ventura S. The Rho Termination Factor of Clostridium botulinum Contains a Prion-Like Domain with a Highly Amyloidogenic Core. Front Microbiol. 2015;6:1516.

[54] Martinez A, Opperman T, Richardson JP. Mutational analysis and secondary structure model of the RNP1-like sequence motif of transcription termination factor Rho. J Mol Biol. 1996;257:895-908.

[55] Wei RR, Richardson JP. Mutational changes of conserved residues in the Q-loop region of transcription factor Rho greatly reduce secondary site RNA-binding. J Mol Biol. 2001;314:1007-15.

[56] Rabhi M, Gocheva V, Jacquinot F, Lee A, Margeat E, Boudvillain M. Mutagenesis-based evidence for an asymmetric configuration of the ring-shaped transcription termination factor Rho. J Mol Biol. 2011;405:497-518.

[57] Skordalakes E, Brogan AP, Park BS, Kohn H, Berger JM. Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. Structure. 2005;13:99-109.

[58] Lawson MR, Dyer K, Berger JM. Ligand-induced and small-molecule control of substrate loading in a hexameric helicase. Proc Natl Acad Sci U S A. 2016;113:13714-9.

[59] Boudvillain M, Walmacq C, Schwartz A, Jacquinot F. Simple enzymatic assays for the in vitro motor activity of transcription termination factor Rho from Escherichia coli. Methods Mol Biol. 2010;587:137-54.

[60] Shi Y, Mowery RA, Ashley J, Hentz M, Ramirez AJ, Bilgicer B, et al. Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. Protein Sci. 2012;21:1197-209.

[61] Seifried SE, Easton JB, von Hippel PH. ATPase activity of transcription-termination factor rho: functional dimer model. Proc Natl Acad Sci U S A. 1992;89:10454-8.

[62] Richardson JP. Activation of rho protein ATPase requires simultaneous interaction at two kinds of nucleic acid-binding sites. J Biol Chem. 1982;257:5760-6.

[63] Ingham CJ, Hunter IS, Smith MC. Isolation and sequencing of the rho gene from Streptomyces lividans ZX7 and characterization of the RNA-dependent NTPase activity of the overexpressed protein. J Biol Chem. 1996;271:21803-7.

[64] Walmacq C, Rahmouni AR, Boudvillain M. Influence of substrate composition on the helicase activity of transcription termination factor Rho: reduced processivity of Rho hexamers during unwinding of RNA-DNA hybrid regions. J Mol Biol. 2004;342:403-20.

[65] Brennan CA, Dombroski AJ, Platt T. Transcription termination factor rho is an RNA-DNA helicase. Cell. 1987;48:945-52.

[66] Rabhi M, Espeli O, Schwartz A, Cayrol B, Rahmouni AR, Arluison V, et al. The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. Embo J. 2011;30:2805-16.

[67] Lau LF, Roberts JW, Wu R. Transcription terminates at lambda tR1 in three clusters. Proc Natl Acad Sci U S A. 1982;79:6171-5.

[68] Qayyum MZ, Dey D, Sen R. Transcription Elongation Factor NusA Is a General Antagonist of Rho-dependent Termination in Escherichia coli. J Biol Chem. 2016;291:8090-108.

[69] Figueroa-Bossi N, Schwartz A, Guillemardet B, D'Heygere F, Bossi L, Boudvillain M. RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. Genes Dev. 2014;28:1239-51.

[70] Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics. 2014;30:2501-2.

[71] Sabate R, Rousseau F, Schymkowitz J, Ventura S. What Makes a Protein Sequence a Prion? PLOS Computational Biology. 2015;11:e1004013.

[72] Frantz JC, McCallum RE. Growth yields and fermentation balance of Bacteroides fragilis cultured in glucose-enriched medium. J Bacteriol. 1979;137:1263-70.

[73] Nishida M, Mine Y, Matsubara T, Goto S, Kuwahara S. Bicyclomycin, a new antibiotic. 3. In vitro and in vivo antimicrobial activity. J Antibiot (Tokyo). 1972;25:582-93.

[74] Vior NM, Lacret R, Chandra G, Dorai-Raj S, Trick M, Truman AW. Discovery and Biosynthesis of the Antibiotic Bicyclomycin in Distantly Related Bacterial Classes. Appl Environ Microbiol. 2018;84.

[75] Raghunathan N, Kapshikar RM, Leela JK, Mallikarjun J, Bouloc P, Gowrishankar J. Genome-wide relationship between R-loop formation and antisense transcription in Escherichia coli. Nucleic Acids Res. 2018.

[76] Leela JK, Syeda AH, Anupama K, Gowrishankar J. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. Proc Natl Acad Sci U S A. 2013;110:258-63.

[77] Canals A, Uson I, Coll M. The structure of RNA-free Rho termination factor indicates a dynamic mechanism of transcript capture. J Mol Biol. 2010;400:16-23.

[78] Bohlin J, Pettersson JH. Evolution of Genomic Base Composition: From Single Cell Microbes to Multicellular Animals. Comput Struct Biotechnol J. 2019;17:362-70.

[79] Levkovich SA, Gazit E, Laor Bar-Yosef D. Two Decades of Studying Functional Amyloids in Microorganisms. Trends Microbiol. 2020.

[80] Strauss M, Vitiello C, Schweimer K, Gottesman M, Rosch P, Knauer SH. Transcription is regulated by NusA:NusG interaction. Nucleic Acids Res. 2016;44:5971-82.

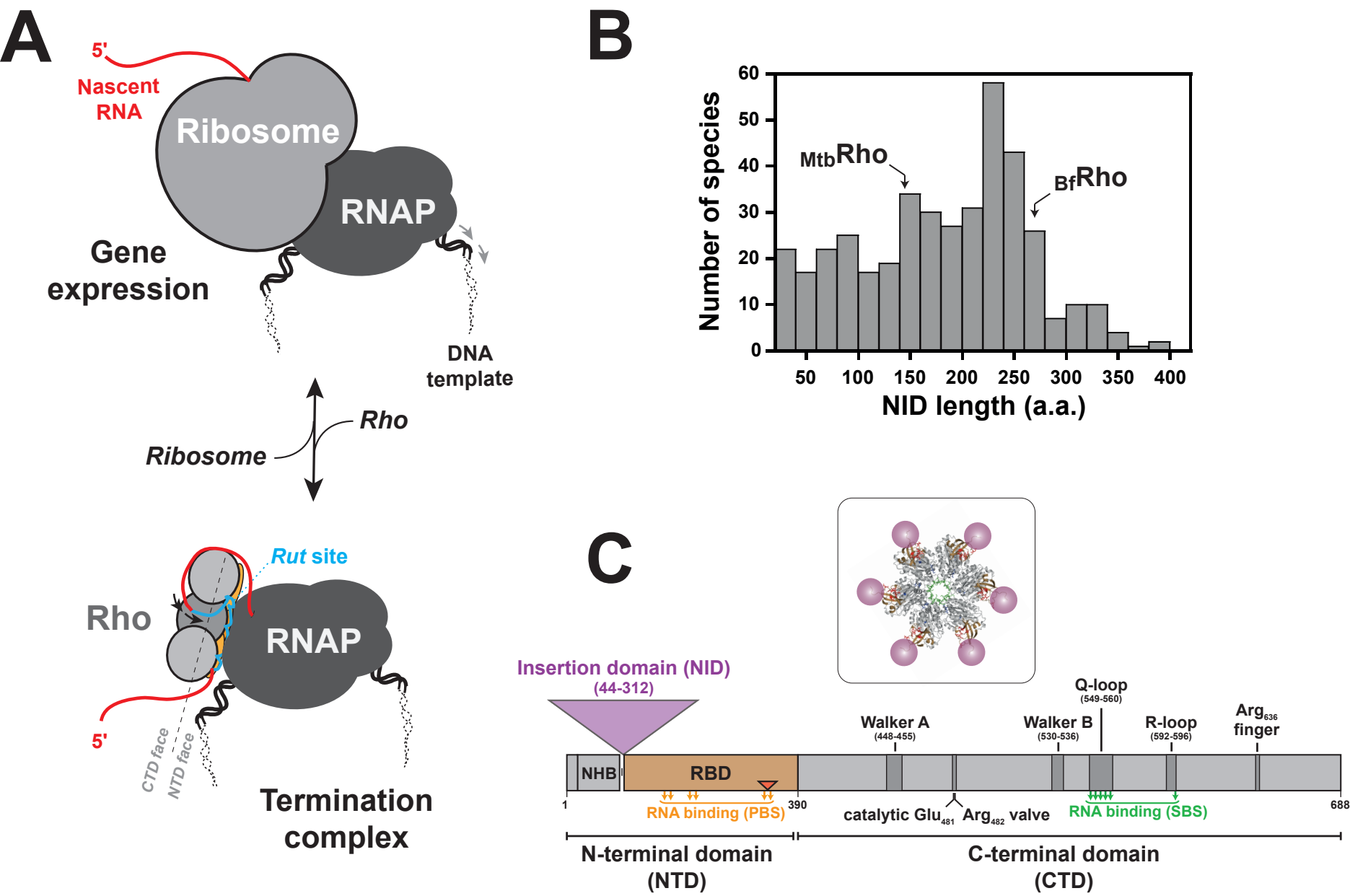
[81] D'Heygere F, Schwartz A, Coste F, Castaing B, Boudvillain M. Monitoring RNA Unwinding by the Transcription Termination Factor Rho from Mycobacterium tuberculosis. Methods Mol Biol. 2015;1259:293-311.

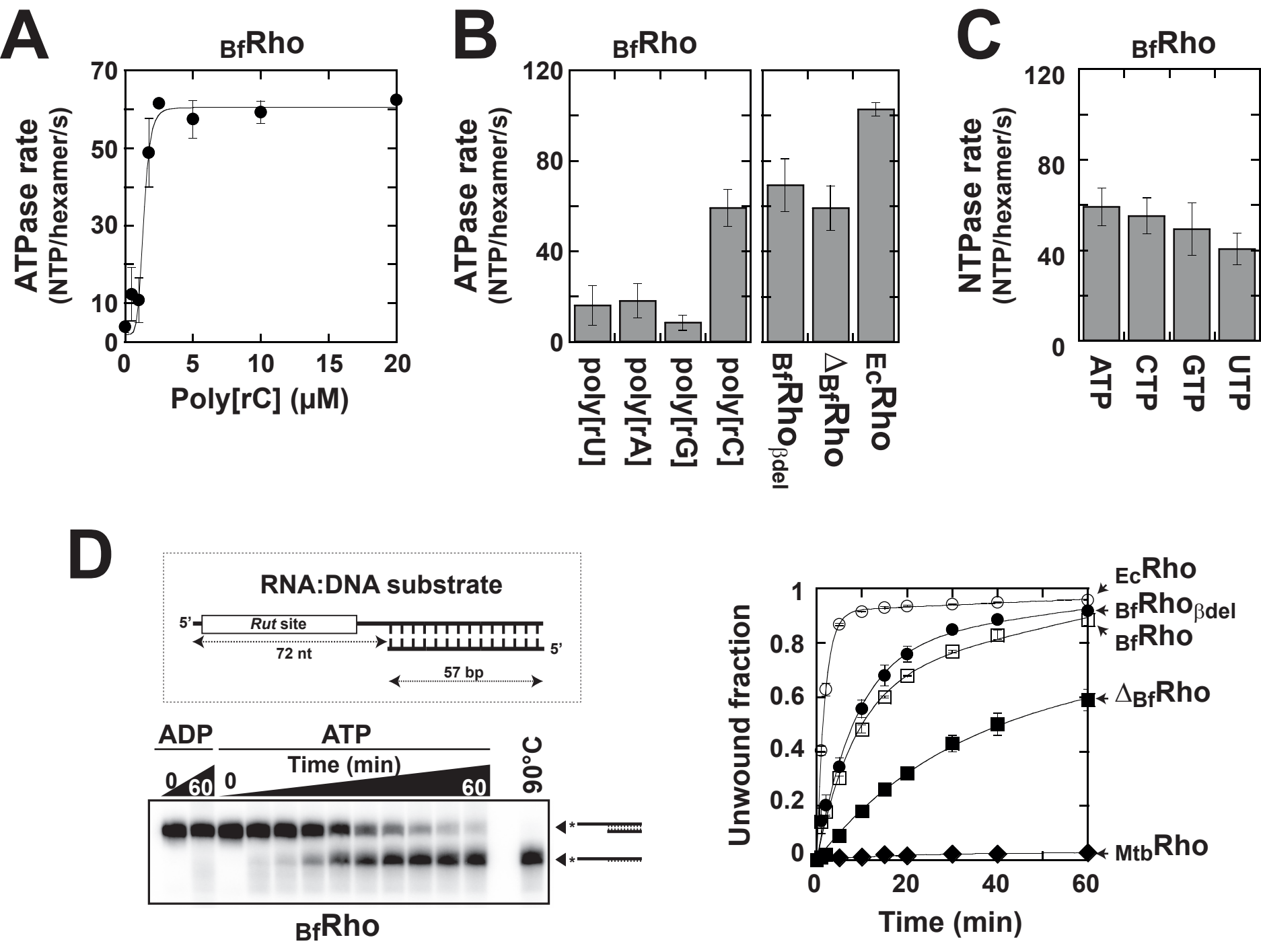
[82] Gan E, Richardson JP. ATP and other nucleotides stabilize the Rho-mRNA complex. Biochemistry. 1999;38:16882-8.

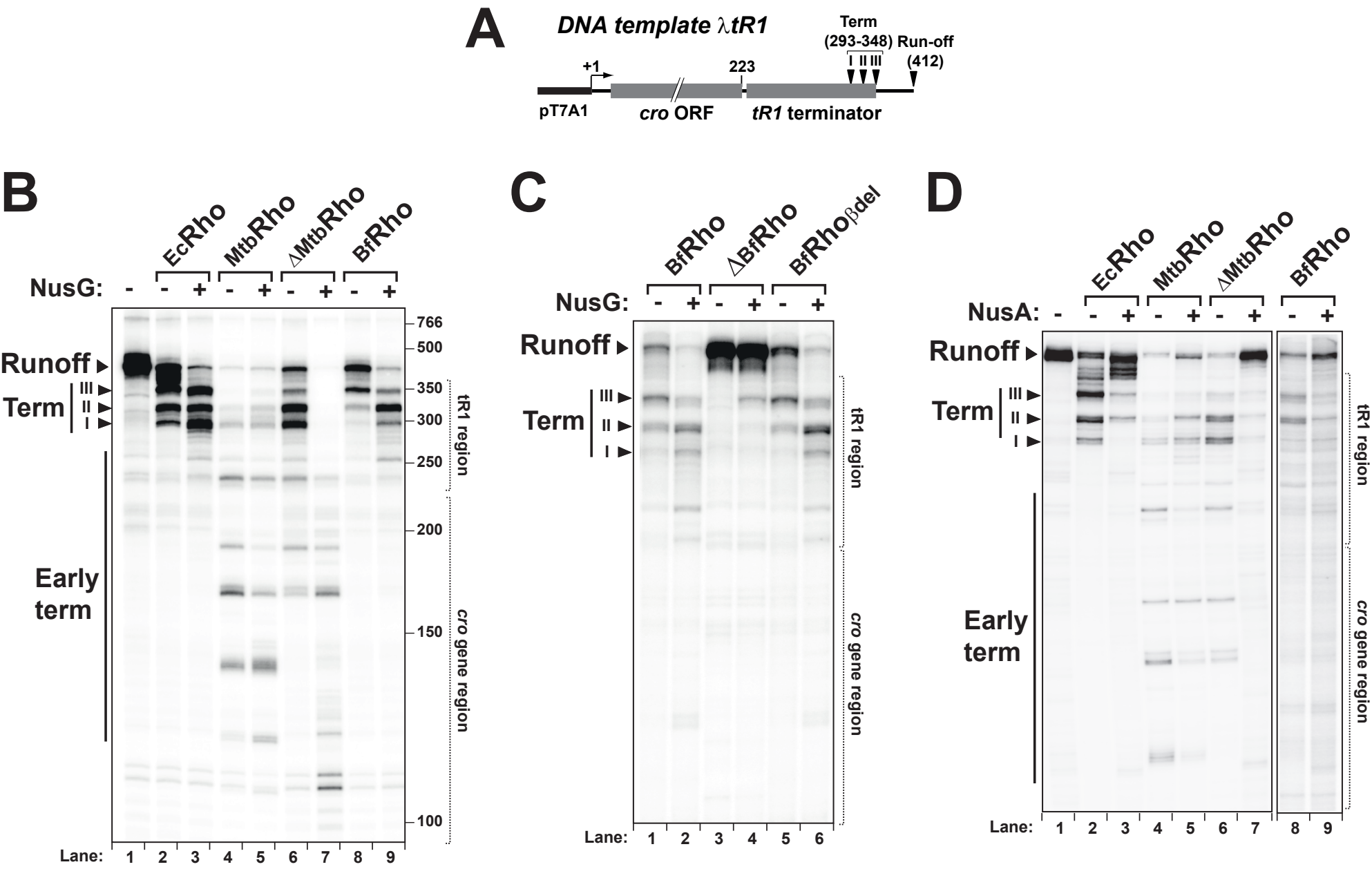
[83] Schwartz A, Walmacq C, Rahmouni AR, Boudvillain M. Noncanonical interactions in the management of RNA structural blocks by the transcription termination rho helicase. Biochemistry. 2007;46:9366-79.

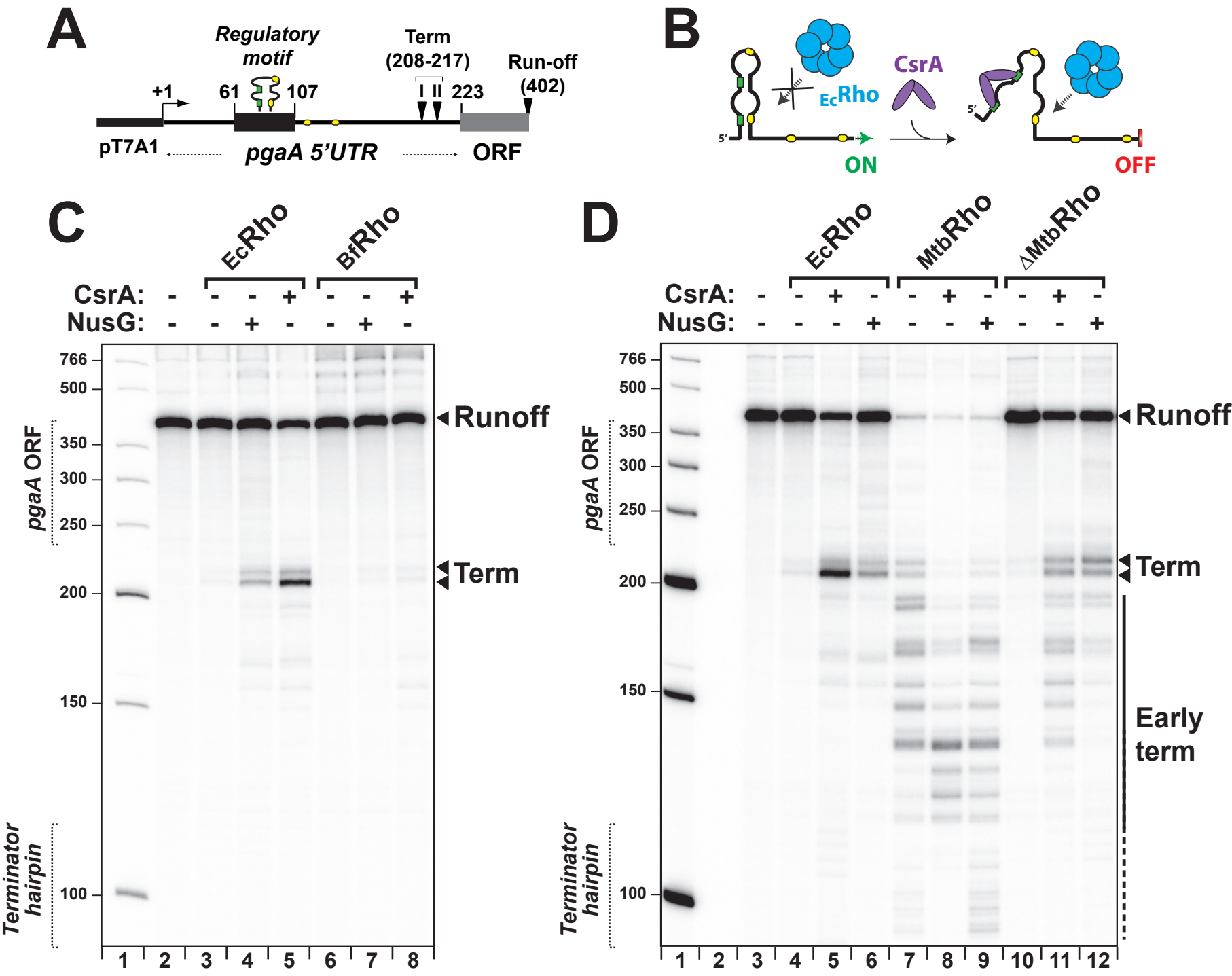
[84] Skordalakes E, Berger JM. Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. Cell. 2003;114:135-46.

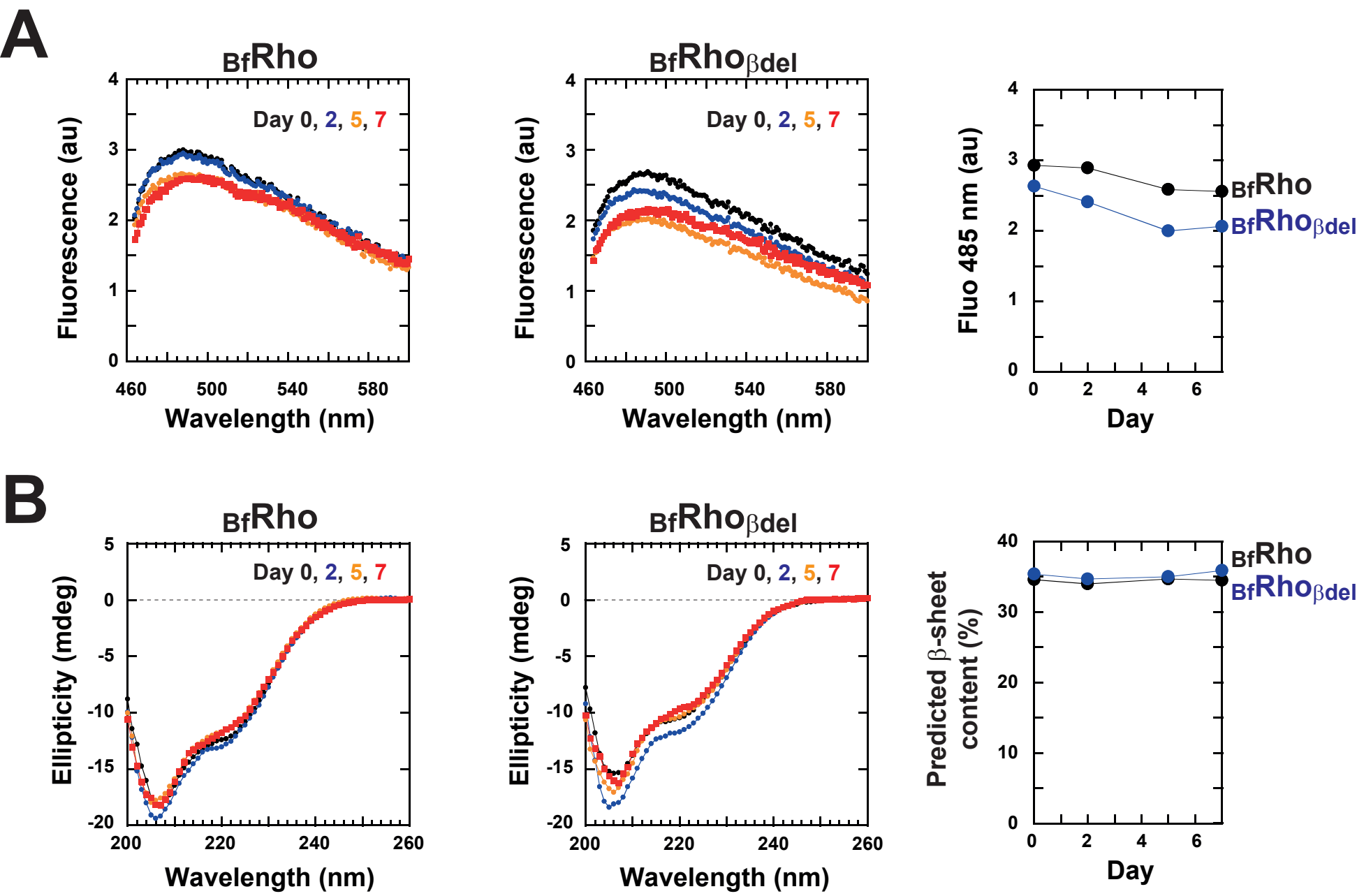
[85] Micsonai A, Wien F, Kernya L, Lee YH, Goto Y, Refregiers M, et al. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc Natl Acad Sci U S A. 2015;112:E3095-103.

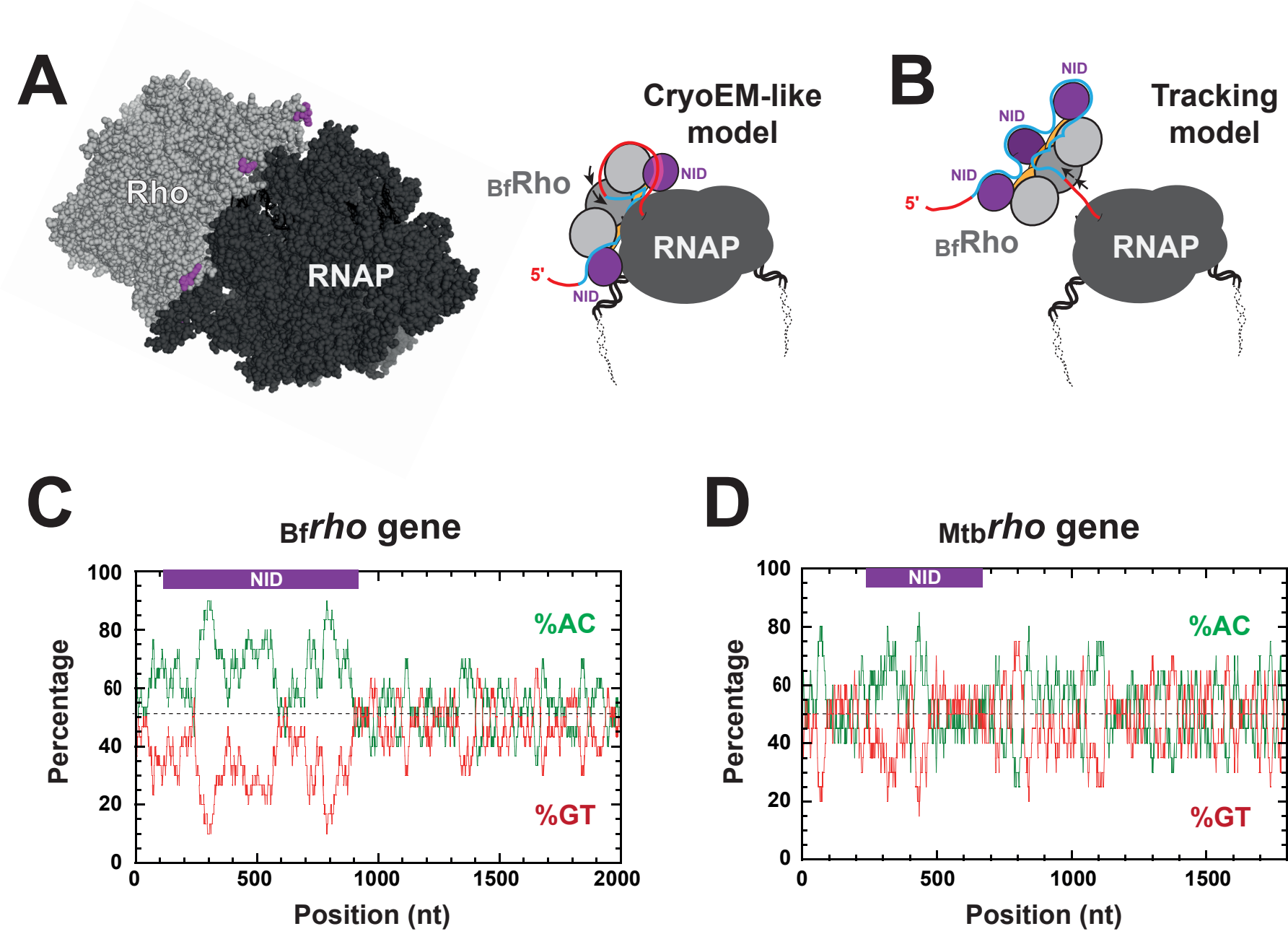

A
5'
Nascent RNA
Ribosome
RNAP
Gene expression
DNA template
Ribosome
Rho
Rho
Rut site
RNAP
5'
CTD face
NTD face
Termination complex
B
Number of species
60
50
40
30
20
10
0
MtbRho
BfRho
50
100
150
200
250
300
350
400
NID length (a.a.)
C
Insertion domain (NID)
(44-312)
Q-loop
(549-560)
Walker A
(448-455)
Walker B
(530-536)
R-loop
(592-596)
Arg636
finger
NHB
RBD
1
390
688
RNA binding (PBS)
catalytic Glu481
Arg482 valve
RNA binding (SBS)
N-terminal domain
(NTD)
C-terminal domain
(CTD)


A
BfRho
ATPase rate
(NTP/hexamer/s)
70
50
30
10
0
0
5
10
15
20
Poly[rC] (μM)
B
BfRho
ATPase rate
(NTP/hexamer/s)
120
80
40
0
poly[rU]
poly[rA]
poly[rG]
poly[rC]
BfRhoβdel
ΔBfRho
EcRho
C
BfRho
NTPase rate
(NTP/hexamer/s)
120
80
40
0
ATP
CTP
GTP
UTP
D
RNA:DNA substrate
5'
Rut site
72 nt
57 bp
5'
ADP
ATP
Time (min)
0
60
0
60
90°C
BfRho
Unwound fraction
1
0.8
0.6
0.4
0.2
0
0
20
40
60
Time (min)
EcRho
BfRhoβdel
BfRho
ΔBfRho
MtbRho


A
DNA template λtR1
Term
(293-348)
I II III
Run-off
(412)
+1
223
pT7A1
cro ORF
tR1 terminator
B
EcRho
MtbRho
ΔMtbRho
BfRho
NusG: - - + - + - + - +
Runoff
Term
III
II
I
Early term
766
500
350
300
250
200
150
100
tR1 region
cro gene region
Lane: 1 2 3 4 5 6 7 8 9
C
BfRho
ΔBfRho
BfRhoβdel
NusG: - + - + - +
Runoff
Term
III
II
I
tR1 region
cro gene region
Lane: 1 2 3 4 5 6
D
EcRho
MtbRho
ΔMtbRho
BfRho
NusA: - - + - + - + - +
Runoff
Term
III
II
I
Early term
tR1 region
cro gene region
Lane: 1 2 3 4 5 6 7 8 9


A
Regulatory motif
Term (208-217)
Run-off (402)
+1
61
107
I II
223
pT7A1
pgaA 5'UTR
ORF
B
EcRho
CsrA
5'
ON
OFF
C
EcRho
BfRho
CsrA: - - - + - - +
NusG: - - + - - + -
766
500
350
300
250
200
150
100
pgaA ORF
Terminator hairpin
Runoff
Term
1 2 3 4 5 6 7 8
D
EcRho
MtbRho
ΔMtbRho
CsrA: - - + - - + - - + -
NusG: - - - + - - + - - +
766
500
350
300
250
200
150
100
pgaA ORF
Terminator hairpin
Runoff
Term
Early term
1 2 3 4 5 6 7 8 9 10 11 12


A
BfRho
Day 0, 2, 5, 7
Fluorescence (au)
Wavelength (nm)
BfRhoβdel
Day 0, 2, 5, 7
Fluorescence (au)
Wavelength (nm)
Fluo 485 nm (au)
Day
BfRho
BfRhoβdel
B
BfRho
Day 0, 2, 5, 7
Ellipticity (mdeg)
Wavelength (nm)
BfRhoβdel
Day 0, 2, 5, 7
Ellipticity (mdeg)
Wavelength (nm)
Predicted β-sheet content (%)
Day
BfRho
BfRhoβdel


A
Rho
RNAP
CryoEM-like model
BfRho
NID
RNAP
5'
NID
B
Tracking model
NID
NID
NID
5'
BfRho
RNAP
C
Bfrho gene
NID
%AC
%GT
Percentage
Position (nt)
0
20
40
60
80
100
0
500
1000
1500
2000
D
Mtbrho gene
NID
%AC
%GT
Percentage
Position (nt)
0
20
40
60
80
100
0
500
1000
1500

A large insertion domain in the Rho factor from a low-G+C Gram-negative bacterium is critical for RNA binding and transcription termination activity

Isabelle Simon, Mildred Delaleau, Annie Schwartz, and Marc Boudvillain

(Supplementary information)

SUPPLEMENTARY METHODS

Single round transcriptions. Single-round 'on-beads' transcription termination experiments were performed as described previously (1) with minor modifications. The biotinylated λ-tR1 DNA templates (30 nM) were mixed with *E. coli* RNAP holoenzyme (36 nM) in transcription initiation buffer (40 mM Tris-HCl, pH 7.9, 50 mM KCl, 5 mM $MgCl_2$, 1mM DTT, 0.05 mg/mL BSA, 0.4U/μL Superase-In [Ambion]) and incubated at 37°C for 5 min. The mixture was supplemented with 12 μM ApU, 6 μM ATP, 6μM GTP, 0.95μM UTP, and 1 μCi/μL [α^{32}P]-UTP, incubated for 10 min at 37°C, and then chilled on ice. The concentration of KCl was raised to 0.1 M before immobilization of the halted transcription complexes on streptavidin-coated magnetic beads (Ademtech, Pessac, France). Then, rNTPs (75 μM each, final concentrations) and rifampicin (25 μg/mL) were added to initiate chase reactions. After incubation for 10 min at 37°C, supernatants and bead pellets were separated on a MagRack (Cytiva) and their transcript compositions analyzed by 8% denaturing PAGE and Typhoon-Trio phosphorimaging.

Supplementary Table 1: plasmids			
Plasmid	**Method**	**Source DNA**	**Primers**
pET28b-BfRho	subcloning of PCR fragment	Genomic DNA from *B. fragilis* type strain NCTC 9343	5'GGCAGCCATATGATGTATAACATCATTCAATTGAACG (BfRho-F) 5'GGTGGTGGTGGTGGTGCTCGAGTTAGCTGTTCATGCTCATCAGG (BfRho-R)
pET28b-ΔBfRho	HiFi DNA Assembly	pET28b-BfRho	5'GGCGGCAGCGGCTATGAATTTGATGATATCCTCACCGG (ΔBfRho-F) 5'GCCGCTGCCGCCTTGTTCGTCGAGGATTTTG (ΔBfRho-R)
pET28b-BfRhoβdel	HiFi DNA Assembly	pET28b-BfRho	5'GCCCACGAGATGCCGCCGAAGCCGCACC (βdelBfRho-F) 5'TTCGGCGGCATCTCGTGGGCGAATCACACGCG (βdelBfRho-R)

Supplementary table 2 : helicase reaction parameters[a]				
		Burst phase		**Linear phase**
Protein	**Condition[b]**	**Amplitude (A)**	k_{exp} **(min^{-1})**	k_{lin} **(X100 min^{-1})**
$_{Ec}$Rho	Standard	0.92 ± 0.03	0.58 ± 0.01	0.08 ± 0.01
$_{Bf}$Rho	Standard	0.67 ± 0.07	0.12 ± 0.02	0.4 ± 0.2
	10X	1.00 ± 0.03	0.07 ± 0.02	-
	Pre-incubation	0.76 ± 0.03	0.13 ± 0.01	0.2 ± 0.1
Δ_{Bf}Rho	Standard	0.54 ± 0.28	0.04 ± 0.02	0.2 ± 0.3
	10X	0.93 ± 0.02	0.14 ± 0.01	-
$_{Bf}$Rho$_{\beta del}$	Standard	0,90 ± 0.02	0.10 ± 0.01	0.02 ± 0.04
	Pre-incubation	0,86 ± 0.06	0.09 ± 0.01	0.2 ± 0.1
$_{Mtb}$Rho	Standard	0.02 ± 0.05	0.06 ± 0.08	0.01 ± 0.05

[a] Reaction data points were fitted to an equation describing the kinetic regimen determined previously for $_{Ec}$Rho: $F_p = A \times \left(1 - e^{-k_{exp}t}\right) + k_{lin} \times t$, where F_p is the fraction of product formed, A is the amplitude of the exponential (burst) phase of the reaction, and k_{exp} and k_{lin} are the rate constants of the exponential and linear phases of the reaction, respectively (2-4).

[b] In "standard" conditions, the RNA-DNA substrate was preincubated with 20 nM Rho for 5 min at 30°C before initiating the helicase reaction with ATP. In "10X" conditions, Rho concentration was raised to 200 nM while in "pre-incubation" conditions, Rho proteins were first incubated for 2h at 37°C (to probe their stability) before being used in the helicase reactions. Note that linear steady-state phases were no longer detectable under "10X" conditions.

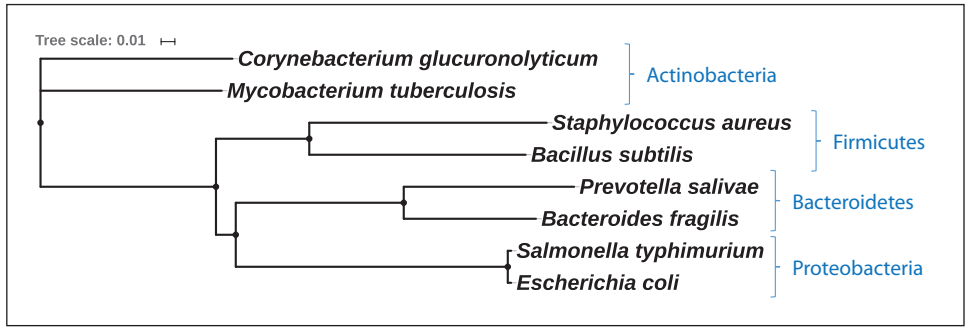
Supplementary Table 3 : conservation of Rho:TEC contacts[a]							
Rho			**Interaction partner**				**Ref.**
Ec	**Bf**	**Mtb**	**identity**	**Ec**	**Bf**	**Mtb**	
S82	S353	S254	RNAP β	K1022	Y1006	del[b]	(5)
R87	K358	R259	RNAP β	D491	D452	T410	(5)
R88	L359	K260	RNAP β	D485	E446	E404	(5)
			RNAP β	Y1018	F1002	del[b]	(5)
			NusA	E136	T143	E113	(6)
R102	R373	R274	RNAP β	D1019	I1003	del[b]	(5)
K105	K376	K277	RNAP β	H1023	K1007	del[b]	(5)
E106	E377	E278	RNAP α	K298	K302	K222	(6)
			RNAP β'	K39	L35	K29	(6)
			RNAP β'	R60	R56	R50	(6)
E108	E379	Q286	RNAP α	K297	K301	Q221	(6)
			RNAP β	K1027	A1011	del[b]	(5)
K115	K386	R293	RNAP β	E1016	L1000	del[b]	(5)
			NusA	E219	G226	G203	(6)
T276	T546	N451	RNAP β	P897	P864	P810	(6)

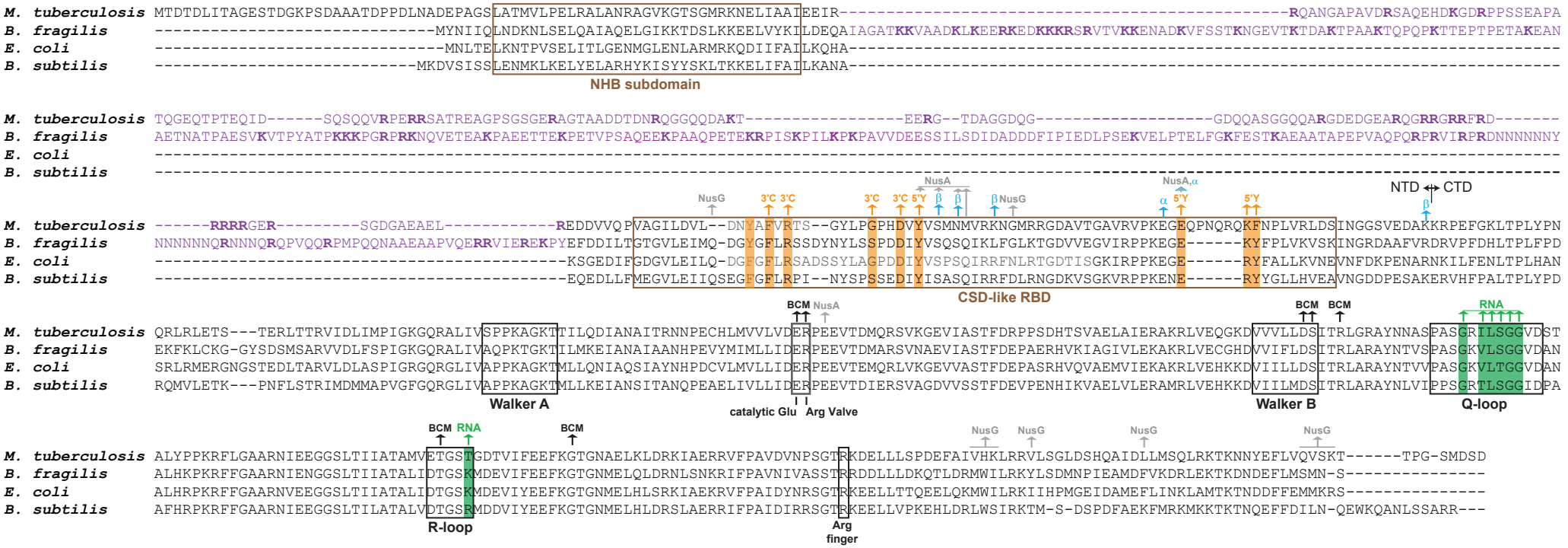
[a] Conservation of residues in proteins from *E. coli* (Ec), *B. fragilis* (Bf), and *M. tuberculosis* (Mtb). Residues not conserved are in red. The loosely defined interactions (e.g. between regions rather than side-chains) reported in refs (5,6) are not listed here, although some involve lineage-specific RNAP domains (5,6)

[b] Deletion. See also Figure S8.

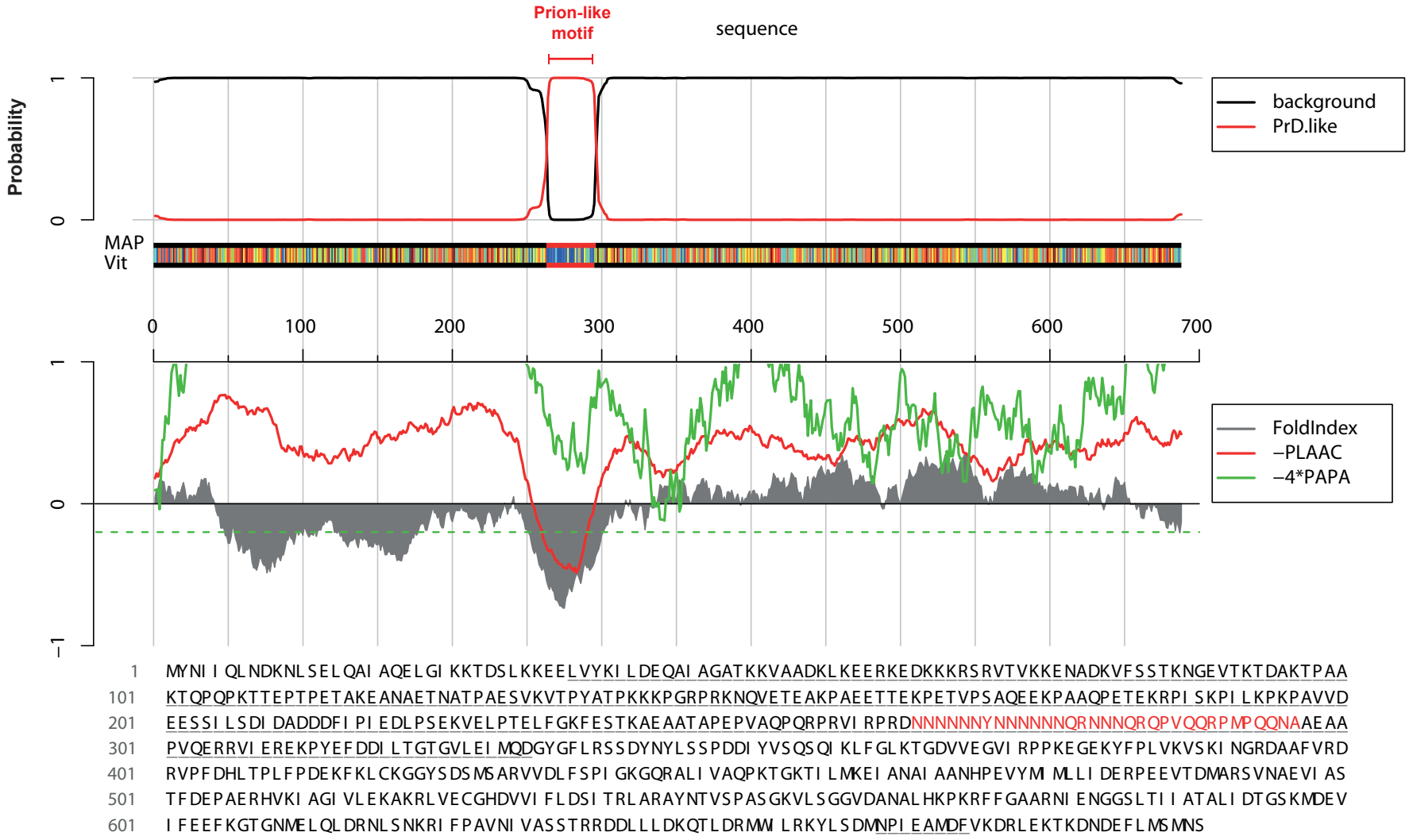
References

1. Figueroa-Bossi, N., Schwartz, A., Guillemardet, B., D'Heygere, F., Bossi, L. and Boudvillain, M. (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev*, **28**, 1239-1251.
2. Walstrom, K.M., Dozono, J.M., Robic, S. and von Hippel, P.H. (1997) Kinetics of the RNA-DNA helicase activity of Escherichia coli transcription termination factor rho. 1. Characterization and analysis of the reaction. *Biochemistry*, **36**, 7980-7992.
3. Walmacq, C., Rahmouni, A.R. and Boudvillain, M. (2004) Influence of substrate composition on the helicase activity of transcription termination factor Rho: reduced processivity of Rho hexamers during unwinding of RNA-DNA hybrid regions. *J Mol Biol*, **342**, 403-420.
4. Walmacq, C., Rahmouni, A.R. and Boudvillain, M. (2006) Testing the steric exclusion model for hexameric helicases: substrate features that alter RNA-DNA unwinding by the transcription termination factor Rho. *Biochemistry*, **45**, 5885-5895.
5. Hao, Z., Epshtein, V., Kim, K.H., Proshkin, S., Svetlov, V., Kamarthapu, V., Bharati, B., Mironov, A., Walz, T. and Nudler, E. (2021) Pre-termination Transcription Complex: Structure and Function. *Mol Cell*, **81**, 281-292 e288.
6. Said, N., Hilal, T., Sunday, N.D., Khatri, A., Burger, J., Mielke, T., Belogurov, G.A., Loll, B., Sen, R., Artsimovitch, I. *et al.* (2021) Steps toward translocation-independent RNA polymerase inactivation by terminator ATPase rho. *Science*, **371**.

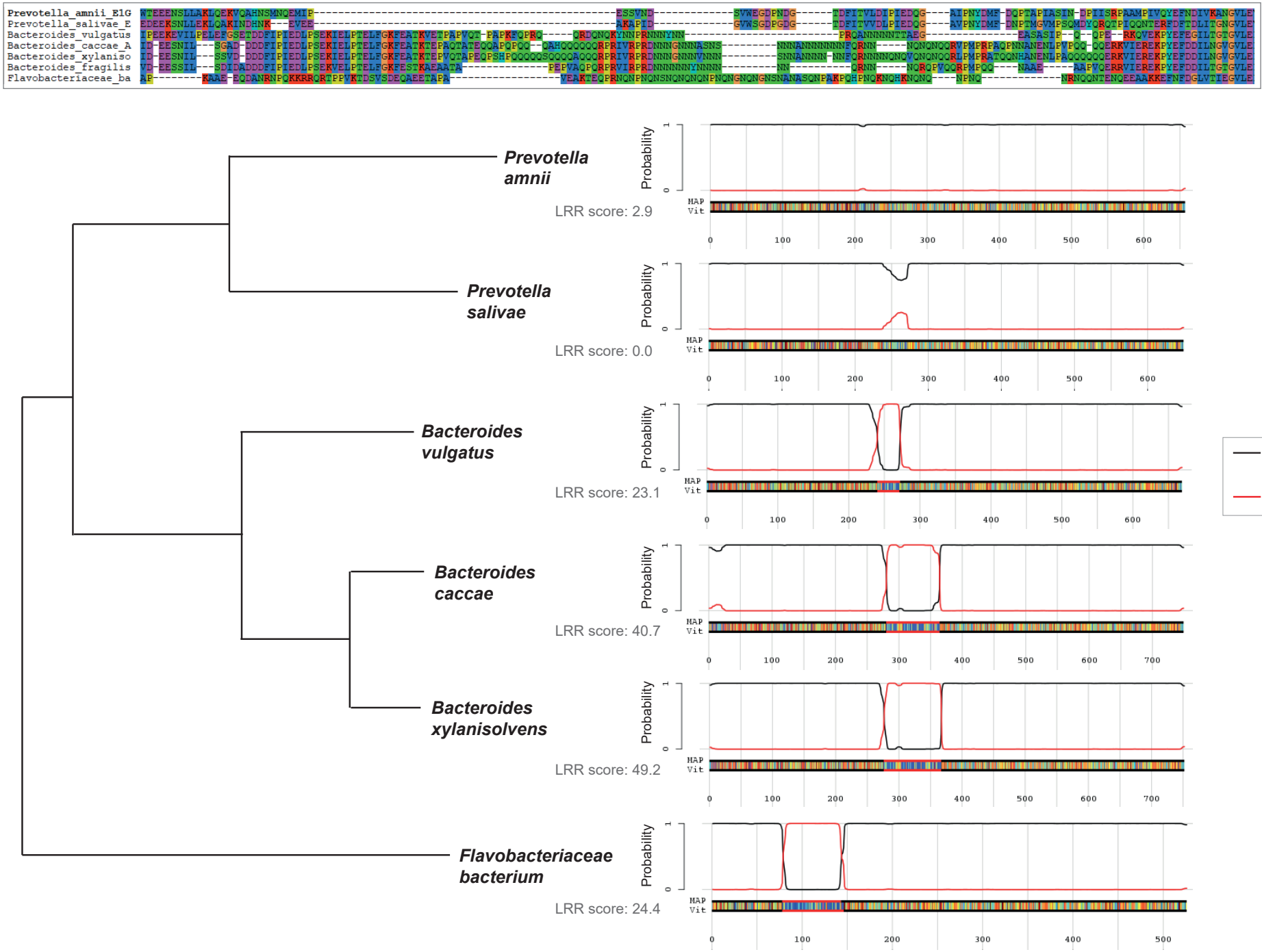




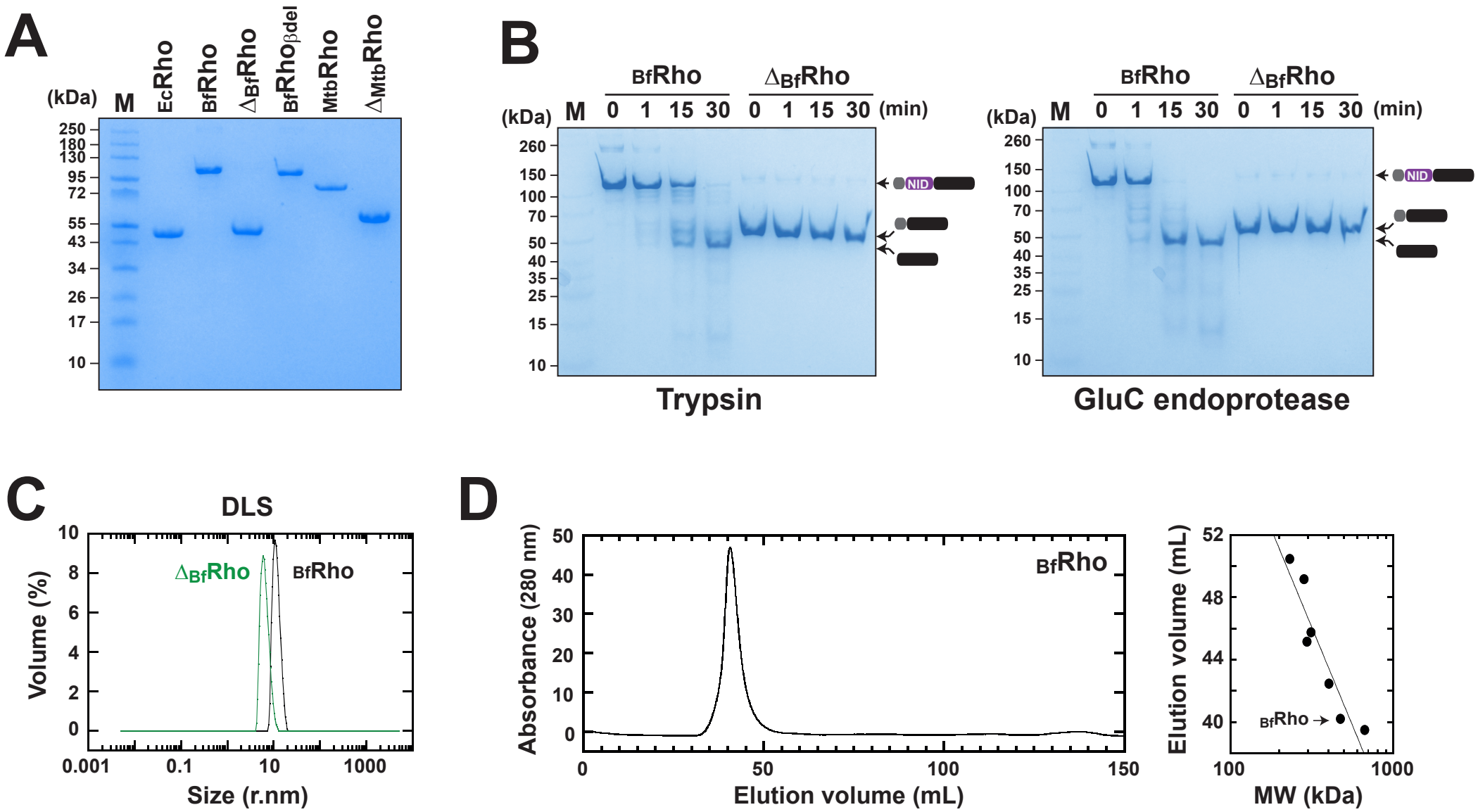
Supplementary Figure 1: Conservation of the Rho sequence across bacterial phyla. The phylogenetic tree and sequence alignment were generated with the Seaview software as described in D'Heygere et al., *Microbiolgy* (2013) 159, 1423-36. The NID sequences of $_{Bf}$Rho and $_{Mtb}$Rho are in purple with basic residues in bold. Features important for the function of $_{Ec}$Rho are highlighted on the sequence alignment. PBS components are highlighted in orange while SBS residues are in green. The $_{Ec}$Rho residues making the most probable contacts with $_{Ec}$RNAP and $_{Ec}$NusA/G factors in recent cryoEM and crystal structures (see main text) are indicated by blue/gray arrows.



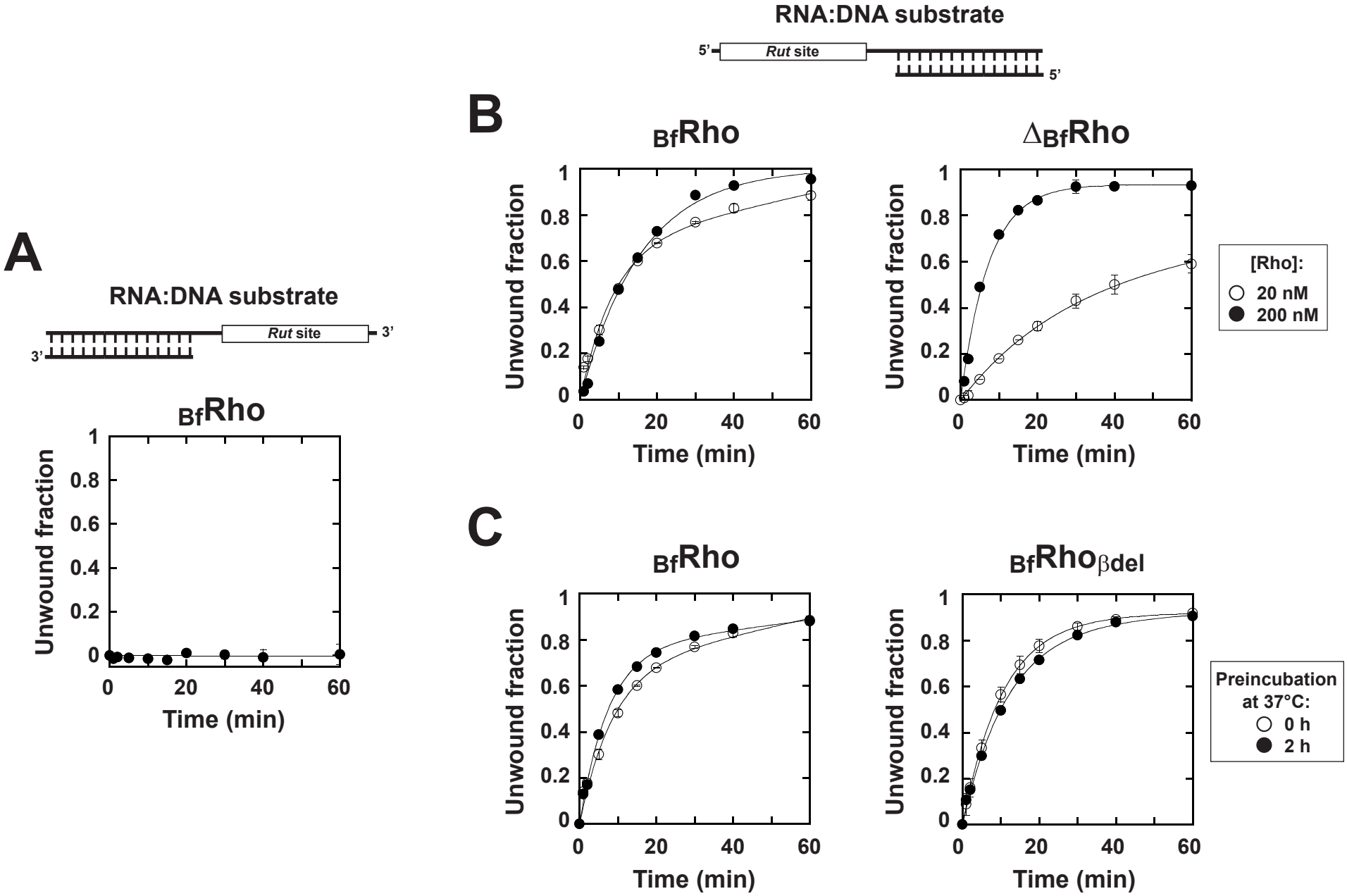
Supplementary Figure 2: Analysis of the prion-like propensity contained in the BfRho sequence using the PLAAC server (Lancaster et al., *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu310). A screenshot of the server output is shown with the pedicted prion-like sequence in the NID in red (http://plaac.wi.mit.edu; default parameters). The prion-likelihood LLR score is 22.6.



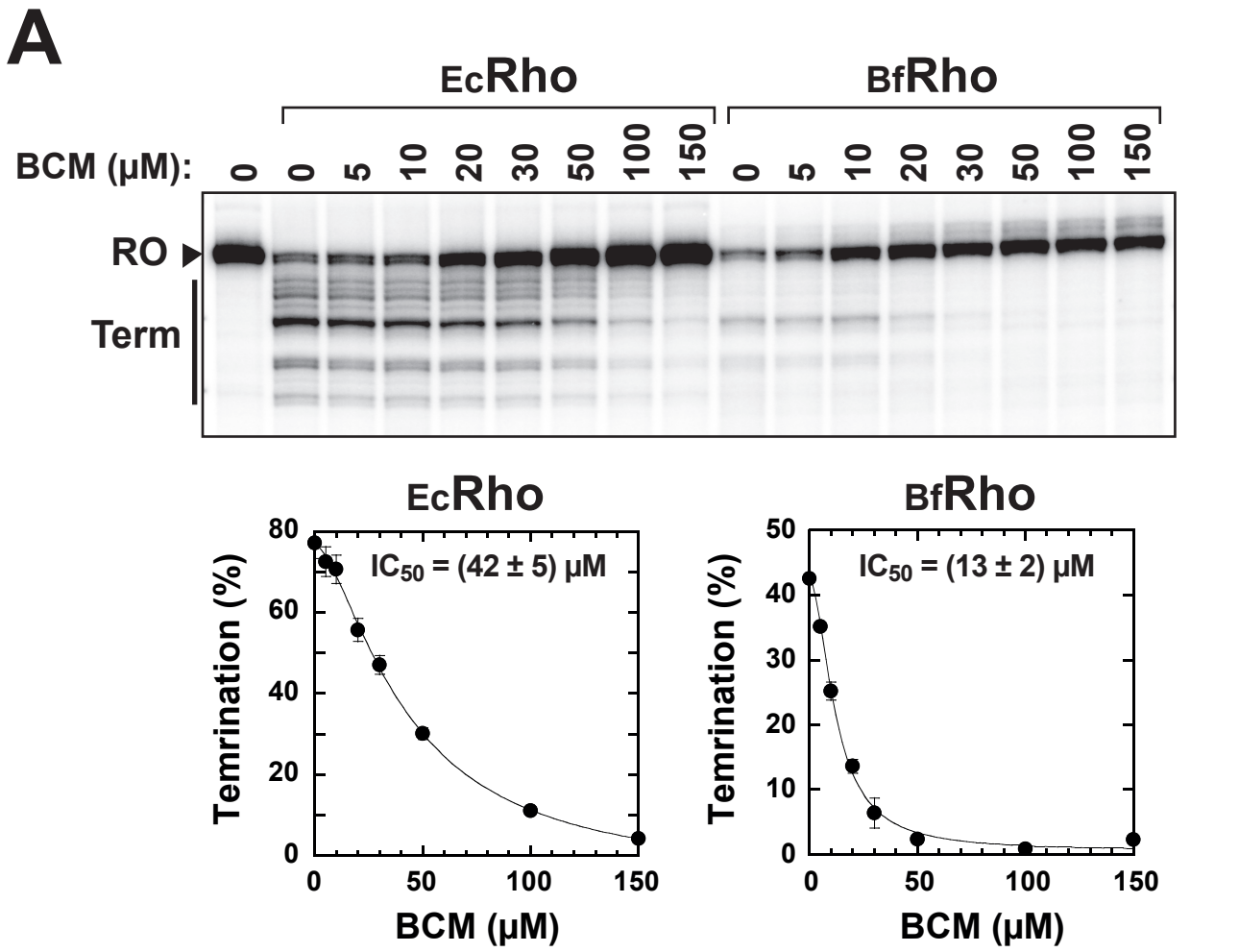
Supplementary Figure 3: Prion-like propensity of representative Bacteroidetes Rho sequences as predicted with the PLAAC server. A screenshot of the Sea-view sequence alignment zoomed on the NID region containing N/Q-rich motifs is shown at the top of the figure.

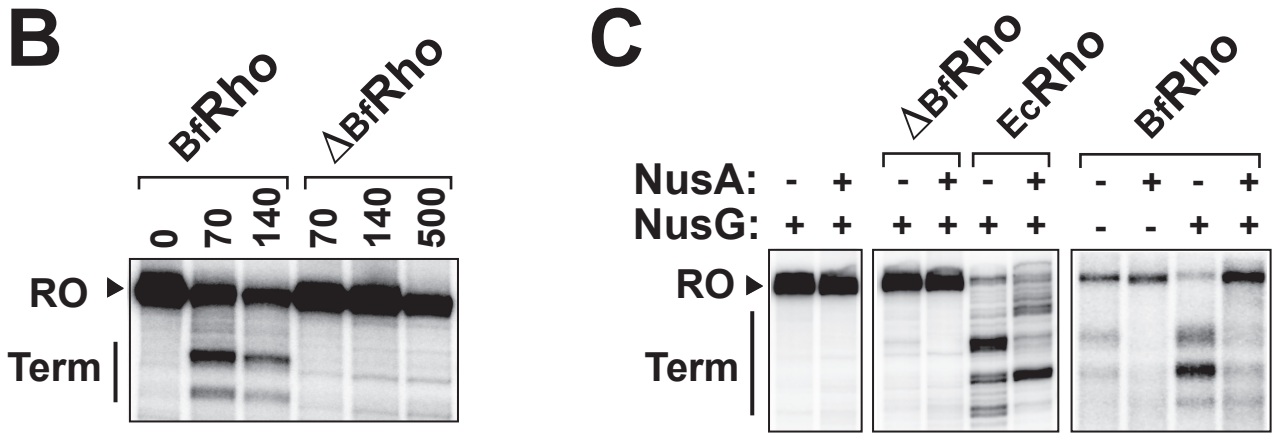


Supplementary Figure 4: (A) SDS-PAGE gel of the purified proteins used in the study. The $_{Ec}$Rho (47.1 kDa) and $_{\Delta Bf}$Rho (49.3 kDa) proteins migrate in the expected size range, the $_{Mtb}$Rho (67.3 kDa), and $_{\Delta Mtb}$Rho (52.2 kDa) display slight gel shifts as expected (Kalarickal et al., J. Mol. Biol., 2010, 395, 966-82), whereas $_{Bf}$Rho (79 kDa) and $_{Bf}$Rho$_{\beta del}$ (75.1 kDa) migrate much more slowly than expected (>20kDa difference in relative mobilities). Lane M corresponds to the broad range 10-250 kDa protein standard (NEB #P7712). **(B)** Protease digests of the $_{Bf}$Rho and $_{\Delta Bf}$Rho proteins support that the NID is intrinsically disordered. Reactions were performed at 37°C with 1:50 (w:w) protease in 100 mM KCl, 10 mM Tris-HCl, pH 8, 50% glycerol, 0.1 mM EDTA, 0.1 mM DTT. **(C)** Diffusion light scattering analysis indicates that $_{Bf}$Rho and $_{\Delta Bf}$Rho form monodisperse species (Zetasizer Nano S instrument). **(D)** The $_{Bf}$Rho factor elutes from a Sephacryl S300 HR column (10-1500 kDa separation range) in a single peak consistent with hexamer formation. MW controls were catalase (232 kDa), $_{Ec}$Rho hexamer (284 kDa), $_{\Delta Bf}$Rho hexamer (296 kDa), $_{\Delta Mtb}$Rho hexamer (314 kDa), $_{Mtb}$Rho hexamer (404 kDa), and thyroglobulin (669 KDa).

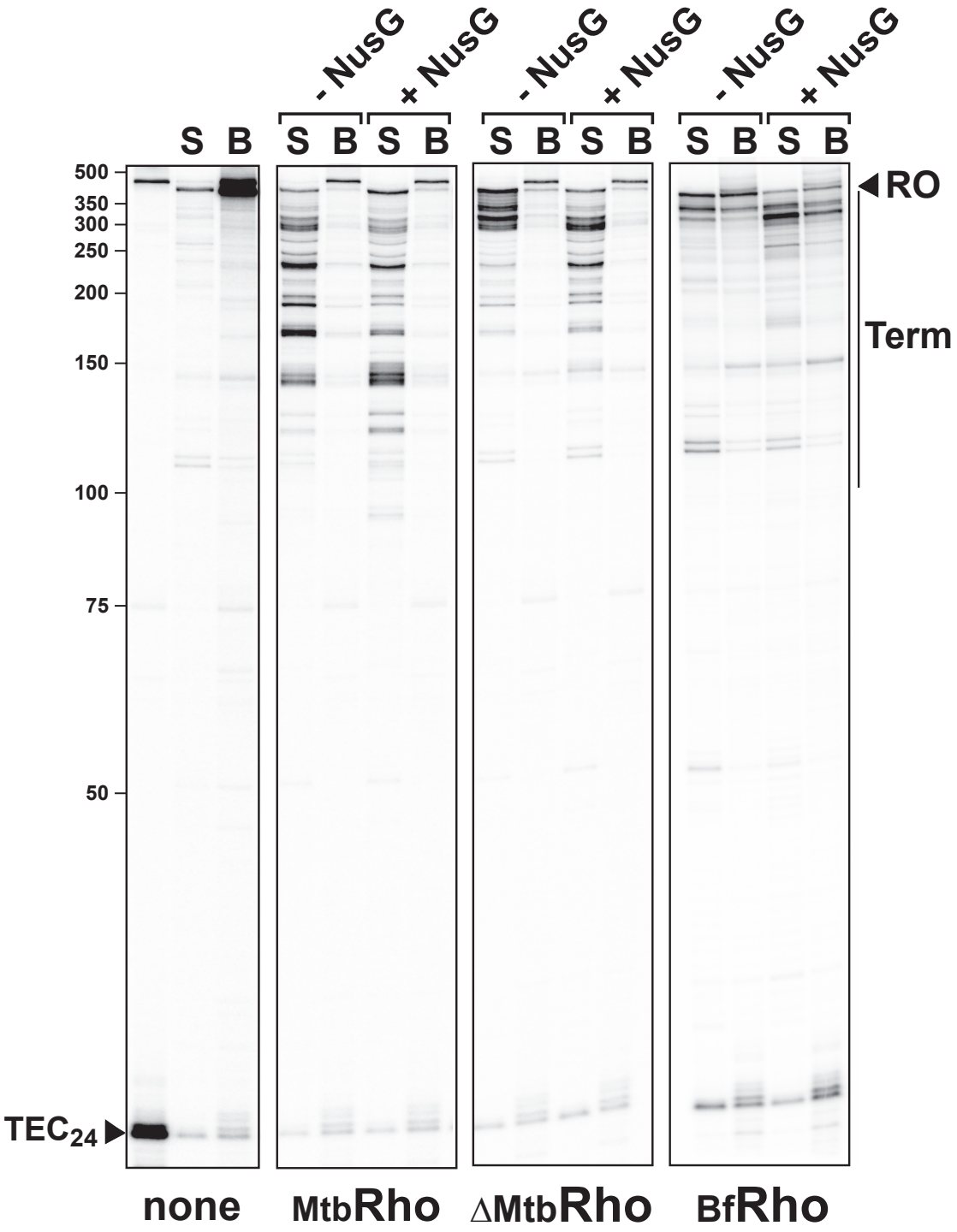


Supplementary Figure 5: RNA-DNA duplex unwinding by the $_{Bf}$Rho factor. (**A)** $_{Bf}$Rho cannot unwind a RNA-DNA duplex located upstream from the single-stranded *Rut* region; (**B)** Increasing the concentration of Δ_{Bf}Rho stimulates duplex unwinding; **(C)** Incubating $_{Bf}$Rho for long periods of time without ATP or RNA does not compromise its helicase activity. Protein stock aliquots (~1 µM) were diluted to 200 nM in 1X helicase buffer before preincubation for 0 or 2h at 37°C.

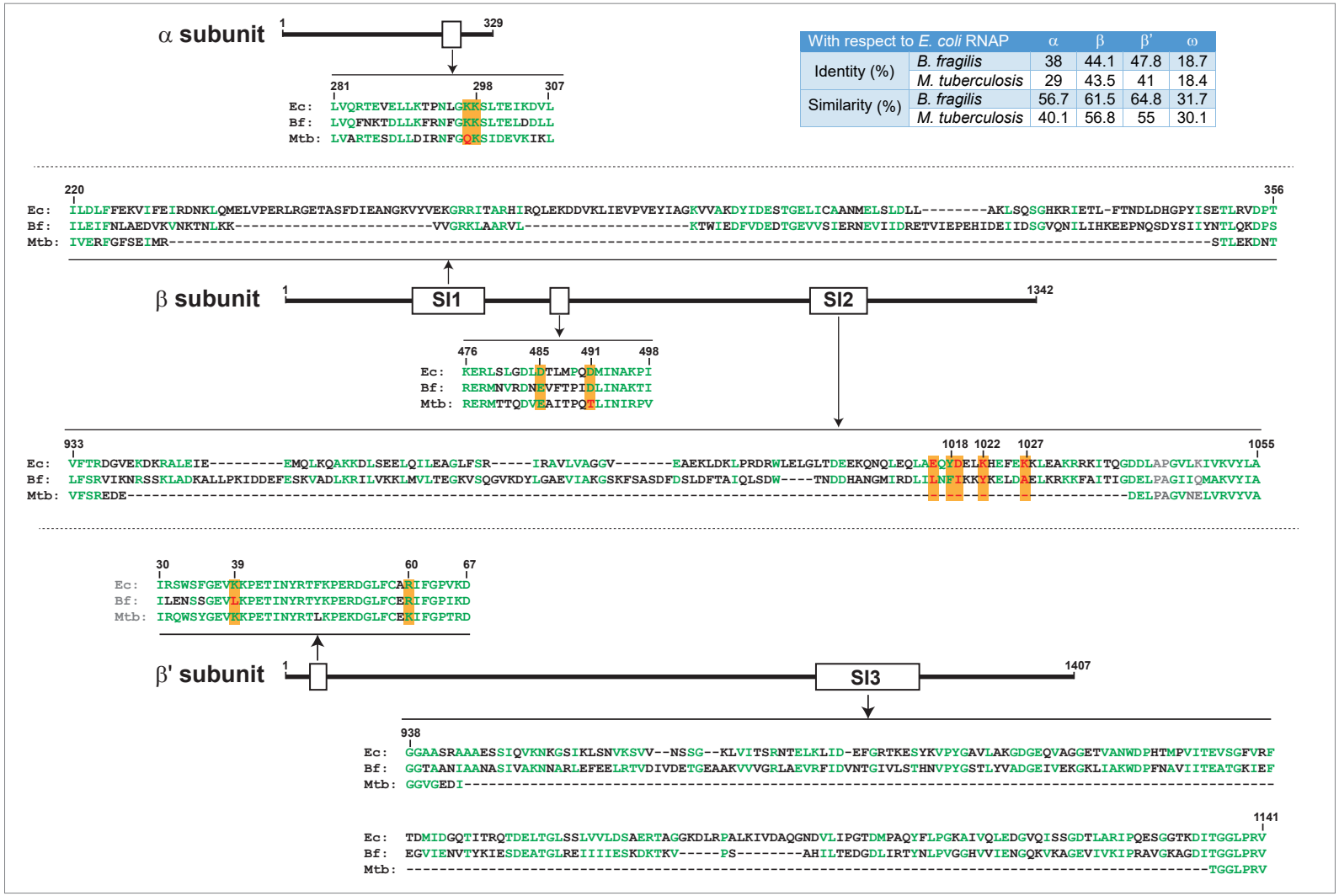




Supplementary Figure 6: Standard transcription termination experiments with the λtR1 template. (**A)** $_{Bf}$Rho is highly sensitive to the bicyclomycin (BCM) inhibitor. (**B)** Increasing the concentration of Δ_{Bf}Rho does not alleviate its RDTT defect. **(C)** RDTT with $_{Bf}$Rho is strongly delayed by $_{Ec}$NusA even in the presence of $_{Ec}$NusG.

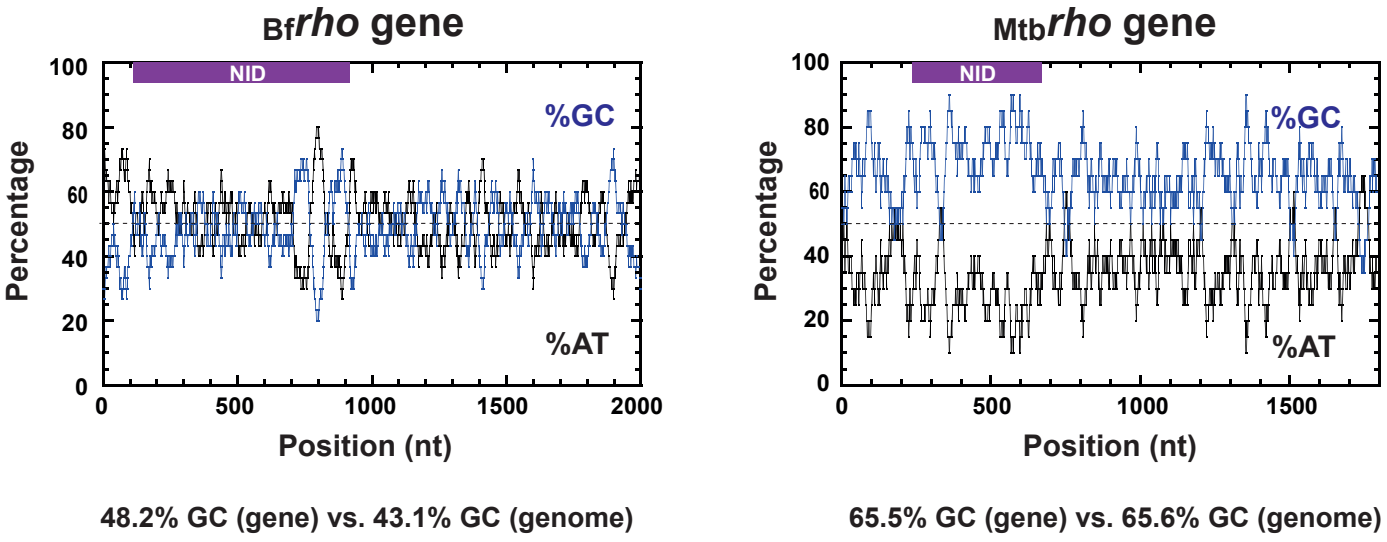


Supplementary Figure 7: Single-round transcription experiments performed with bead-affixed TECs containing biotinylated λtR1 templates. Reactions were performed in presence of 100 mM KCl, as in figure 3. In each case, transcripts released in the supernatant (S lanes) or still bound to the beads (B lanes) are compared.



With respect to *E. coli* RNAP		α	β	β'	ω
Identity (%)	*B. fragilis*	38	44.1	47.8	18.7
	M. tuberculosis	29	43.5	41	18.4
Similarity (%)	*B. fragilis*	56.7	61.5	64.8	31.7
	M. tuberculosis	40.1	56.8	55	30.1

Supplementary Figure 8: Conservation of the core RNAP and NusA factor among *E. coli*, *B. fragilis*, and *M. tuberculosis*. Alignments correspond to regions involved in contacts between $_{Ec}$RNAP and $_{Ec}$Rho (contact residues boxed in orange with non-conserved side-chains in red) and to the lineage-specific sequence insertions present in $_{Ec}$RNAP (SI1, SI2, SI3; see Artsimovitch et al., J. Biol. Chem, 2003, 278,12344-55). Note that similar sequence insertions are found in $_{Bf}$RNAP but not in $_{Mtb}$RNAP. Numbering is for the *E. coli* sequences.



Supplementary Figure 9: Distributions of G and C residues versus A and T residues along the sequences of the genes encoding $_{Bf}$*Rho* and $_{Mtb}$*Rho*. Percentages of individual bases were calculated with EMBOSS freak (Rice et al., Trends Genet., 2000, 16, 276-7) using a 20 nt sliding window.