Can Activation of Rho Factor Through RNA Binding-Induced Ring Closure be Identified by Cryo-EM?

Jonathan Rangel, Mehmet Sen, William Widger

Department of Biology & Biochemistry, University of Houston

Abstract

The rho transcription termination factor is an ATP-dependent, hexameric helicase responsible for terminating up to 50% of mRNA synthesis in *Escherichia coli*. Rho activates by binding to *rut* (*r*ho *ut*ilization) sites on nascent RNA and translocating to the RNAP complex, thereby terminating transcription. Rho has been proposed to have an inactive 'lock-washer' structure that is poised to load mRNA, and an active closed-ring structure responsible for translocation. *rut* sites are known to be at least 70 to 80 nt in length and have a high content of cytosine residues but no consensus sequence have been identified amongst known *rut* sites. We hypothesize Rho binds cytosine at each of the six primary binding sites and there is a minimum of 11 nucleotides between each cytosine residue to bind and activate Rho. Rho is not an on/off switch, but a rheostat that preferentially terminates strong *rut* sites, based on the cytosine spacing on the RNA, based upon the transition of Rho to its closed-ring state. To determine the activity of Rho, we have devised several synthetic RNAs with differing cytosine spacings that will be used to test the degree of ring closure by cryo-EM.

Introduction

The *Escherichia coli* Rho protein is responsible for terminating up to half of all RNA transcription synthesis and has been shown to be essential for viability. Experiments using bicyclomycin, a specific inhibitor of Rho, or Rho-null mutants of other species have shown that Rho plays an important role in regulating global gene expression (1,2). Rho has shown to be involved in regulating virulence genes in pathogenic bacteria (3), sporulation and biofilm formation in *Bacillus subtills* (4), and stress response systems (5). Rho factor activates by binding to C-rich nascent RNAs with low secondary structure known as *rut* (*r*ho *ut*ilization) sites. The *rut* sites can be located at the end of operons to indicate the end of transcription, but they can also be located before operons, within coding regions, and on the antisense strand. *rut* sites located before operons act as attenuation controls, preventing transcription unless Rho is inhibited. *rut* sites located within coding regions are responsible for polarity mutations, whereby missense and nonsense mutations lead to a decrease in downstream expression. Finally, antisense *rut* sites may act to prevent pervasive transcripts which can act as RNA silencers. Identifying genes under Rho control has proven difficult since no consensus exists among the known *rut* sites. Maximal activity of Rho can be achieved by using poly(C) RNA, but not with other poly-nucleotides (6), suggesting that Rho binds to the cytosine residues. We theorize that the 'consensus sequence' is a repeating pattern of C-$(X)_n$ that place cytosine residues at preferred orientation for binding activation of Rho. Based on the structure of Rho (7), the distance between two primary binding sites (Phe62) is 38.36 Å, suggesting a minimum of 11 nucleotides between the C residues is required to bind and activate Rho's ATP hydrolysis activity. An algorithm looking for the most abundant Y-$(X)_n$ gap sequence with a six anchor points revealed that the C-$(X)_{11}$ sequence was the most abundant pattern. To understand the cytosine placement pattern for Rho we have created algorithms to search and score *rut* sites using 10, 11, and 12 gap sizes in the *E. coli* genome.

Rho-Dependent Transcription Termination

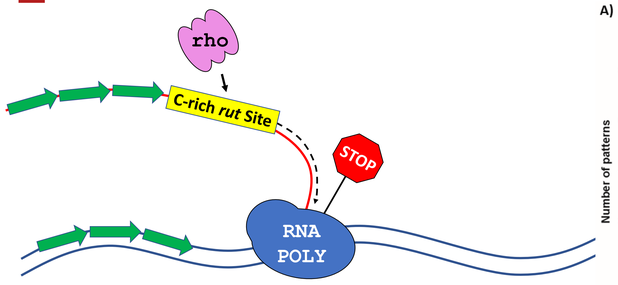

Figure 1. Diagram of rho mediated transcription termination
Rho is a hexameric ATPase motor protein that binds to C-rich nascent RNA without secondary structure. It translocates along the RNA until reaching the RNAP:DNA:RNA elongation complex after which it disrupts and dislodges the RNA polymerase through a yet unknown mechanism.

Background: Identifying the C-$(X)_{11}$ Gap Sequence in the *trp t'* Mutants

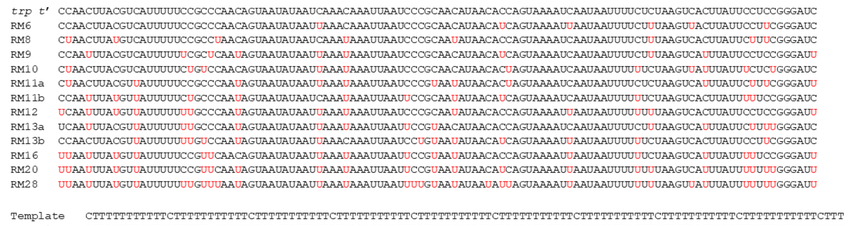

Figure 2. Uracil substitution of cytosine of the *trp t' rut* site leads to decreased Rho activity
A) Uracil was substituted for cytosine to determine how the loss of cytosine affects rho function. The 5' sequence of a known *rut* site called *trp t'* (112 bp) is shown along with the cytosine to uracil substitutions. The position of the cytosine-to-uracil mutations are indicated in red. Mutant *trp t'* sequences are arranged by decreasing cytosine content (Zalatan and Platt, 1992). B) The optimal interval of cytosine was assumed to be every twelve residues. This template was used to score the cytosine residue placement on the mutated *trp t'* RNA sequences.

Background: Scoring the C-$(X)_{11}$ Gap Sequence in the *trp t'* Mutants

	1	2	3	4	5	6	7	8	9	10	Score
trp t'	5	4	8	8	7	6	5	6	5	5	59
RM6	5	3	6	5	5	6	5	6	4	4	49
RM8	4	4	7	7	6	5	4	3	4	3	47
RM9	4	2	5	3	4	3	4	5	5	5	40
RM10	3	2	6	7	6	5	4	4	3	4	44
RM11a	3	2	4	5	6	5	4	2	4	4	39
RM11b	3	2	6	6	5	3	3	4	4	3	39
RM12	3	1	3	2	3	3	3	3	4	3	28
RM13a	3	2	5	4	5	3	3	2	3	4	34
RM13b	3	1	5	5	6	5	3	4	3	4	39
RM16	1	1	3	4	3	3	3	3	3	2	26
RM20	1	0	1	2	2	3	3	2	2	1	17
RM28	0	0	0	0	0	0	0	0	0	0	0

	Score	% term. *in vivo*	% term. *in vitro*	K_{diss} (nM)	K_{assoc} (nM)	ATPase Activity (moleATP/min/mol rho)
trp t'	59	99.8	92	0.02	50.00	228
RM6	49	99	84	0.047	21.28	108
RM8	47	98.7	82	0.063	15.87	92
RM9	40	96	84	0.059	16.95	60
RM10	44	99.3	82	0.072	13.89	116
RM11a	39	79.5	14	0.096	10.42	58
RM11b	39	73.5	79	0.084	11.90	94
RM12	28	99.5	75	0.086	11.63	48
RM13a	34	33	71	0.085	11.76	50
RM13b	39	48	80	0.075	13.33	56
RM16	26	45	34	0.102	9.80	48
RM20	17	0	0	0.186	5.38	38
RM28	0	0	0	1.98	0.51	0

Figure 3. Termination sites based on (*rut*) *rho ut*ilization site location
A) The 112 bp template was used to score target RNA sequences. The target sequence was aligned to the scoring template sequence beginning at the first residue being score. Cytosine present on the target sequence at the position corresponding to the C in the scoring template or a position +- one from the template C was given a subscore of 1 and summed over the nine positions. The scoring template was shifted one residue to the right and score calculated again. This was repeated 10 times and the subtotals summed. The subscores and the total scores for the *trp t'* and the *trp t'* mutant templates are presented in (A). The total score for *trp t'* and all the variants decreased with a decreasing number of cytosine residues. B) The relation of the scoring parameter among all the variants with the % in vivo and in vitro transcription termination, RNA-rho dissociation constant (K_{diss}), RNA-rho association constant (K_{assoc}), and rho ATPase activation is presented.

Location of *rut* Sites

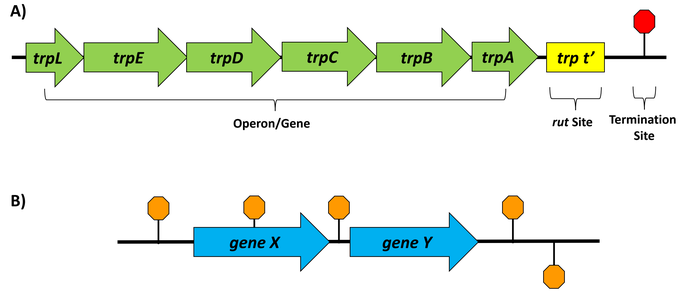

Figure 4. Termination sites based on (*rut*) *rho ut*ilization site location
A) The *trp t' rut* site is located at the end of the trp operon. Termination sites are located downstream of the *rut* binding site. B) Rho-dependent termination sites are not only located at the end of genes. They can also be located before genes for attenuation, within genes for polarity, at the end of genes for classical termination, and on the antisense strand.

No Consensus Sequence to Known *rut* Sites

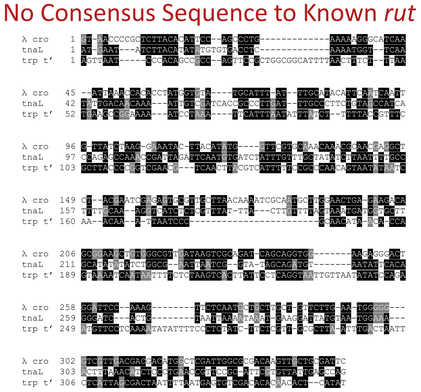

Figure 5. No consensus sequence among the known *rut* sites
The λ cro, tnaL, and *trp t' rut* sites have no discernable consensus sequence. An explanation to this could be that the sequences are actually repeating patterns that are not easily identifiable through sequence alignment.

Creating Synthetic *rut* Sites

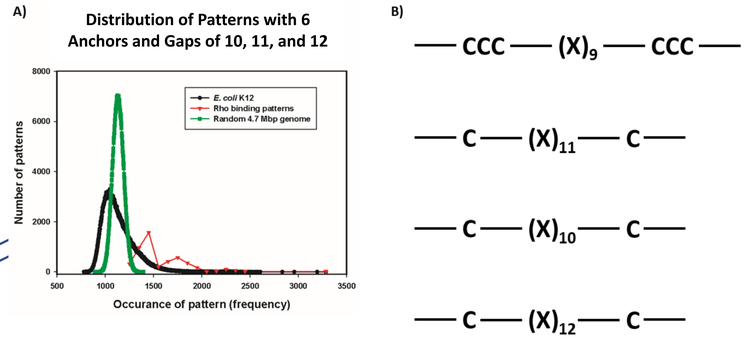

Figure 6. Distribution of rho-binding patterns is the most common pattern
A) An algorithm identified the frequency of 6 anchors with gaps of 10, 11, or 12 nucleotides and found that the most common pattern was having C-anchors. B) Four gap patterns were designed for synthetic *rut* sites. The first contains three C-anchors at each anchor point with a gap of 9 nucleotides between the central C-anchors, keeping the central anchors 11 nucleotides apart. The other patterns have singular C-anchors with different gap spaces of either, 10, 11, or 12 nucleotides.

RNA-Induced Ring Closure

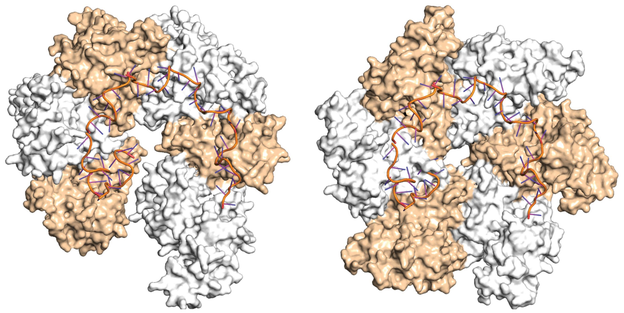

Figure 7. Rho has an RNA-loading open form and a RNA-translocating closed form
Structural analysis of the rho protein shows that it has two conformations. The open lock-washer form is poised for RNA loading of the *rut* site at its primary binding site. Binding of the *rut* site and loading of the RNA onto the secondary binding site within the loop induces a conformational change. The closed-ring form has the RNA loaded through the ring and is optimal for RNA translocation. We believe that different cytosine spacing arrangements induce different degrees of ring closure, thereby linking the *rut* site directly to the ATPase activity.

Conclusion

The algorithms have determined that the most frequent pattern in the *E. coli* genome to be a repeating C-$(X)_{11}$ pattern (8). A second program using a C-$(X)_{11}$, C-$(X)_{10}$, and C-$(X)_{12}$ pattern has identified over 9000 possible *rut* sites. We have created synthetic *rut* sites with differing cytosine spacing that will be used in ATP hydrolysis assays and in vitro termination assays. We deduce that the degree of ring closure is directly proportional to the binding affinity and the activation of ATP hydrolysis. A "strong" *rut* site, would have all C-anchors ideally spaced to bind to the primary binding site on Rho, would have maximal ATP hydrolysis activity, and would induce the greatest degree of ring closure. "Weak" *rut* sites on the other hand, would have less activity, and would induce a weaker degree of ring closure. The control of termination efficiency is a proposed method of transcriptional control over gene expression. Creating a Rho-RNA complex with the synthetic *rut* sites with specific gap sizes could reveal the mechanism by which Rho loads RNA and induces its conformational change to the closed-ring state.

References

1. Peters, J. M., Mooney, R. A., Kuan, P. F., Rowland, J. L., Keles, S., and Landick, R. (2009) Rho directs widespread termination of intragenic and stable RNA transcription. *Proc. Natl. Acad. Sci. USA* **106**, 15406-15411
2. Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E., and Nudler, E. (2008) Termination factor Rho and Its cofactors NusA and NusG silence foreign DNA in *E. coli*. *Science* **320**, 935-938
3. Botella, L., Vaubourgeix, J., Livny, J., and Schnappinger, D. (2017) Depleting Mycobacterium tuberculosis of the transcription termination factor Rho causes pervasive transcription and rapid death. *Nature communications* **8**, 14731
4. Bidnenko, V., Nicolas, P., Grylak-Mielnicka, A., Delumeau, O., Auger, S., Aucouturier, A., Guerin, C., Francis, R., Bardowski, J., and Aymerich, S. (2017) Termination factor Rho: from the control of pervasive transcription to cell fate determination in Bacillus subtilis. *PLoS genetics* **13**, e1006909
5. Italiani, V. C. S., Zuleta, L. F. G., and Marques, M. V. (2002) Transcription termination factor rho is required for oxidative stress survival in *Caulobacter crescentus*. *Mol. Microbiol.* **44**, 181-195
6. Richardson, J. P. (1982) Activation of rho protein ATPase requires simutaneous interaction at two kinds of nucleic acid-binding sites. *J. Biol. Chem.* **257**, 5760-5766
7. Skordalakes, E., and Berger, J. M. (2003) Structure of the Rho transcription terminator: Mechanism of mRNA recognition and helicase loading. *Cell* **114**, 135-146
8. Zalatan, F., and Platt, T. (1992) Effects of decreased cytosine content on rho interaction with the rho-dependent terminator *trp t'* in *Escherichia coli*. *J. Biol. Chem.* **267**, 19082-19088