Mechanism of action of bacterial transcription terminator Rho
Passong Immanuel R Chhakchhuak, Ajay Khatri and Ranjan Sen
Laboratory of Transcription, Center for DNA Fingerprinting, Inner Ring Road, Uppal, Hyderabad-39.
e.mail: Ranjan Sen: rsen@cdfd.org.in; Passong Immanuel R Chhakchhuak: pimmanuel@cdfd.org.in; Ajay Khatri: ajaykhatri@cdfd.org.in. Phone: +91 9866122652; Fax: 040-27216006.
*Corresponding Author
©Equal Contribution.
Running Title: Transcription terminator Rho
Key words: Bacterial transcription, Rho, NusG, NusA, transcription termination.

Abstract
The bacterial transcription terminator, Rho, is a well conserved protein among the prokaryotes. It was discovered about 50 years back and still, a significant number of the researchers across the globe are actively engaged in understanding the function of this protein. Rho is a hexameric RNA-dependent helicase that dislodges transcribing RNA polymerase. Its binding site on the RNA is quite degenerated, and thereby many mRNAs are its target, which leads to regulation of a wide range of operons by the Rho-dependent termination. This genome-wide control by Rho brings in pleiotropic effects in the cell physiology, and hence if this termination process is perturbed, many biological processes get directly affected. In this review, we will cover various mechanistic aspects of the Rho-dependent transcription termination highlighting the significant contributions made by us over the last 12 years. We shall also describe some aspects of the cell physiology that is under the control of this factor-dependent termination process. The review includes the biochemistry and the structural analyses of the Rho, its mechanism of action, its regulation by other cellular factors and briefly cellular events controlled by this protein.
Bacterial Transcription

Transcription is the first level of gene regulation that involves multiple steps, namely, initiation, elongation, and termination (Figure 1). In a typical transcription reaction, the bacterial multi-subunit RNA polymerase (RNAP) recognizes the promoter sequences located upstream of the operons and initiates the mRNA polymerization reactions. During the process of RNA polymerization, the RNAP makes a stable yet dynamic complex with the DNA and the RNA called the elongation complex (EC) (Mooney et al., 1998; Wilson and von Hippel, 1994). The EC then extends the RNA chain till it reaches a pause site or a termination signal, where it becomes unstable and dissociates (Gusarov and Nudler, 1999; von Hippel and Yager, 1992) from the DNA template that marks the end of the transcription process. The termination occurs via two ways. 1. Intrinsic (Rho-independent) termination. 2. Extrinsic (Rho-dependent) termination. Intrinsic termination occurs at a DNA sequence that transcribes a RNA forming a hairpin loop. The Rho-dependent termination requires the hexameric motor protein Rho that dislodges the EC.

Figure 1. The bacterial transcription cycle: Four stages: promoter recognition, initiation, elongation, and termination. Bacterial holoenzyme is made up of a core consists of $\alpha$, $\beta$, $\beta'$ and $\omega$ subunits. The core enzyme selects one of the seven sigma factors to form a holoenzyme. Once RNAP-sigma factor holoenzyme recognizes promoter, it leads to the formation of the open complex from the closed complex, and the RNA synthesis initiates by the formation of the transcription initiation complex (TIC). After the release of the sigma factor, the transcription cycle proceeds to elongation by the formation of the transcription elongation complex (TEC) that leads to the RNA synthesis. During the elongation, the EC may pause or is arrested depending on the DNA sequences. At the end, the transcription termination occurs via two ways. 1. Intrinsic (Rho-independent) termination. 2. Extrinsic (Rho-dependent) termination. Intrinsic termination occurs at a DNA sequence that transcribes a RNA forming a hairpin loop. The Rho-dependent termination requires the hexameric motor protein Rho that dislodges the EC.
through two types of mechanism in *E. coli*; extrinsic termination and intrinsic termination. The extrinsic termination requires a protein called Rho. Rho is a RNA-dependent ATPase that binds to the mRNA sequence called *rut* (Rho utilization; C-rich region) site following which it translocates along the RNA towards the EC by its ATP dependent helicase activity, and eventually dislodges the RNA from the EC that leads to the transcription termination (Banerjee *et al*., 2006; Mitra *et al*., 2017). The intrinsic termination does not require any *trans* factors; it occurs at the mRNA site where a GC rich stem loop is formed followed by a poly U tract (Peters *et al*., 2011). This RNA-signal causes transcriptional pause, hairpin nucleation, EC disruption and finally the EC dissociation (Peters *et al*., 2011). Both the intrinsic and the extrinsic terminations are also influenced by certain factors, called Nus factors (Sen *et al*., 2014).

The Rho protein was discovered by J.W Roberts (1969). It is a homo-hexamer with each protomer of 46.8 kDa. Rho remains conserved in many species of bacteria and prevents unwanted expressions of downstream genes, which influences many cellular processes. In recent years various physiological roles of Rho has been identified that are comprehensively described by Mitra *et al* (2017). Here, we focus mainly on the mechanism of action of the Rho protein, its role in termination, factors that interact with Rho: and briefly state its physiology.

**Biochemistry and structure of the Rho protein**

Rho protein comprises of 419 amino acids and its functional state is a homohexamer (Figure 2A). However, it may exist in various oligomeric states, depending on the protein concentrations, the ionic strength and the presence of the cofactors like, RNA or ATP. The homo-hexameric form was found to be the principal state in the presence of the cofactor ATP (Geiselmann *et al*., 1992). The crystal structures of Rho revealed that the homohexamer exists as a closed or open hexamer.

**Figure 2. The Rho structure.** (A) Hexameric closed complex and (B) Monomeric structure of the Rho protein showing its PBS, P, Q and R loops structures (PDB ID 3ICE).
The Rho protein has two major domains; N-terminal (NTD) and C-terminal Domains (CTD). The NTD contains the primary RNA- binding site (PBS) encoded by 22–116 amino acids. The amino acid residues, 103-110, form a hydrophobic pocket that binds to the nucleic acids (Skordalakes and Berger, 2003), via van der Waals interactions with the nitrogenous bases of the nucleic acids. This pocket is better fitted for the pyrimidine bases than the purines. This may explain the requirement of Rho for cytosines (Banerjee et al., 2006). The CTD comprises of P, Q and R loops (Figure 2B). The Q-loop and R-loop form the secondary RNA binding sites (SBS) and the P-loop binds the ATP (Banerjee et al., 2006). It is accepted that the RNA initially binds to the PBS in an open hexamer complex (Skordalakes and Berger, 2003) and subsequently gets isomerized into a closed hexamer complex, once the SBS is filled with the RNA and the ATPase activity is activated.

**Rho-dependent termination**

![Figure 3: Rho isomerization](image)

Interconversion of Rho open complex into the closed complex in the presence of NusG, ATP, RNA (PDB ID 3ICE). Closed complex is formed when the RNA reached the SBS. This conversion process is accelerated in the presence of ATP and NusG. The closed complex is competent for the translocase activities. PBS- primary binding sites; SBS- secondary binding site.

The mechanism of Rho-dependent termination has been studied in detailed, and various models have been put forward (Mitra et al., 2017; Peters et al., 2011). The classical model comprises of the following mechanistic steps. (i) Loading of Rho onto the rut sites via its PBS as an open ring structure that isomerizes into a closed complex upon occupancy of the SBS through an unknown mechanism. This closed complex is translocation-competent. (ii) The Rho translocation along the mRNA, which should be and is kinetically, coupled to catch-up the elongating RNA polymerase. (iii) Upon catching-up the EC, the Rho protein disrupts the RNAP-DNA, RNAP-RNA and the RNA:DNA hybrid interactions either by direct collisions with the EC (Dutta et al., 2008) or by pulling the RNA out of the active center of the EC (Koslover et al., 2012).

*Rho loading onto RNA:* The Rho binding sites on the transcripts are called Rho utilization (rut) sites that are recognized by the PBS of the Rho protein. These sites are the RNA sequences
located upstream of the termination zone. It is comprised of C-rich and G-poor sequences and lacks secondary structures (Banerjee et al., 2006; Ciampi, 2006; Peters et al., 2011). In agreement with that, Rho has the highest affinity for poly(C) RNA sequences. The 60-90 nt sequences of the rut site is sufficient to fill the six PBSs, one per protomer. The structural studies have suggested that Rho is loaded onto the rut site of RNA by its N-terminal domain (NTD) as an open-ring form (Figure 3) (Skordalakes and Berger, 2003; Gogol et al., 1991). This event stimulates the following steps where RNA is guided into the central channel of the Rho C-terminal domain (CTD) interacting with the P-, Q- and R- loops, which leads to the formation of a closed ring form (Figure 3). This state activates the Rho ATPase activity and the translocase activity ensues (Thomsen and Berger, 2009). In recent studies, Rho binding was observed to be unaffected by large secondary structures if the later do not interfere with the rut sites (Hollands et al., 2012, 2014; Schwartz et al., 2007b).

The translocation of Rho along the mRNA

Figure 4. Translocation models of Rho along the mRNA. (A) Simple tracking model: Rho leaves the rut site and tracks along RNA. (B) Looping model: Rho remains attached to the rut site and loops out to catch the elongation complex (EC). (C) Tether tracking model: Rho remains attached to the rut site but tracks along the RNA in a zipper-like manner

Several models have been proposed to explain the mechanism of the translocation events of Rho along the RNA (Koslover et al., 2012; Steinmetz and Platt, 1994). The models proposed so far are, simple tracking model, looping model and tethered tracking model (Figure 4). In the simple tracking model, Rho leaves the rut site and moves onto the downstream sequences (Figure 4A). The looping model suggests that, Rho remains bound to rut site on the RNA and the intervening RNA loops out as Rho approaches the EC (Figure 4B). Among the different models, the most likely model could be the tethered tracking model (Figure 4C). According to this model, the Rho PBS does not leave the RNA rut site and tracks along the transcript in a zipper-like manner (Steinmetz and Platt, 1994). This model was supported by single-molecule force-clamp and magnetic tweezer experiments (Gocheva et al., 2015; Koslover et al., 2012). It has
also been demonstrated that Rho makes a 7-nucleotide step during the translocation event along the mRNA (Schwartz et al., 2009).

**Rho-RNA Polymerase interaction**

![Mechanistic models of Rho-dependent terminations](image)

**Figure 5**: Mechanistic models of Rho-dependent terminations (A) RNA-dependent pathway: Rho loads onto the rut site of the RNA transcript and translocates towards the EC. (B) RNAP-dependent model: Rho directly attaches to the RNAP and then is transferred to the rut site of RNA transcript and following which the translocation ensues. In both the pathways, Rho contacts EC by RNA exit channel, which leads to EC dissociation. RNA:RNA polymerase; Pi: inorganic phosphate; EC: elongation complex.

The RNAP usually pauses in the termination zone located downstream of the rut sites, where termination is induced by the Rho most likely by a direct collision with the RNAP approaching via the RNA exit channel. This model of transcription termination based on the RNA-dependent pathway (Figure 5A), invokes a kinetic coupling between the two moving machines, the Rho and the EC. The concept of kinetic coupling involves a direct competition between the translocating Rho on the transcript and the EC. The slowly transcribing or pause-susceptible RNAPs are more amenable to the Rho-dependent termination; likewise the faster moving RNAP is less prone to this termination (Jin et al., 1988, 1992; Jin and Gross, 1988; Shashni et al., 2012).

In a genome-wide ChIP assay, Rho was observed to be associated with RNAP at the promoter region in the absence of mRNA (Mooney et al., 2009a). Later an in vitro study reported a Rho-RNAP interaction (Epshtein et al., 2010; Figure 5B). However, requirement of RNA is a prerequisite for the Rho to be associated with the EC was shown in a subsequent study (Kalyani et al., 2011). Even though a stable association of the Rho and the RNAP could not be proven unequivocally, it could be perceived that a dynamic Rho-RNAP interaction occurs during the transcription termination cycle. To solve this puzzle, it is important to delineate the Rho-interacting domain(s) of RNAP by genetic screening and by fast kinetic measurements in vitro.

**Releasing of RNA transcripts** How Rho releases the RNA transcript from the DNA:RNA hybrid and dislodges the EC is not very clearly understood. However, many models have been...
proposed. According to the hyper-translocation model, Rho applies brute force that forces RNAP to move forward on the DNA template without adding nucleotides. This may cause the collapse of the transcription bubble release of RNA (Dutta et al., 2008). This model is consistent with Rho’s capability to translocate the EC in the presence of a protein roadblock (Park and Roberts, 2006). In the hybrid-shearing model, the translocating Rho breaks the stable DNA:RNA hybrids by pulling out the RNA out of the active center of the EC (Richardson, 2002). Although Rho is known to generate sufficient force to shear a streptavidin-biotin interaction (Schwartz et al., 2007a), we do not know if Rho is capable of generating enough force to shear a non-U-tract hybrid, which is stabilized with the help of RNAP interactions.

![Figure 6. NusG and NusA. Various functional domains of E. coli: (A) NusG and (B) NusA as indicated with the boundaries of each domain are shown by amino acids numbers. Below each of the domain boundaries, homology modeling of the structures of the proteins are shown (NusG: PDB ID 2JVV; NusA: PDB ID 6FLQ). The structure of NusA is taken from its complex with the EC.](image)

**Regulation of Rho-dependent termination**

**Nus Factors**

Nus factors are a set of proteins that have a prominent role in bacterial transcription, translation and DNA repair process; the factors are NusA, NusB, NusE and NusG. They were named Nutilization substances (Nus) for their involvement in the bacteriophage N-mediated antitermination (Sen et al., 2014). These factors were found to be essential for the survival of the *E.coli* with the exception of NusB. NusB deleted strains were found to be conditionally lethal (Bubunenko et al., 2007). Here, we discuss about NusG and NusA that play important roles in Rho dependent termination.

**NusG**

NusG was first discovered as transcription elongation factor in 1992 (Li et al., 1992; Sullivan and Gottesman, 1992). It is a 21 kDa protein having two domains, the CTD and the NTD (Figure
6A). The NTD (1-116 residues) comprises of α-helices surrounding four anti-parallel β-sheets connected to the β-barrel of CTD (123-181). These two domains are linked by a flexible linker region. NusG-NTD interacts with RNAP through the RNAP β’- clamp helices (Belogurov et al., 2009; Mooney et al., 2009b). The CTD of the NusG interacts with NusE protein to play a role in transcription-translation coupling, and also interacts with Rho to help in transcription termination. This domain is also predicted to interact with the RNA through its 27 residue KOW motif (Burmann et al., 2010; Chalissery et al., 2007, 2011; Knowlton et al., 2003; Steiner et al., 2002; Sullivan and Gottesman, 1992; ). While the NTD is bound to the EC, the CTD could interact with different proteins; hence EC-bound NusG may act as a linker bridge between various factors.

NusG is involved in Rho dependent termination both in vivo (Cardinale et al., 2008; Sullivan and Gottesman, 1992) and in vitro (Chalissery et al., 2011; Mooney et al., 2009b). NusG was found to be required for efficient termination in ~20% of the sense and antisense Rho dependent terminators in vivo (Peters et al., 2012). In a purified system, NusG enhances the termination efficiency of the Rho dependent termination (Burns et al., 1999; Li et al., 1993). NusG has been seen to cause early termination in various terminators and enhance termination in certain Rho mutants that are defective for termination. The early termination induced by NusG is due to the increase in the rate of isomerization of open to close complex formation by Rho at the Rho-loading site (Valabhoju et al., 2016) (Figure 7A), and at many weaker terminators, this isomerization rate is enhanced by NusG (Shashni et al., 2014). The NusG assisted Rho-RNA interaction seems to be necessary for terminators, which have suboptimal rut sites. Some terminators are highly dependent on NusG for Rho to function properly, such as the t\textsubscript{rac} terminator (Peters et al., 2012; Shashni et al., 2014). Thus, NusG seems to be responsible for efficient termination in which the terminator is suboptimal for Rho. It is important to identify those terminators and document their common characteristics. It should be noted that NusG does not improve the helicase or ATPase activity (Nehrke et al., 1993). Further detailed analyses of the mechanism of action NusG in the Rho-dependent termination is required.
**Figure 7. Proposed mechanisms of action of NusA and NusG:** (A) Kinetic scheme depicting the isomerization steps of the OC to CC formation in the presence of NusG and NusA. NusG accelerates the forward reaction, whereas NusA competes at the rut site recognition step. Kinetic constants are indicated. (B) Cartoon depicting mode of NusA-Rho competition for the overlapping rut/nut sites.

**NusA**

The *nusA* gene was discovered through the mutations in *E. coli* that prevented λN-dependent bacteriophage λ growth. NusA is a 55 kDa protein conserved in both the archaea and the prokaryota. It consists of 495 amino acid residues (Nudler and Gottesman, 2002; Sen *et al.*, 2014). NusA interacts with RNAP through its NTD (1-137 residue) and has a flexible linker that links NTD to the three subdomains that are capable of binding to RNA: S1 (138-201), KH1 (202-276) and KH2 (277-344) (Borukhov *et al.*, 2005; Gibson *et al.*, 1993; Mah *et al.*, 1999). These are called the SKK domains. C-terminal to the SKK are the acid repeats 1 and 2 [AR1 (345-426) and AR2 (427-495)] (Figure 6B). The AR1 domain interacts with λN to form the λN:AR1 complex which might be involved in antitermination (Bonin *et al.*, 2004). AR2 regulates RNAP binding to upstream promoter elements by forming a complex with the CTD of α-subunit of RNAP (Mah *et al.*, 2000; Schweimer *et al.*, 2011).

During factor-independent transcription termination, NusA interacts with the terminator hairpin structures and stabilizes it (Touloukhonov *et al.*, 2001) and that increases the efficiency of termination. It enhances pausing by stabilizing the pause-hairpin structures (Artsimovitch and Landick, 2000).

NusA mutant showed termination defect at a Rho dependent terminator, *H-19B tR1* (Saxena and Gowrishankar, 2011) that indicates its involvement in the Rho-dependent termination. One may
hypothesize that as NusA binds to both the RNA and the RNAP, it could affect Rho dependent termination. A genome wide expression study indicated that the NusA-deleted and the NusG-deleted strains show similar expression patterns (Cardinale et al., 2008). NusA was observed to inhibit Rho-dependent termination in vitro (Burns et al., 1998; Kassavetis and Chamberlin, 1981). Another mode of action of NusA could be that the NusA binding region on RNA may overlap with that of the Rho binding region. A recent study described that NusA mutants in SKK domain with enhanced RNA binding affinity were able to inhibit Rho dependent termination at the $\lambda_{T_R1}$ (having an overlapping nut site) more efficiently by competing with the Rho loading step (Qayyum et al., 2016). It is expected that there exist overlapping NusA and Rho binding sites on many different mRNAs. Hence, NusA may function as a general negative regulator of the Rho-dependent termination (Figure 7B).

**Cellular events under the control of Rho-dependent termination.**

Genomic analyses such as, microarrays, ChIP-Seq, and proteomics assays (Cardinal et al., 2008; Mooney et al., 2009; Peters et al., 2012; Shashni et al., 2014) revealed that the expressions of about one-third of the operons in the log phase of *E. coli* are suppressed by the Rho dependent termination. Due to the presence of degenerated binding sites on RNA, the Rho protein has acquired a pervasive mode of action leading to controlling many cellular events. The notable ones are repression of unwanted transcription, regulation of RNA remodeling, maintenance of chromosomal integrity and Mg-homeostasis. We hypothesize that Rho would emerge as a pleiotropic master regulator establishing a new paradigm, where pleiotropy is favored over specificity.

In a seminal paper, Cardinale et al. (2008) showed that Rho-dependent termination represses toxic gene expressions (xenogenic gene expressions) from the cryptic prophages residing in the *E. coli* genome. A Rho dependent terminator, $t_{rac}$ (Cardinale et al., 2008; Shashni et al., 2014), regulates the expression of the *kill* gene of the rac prophage present in the *E. coli* genome. In addition to these transcripts from the prophages, pervasive transcriptions from the anti-sense strand are quite wide-spread in bacteria. Genome-wide transcriptome analyses revealed that these unwanted transcriptions are also suppressed by Rho-dependent termination (Peters et al., 2009, 2012).

Binding of Rho to an upstream site could modulate the secondary structures of the immediate downstream regions of the mRNA that could have “domino effect” on the various molecular processes. A famous example in this category is the modulation of the riboswitch RNA structure by Rho-binding. The *Salmonella* Mg$^{2+}$-sensing *mgtA* gene is under the riboswitch control. It was found that under high concentration of Mg$^{2+}$, the riboswitch of *mgtA* in the 5’-UTR assumes a conformation that promotes Rho binding, which in turn leads to repression of transcription of the *mgtA* coding region (Hollands et al., 2012, 2014). Rho is also reported to be involved in the regulatory mechanisms of the Flavin mononucleotide–sensing, *ribB* riboswitch of *E. coli* (Hollands et al., 2012).

The R-loops, a three stranded nucleic acids structure may form during the transcription event, which is harmful for the cell, and is required to be removed. It had been claimed that Rho plays a role in removal of genome wide deleterious R-loops formation (Leela et al., 2013). It is possible that Rho helicase activity could be instrumental in resolving this three stranded structure.
The Bacterial replication is ~20 times faster than transcription, and hence, there are possibilities of head-on collisions between a replication fork and a stalled or slow-moving transcribing RNAP, which could lead to fork collapse and double-strand breaks (DBS) (Washburn and Gottesman, 2011; Dutta et al., 2011). Rho dissociates stalled TEC located in the path of the replication fork, thereby prevents DSBs and maintains the genomic integrity.

**Future directions**

In the last few decades, our understanding of Rho-dependent transcription termination has improved tremendously, but still there are various aspects that are not understood. We sum-up various questions that should be addressed in near future and we believe answering some of these questions are in pursuance in different laboratories across the globe.

Does Rho interact with RNAP during the termination event? If so, is it a specific interaction? How and when it interacts with RNAP during the transcription cycle. Where is the Rho-interaction site(s) on the EC? Does the EC undergo various conformational changes in different stages of the termination events? Does Rho recognize a particular conformation of the EC?

To answer these questions, a detailed structural study of the Rho-EC complex is required to be solved by the cryo-EM techniques. It would be required to employ fast kinetics tools to map the interaction sites of Rho on the EC in a time-dependent manner during the transcription cycle. A genetic screen to identify RNAP mutants defective for interaction with Rho would complement structural and kinetics experiments.

The transcription elongation factor NusG plays very important role in the Rho-dependent termination, especially in vivo. Detailed mechanism of action of NusG during the termination process is not clear to us. We and many other laboratories are asking the following questions. Does it recruit Rho into the rut site as well as to the EC? Does it always remain in contact with the EC in vivo? Does NusG remain in complex with Rho in vivo? What are the functional sites of Rho for the NusG-CTD during the termination event?

NusG is required at various terminators with the suboptimal Rho-binding sites (rut sites). The properties of those NusG-dependent rut sites are not known, which is a pertinent question that needs to be solved. A combination of genomics and bioinformatics techniques would be needed to address this issue. A mechanistic question that remains to be answered is the functional stoichiometry of the Rho-NusG complex and the basis of recognition of a single site for NusG-binding out of the six equivalent sites on the Rho hexamer. Does NusG prefer the unsymmetrical open complex of Rho? Other fundamental question that needs to be answered is the nature of information that is transmitted to the EC via the NusG-NTD upon Rho-NusG CTD interactions.

**Acknowledgement**

We are grateful to Dr. Amit Kumar for preparing different structural models of Rho, NusA and NusG proteins. We also thank all the members of the laboratory to go through the manuscript. PIRC and AK are the DBT junior research fellows. The research activities of the laboratory of RS is supported by the SERB grant: EMR/2015/001620 and DBT grant: BT/PR10466/BRB/10/1271/2013 and by the CDFD intramural grants.
References


