Published in final edited form as:

Science. 2008 May 16; 320(5878): 935–938. doi:10.1126/science.1152763.

Termination Factor Rho and Its Cofactors NusA and NusG Silence Foreign DNA in *E. coli*

Christopher J. Cardinale^{1,*}, Robert S. Washburn^{2,*}, Vasisht R. Tadigotla^{3,6}, Lewis M. Brown⁴, Max E. Gottesman^{2,5,**}, and Evgeny Nudler^{1,**}

¹Department of Biochemistry, New York University School of Medicine, New York, NY 10016, USA

²Department of Microbiology, Columbia University Medical Center, New York, NY 10032, USA

³BioMaPS Institute for Quantitative Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

⁴Comparative Proteomics Center, Department of Biological Sciences, Columbia University, New York, NY 10027

⁵Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA

Abstract

Transcription of the bacterial genome by the RNA polymerase must terminate at specific points. Transcription can be terminated by Rho factor, an essential protein in enterobacteria. We utilized the antibiotic bicyclomycin, which inhibits Rho, to assess its role on a genome-wide scale. Rho is a global regulator of gene expression that matches *E. coli* transcription to translational needs. We found that genes in *E. coli* that are most repressed by Rho are prophages and other horizontally-acquired portions of the genome. Elimination of these foreign DNA elements increases resistance to bicyclomycin. Although *rho* remains essential, such reduced-genome bacteria no longer require Rho cofactors NusA and NusG. Deletion of the cryptic *rac* prophage in wild-type *E. coli* increases bicyclomycin resistance and permits deletion of *nusG*. Thus, Rho termination, supported by NusA and NusG, is required to suppress the toxic activity of foreign genes.

The complete genome sequence of *E. coli* revealed that 90% of its nucleotide sequence could encode protein (1). The remaining non-coding genome is densely packed with regulatory signals for transcription initiation and termination. This high information density requires that transcription terminate precisely at operon ends to avoid interference with neighboring transcription units. Based on sequence, approximately half of the transcription units, or operons, in *E. coli* are predicted to end with a specific structure, an intrinsic terminator, consisting of a hairpin followed by several U residues at the 3' terminus of the RNA. This structure alone is sufficient to dissociate the polymerase elongation complex *in*

^{**}Corresponding authors: evgeny.nudler@med.nyu.edu (Evgeny Nudler) and meg8@columbia.edu (Max Gottesman).

⁶Current address: Department of Physics, Boston University, Boston, MA 02215

*These authors contributed equally to this work.

vitro (2, 3). In contrast, transcription termination of the remaining half of operons could not be predicted from DNA sequence, and has been generally assumed to rely on an ATP-dependent RNA-DNA helicase known as Rho factor. In the decades since its discovery (4) Rho has been well-studied biochemically and structurally (2, 5–8), but its role as a biological regulator is still unclear. Rho factor recognizes no specific consensus but rather binds to naked, untranslated RNA, favoring C-rich sites that contain little secondary structure (9–11). Rho-dependent termination sites occur frequently in operons. For example, Rho can stop transcription when the end of the coding information is reached (12), attenuate transcription conditionally at the beginning of operons (13), and even within open reading frames when messenger RNA is uncovered by a nonsense mutation (14). In each case, the hypothesized roles of Rho are to prevent transcription from impinging on neighboring operons, to prevent the wasteful production of unusable transcripts, and to recycle polymerases promptly to locations where they are needed. However, since only a handful of Rho terminators (less than ten) have been actually located and characterized (15), there is still much to be learned about the role of Rho-dependent termination *in vivo*.

To investigate the biological role of Rho, we assayed gene expression using the Affymetrix *E. coli* Genome 2.0 array, an in-situ synthesized oligonucleotide array covering the entire genome of four evolutionarily divergent *E. coli* strains: the laboratory strain K-12 MG1655, the enterohemorrhagic strains O157:H7 (EDL933 and Sakai), and uropathogenic CFT073. Specific and potent inhibition of Rho can be achieved rapidly by treatment with the antibiotic bicyclomycin (BCM) (16). An advantage of chemical over genetic intervention is that the transcriptome content of control and experimental cultures remains identical until the moment the inhibitor is added. Indeed, total inhibition of Rho termination activity cannot be achieved by genetic manipulation since *rho* is an essential gene (17). BCM is highly specific to Rho; it rapidly permeates cells and has no other known *in vivo* targets (16, 18). Changes in gene expression in response to BCM reflect, therefore, a snapshot of Rho activity. Treatment of MG1655 with a series of concentrations of BCM for short time intervals revealed a pervasive change in gene expression (Fig. 1). One theme which emerges from the array data is a widespread increase in the expression of genes derived from recent horizontal transfer into the genome of K-12 from other species or from defective bacteriophage (Fig. 1B, 1D, and S2B). Based on whole-genome alignment, approximately 14–18% of the K-12 genome differs from other families of *E. coli*, tending to occur in contiguous blocks known as K-islands (19, 20). K-islands are characterized by an altered GC/AT content, distinct codon preference, and reduced evolutionary conservation. The genomic islands are enriched in defective prophages, transposons, and insertion sequences (21). Comparing MG1655 with the enterohemorrhagic strain O157:H7 (EDL933) shows that the two strains possess 3658 genes that are nearly identical in sequence, as well as 648 and 1769 unique genes, respectively (22). As shown in Fig. 1B and D, the genes unique to each strain and prophage genes tended to be upregulated, with half of these genes increasing expression more than threefold. By contrast, a quarter of orthologous genes, common between the two strains, were upregulated greater than threefold (compare orthologous and K-12-specific genes in Fig 1E).

We find that expression of the non-coding intergenic (IG) regions is in general increased by Rho inhibition (Fig. 1F). Of the IG probes which were reproducibly measured, as selected by Significance Analysis of Microarrays at the 1% false discovery rate (23), 72% were increased at least three-fold and only 1% were decreased at least three-fold. The general upregulation of IG regions confirms that Rho has a global role in preventing synthesis of untranslated transcripts. Taken together, the array data from BCM treatment of *E. coli* indicate that Rho is intimately involved in operon regulation throughout the genome and is not only acting on a rare subset of genes or when translation terminates abnormally.

We next sought to determine if this extensive perturbation in the transcriptome was reflected in the proteome. We utilized the technique of difference gel electrophoresis (DIGE) to analyze the protein complement of MG1655 cells treated under the same conditions used in the microarray experiments (24). The workflow for this analysis is shown in Fig. S6A. Two-dimensional gels of fluorescently-labeled proteins show that of 3341 unique spots analyzed, 101 were increased greater than two-fold and 8 were decreased greater than two-fold by BCM treatment. Altered spots were robotically excised from gels and the proteins were identified by mass spectrometry (Fig. 2). As shown in Fig. 2 and Tables S1 and S2, among the most affected unique proteins is Rho itself and the RecE protein of the *Rac* prophage. For reasons not understood, most of the other proteins identified are involved in anaerobic metabolism and the response to acidic pH. Based on the microarray result, that many *de novo* transcripts of unique genes were being produced, we expected to see many new spots appearing on the gel from the BCM treated sample. However, this did not occur. The proteomic results corroborate the role of Rho as a general inhibitor of transcription under normal growth conditions. There is a profound excess of transcriptional output over translational needs when Rho activity is reduced. The lack of perturbation of the proteome also suggests that protein expression is frequently controlled post-transcriptionally.

Since Rho strongly represses transcription of horizontally-transferred genes, we investigated a synthetic *E. coli* strain, MDS42, that lacks these genes. 14% of the MDS42 genome has been removed by targeted deletion of prophages, IS elements, and K-island clusters (25, 26). Fig. 3B shows that MDS42 was $\sim 10^4$ -fold more resistant to BCM (25 $\mu\text{g/ml}$) than the parent strain, MG1655. MG1655 contains the remnants of a lambdoid bacteriophage known as *rac*. This defective prophage carries a *kil* gene encoding an inhibitor of cell division. Deletion of *rac* alone produced levels of BCM resistance comparable to the MDS42 strain. The resistance was conferred by deletion of *kil* and the remaining downstream operon, but not DNA downstream of *kil* (Fig. 3B).

Studies of λ phage revealed that endogenous host proteins NusA, NusB, NusE (ribosomal protein S10) and NusG were required for the λ N protein to suppress transcription termination on the phage chromosome (27, 28). NusA and NusG have been implicated in both Rho-dependent and intrinsic termination and are essential for *E. coli* to survive (29, 30). It was possible to delete both *nusA* and *nusG* in strain MDS42, although stationary phase survival was poor, and the strains were highly sensitive to BCM (Fig. 3B). Deleting *nusA* or *nusG* also adversely affected growth rate, increasing the doubling time in rich media from 32 min to 57 min and 68 min, respectively. Unexpectedly, it was possible to transfer

the *nusG* knockout allele to a wild-type MG1655 strain lacking the *rac* prophage alone, indicating that suppression of *rac* gene expression is the critical function of NusG.

Strains lacking NusA or NusG are highly similar in their overall pattern of gene expression as shown by the hierarchical cluster analysis in Fig. 3A and the scatterplot of intergenic region expression in supplemental Fig. S2C and D. We therefore conclude that these proteins normally act in concert, recognizing the same elongation and termination signals.

To understand the basis of how Rho inhibition could affect gene expression on such a pervasive scale we examined two specific operons, one in the *rac* prophage and one in λ . The maps in Fig. 4A and B show that the leftward operons of *rac* and λ are homologous, implying that there should be a Rho-dependent terminator (t_{imm}) in *rac* following the *racR* gene, as there is in λ . Addition of BCM enables the RNA polymerase to continue through this terminator and express downstream genes, including the toxic *kil* gene. RT-PCR analysis reveals the elongated transcript (Fig. 4C). Similarly, the leftward operon of λ (Fig. 4B) exhibits readthrough of the t_{imm} terminator into the downstream *N::lacZ* reporter fusion in the presence of BCM (Fig. 4D). The HK022 Nun termination protein, which blocks transcription elongation at the λ *nutL* site, prevents reporter gene expression, consistent with a transcript originating from the P_{RM} promoter (31).

As shown by the maps of prophages in Fig S4, genes that are upregulated by treatment with BCM tend to occur in consecutive series in the same strand orientation, suggesting that preventing read-through into neighboring operons is an important function of Rho. Rho's bias toward suppressing foreign DNA could be related to the lower density of the Rho-independent intrinsic terminators in the K-island regions. Utilizing the terminator-prediction model of Lesnik et al. (32), there is an intrinsic terminator on average every 4.0 kb in the conserved regions of the genome, but only every 8.5 kb in the K-islands (Table S7) (32). Moreover, the genes upregulated by treatment with BCM tend to be more AT rich than the genome as a whole (Fig. S1) and have a lower Codon Adaptation Index (Fig. S5B). The lower secondary structure of AT-rich RNA could make it a favored target of Rho-dependent termination in spite of Rho's *in vitro* binding affinity for C-rich RNA, whereas the suboptimality of translation in genes with poor codon preference leaves them open to Rho.

In summary, our results reveal Rho factor as a global regulator of bacterial gene expression under normal growth conditions. Rho serves the crucial role of maintaining transcriptional boundaries throughout the genome. In particular, Rho is responsible for silencing horizontally transferred DNA elements, some of which are detrimental to the host. Recently, H-NS protein has been implicated in selective silencing of foreign DNA in *Salmonella* by acting at the level of promoter initiation (33). Rho-dependent termination may represent a separate "immunity" system that protects bacterial cells from the harmful activity of certain foreign genes. The existence of such different defensive tools against new acquisitions to the genome underscores the importance of this phenomenon for bacterial evolution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the NIH grants R01 GM58750 and GM72814 (E.N.), R01 GM37219 (M.G.), and Office of the Vice President for Arts and Sciences, Columbia University (L.B.). The authors are grateful to the NYU Cancer Institute Genomics Facility for processing of arrays. Microarray data discussed in this paper is available from the NCBI Gene Expression Omnibus accession number GSE10345.

References and Notes

1. Blattner FR, et al. *Science*. Sep 5.1997 277:1453–74. [PubMed: 9278503]
2. Nudler E, Gottesman ME. *Genes Cells*. Aug.2002 7:755–68. [PubMed: 12167155]
3. Platt T. *Annu Rev Biochem*. 1986; 55:339–72. [PubMed: 3527045]
4. Roberts JW. *Nature*. Dec 20.1969 224:1168–74. [PubMed: 4902144]
5. Browne RJ, Barr EW, Stitt BL. *J Biol Chem*. Apr 8.2005 280:13292–9. [PubMed: 15703178]
6. Geiselmann J, Wang Y, Seifried SE, von Hippel PH. *Proc Natl Acad Sci U S A*. Aug 15.1993 90:7754–8. [PubMed: 7689228]
7. Richardson JP. *Cell*. Jul 25.2003 114:157–9. [PubMed: 12887917]
8. Skordalakes E, Berger JM. *Cell*. Nov 3.2006 127:553–64. [PubMed: 17081977]
9. Hart CM, Roberts JW. *J Biol Chem*. Dec 15.1991 266:24140–8. [PubMed: 1721066]
10. Platt T. *Mol Microbiol*. Mar.1994 11:983–90. [PubMed: 8022288]
11. Richardson LV, Richardson JP. *J Biol Chem*. Aug 30.1996 271:21597–603. [PubMed: 8702947]
12. Court D, et al. *J Mol Biol*. Apr.1980 138:231–54. [PubMed: 6447791]
13. Stewart V, Landick R, Yanofsky C. *J Bacteriol*. Apr.1986 166:217–23. [PubMed: 2420781]
14. Adhya S, Gottesman M. *Annu Rev Biochem*. 1978; 47:967–96. [PubMed: 354508]
15. Ciampi MS. *Microbiology*. Sep.2006 152:2515–28. [PubMed: 16946247]
16. Zwiefka A, Kohn H, Widger WR. *Biochemistry*. Apr 13.1993 32:3564–70. [PubMed: 8466900]
17. Das A, Court D, Adhya S. *Proc Natl Acad Sci U S A*. Jun.1976 73:1959–63. [PubMed: 132662]
18. Vincent F, et al. *Biochemistry*. Aug 8.2000 39:9077–83. [PubMed: 10924100]
19. Ochman H, Lawrence JG, Groisman EA. *Nature*. May 18.2000 405:299–304. [PubMed: 10830951]
20. Perna NT, et al. *Nature*. Jan 25.2001 409:529–33. [PubMed: 11206551]
21. McClelland M, et al. *Nature*. Oct 25.2001 413:852–6. [PubMed: 11677609]
22. Glasner JD, et al. *Nucleic Acids Res*. Jan.2008 36:D519–23. [PubMed: 17999997]
23. Tusher VG, Tibshirani R, Chu G. *Proc Natl Acad Sci U S A*. Apr 24.2001 98:5116–21. [PubMed: 11309499]
24. Rivard CJ, et al. *J Biol Chem*. Mar 2.2007 282:6644–52. [PubMed: 17200116]
25. Kolisnychenko V, et al. *Genome Res*. Apr.2002 12:640–7. [PubMed: 11932248]
26. Posfai G, et al. *Science*. May 19.2006 312:1044–6. [PubMed: 16645050]
27. Mah TF, Li J, Davidson AR, Greenblatt J. *Mol Microbiol*. Nov.1999 34:523–37. [PubMed: 10564494]
28. Zhou Y, Filter JJ, Court DL, Gottesman ME, Friedman DI. *J Bacteriol*. Jun.2002 184:3416–8. [PubMed: 12029062]
29. Schmidt MC, Chamberlin MJ. *J Biol Chem*. Dec 25.1984 259:15000–2. [PubMed: 6096352]
30. Sullivan SL, Gottesman ME. *Cell*. Mar 6.1992 68:989–94. [PubMed: 1547498]
31. Washburn RS, Wang Y, Gottesman ME. *J Mol Biol*. Jun 13.2003 329:655–62. [PubMed: 12787667]
32. Lesnik EA, et al. *Nucleic Acids Res*. Sep 1.2001 29:3583–94. [PubMed: 11522828]
33. Navarre WW, et al. *Science*. Jul 14.2006 313:236–8. [PubMed: 16763111]
34. Eisen MB, Spellman PT, Brown PO, Botstein D. *Proc Natl Acad Sci U S A*. Dec 8.1998 95:14863–8. [PubMed: 9843981]

Fig. 1. Genomic response of divergent *E. coli* strains to Rho inhibition

(A) Hierarchical cluster analysis of a concentration gradient of BCM (doses of 10, 25, and 100 $\mu\text{g/ml}$) in *E. coli* K-12 strain MG1655 showing only genes orthologous between K-12 and enterohemorrhagic *E. coli*. Arrays (columns) are shown in biological duplicates, normalized so that average of each gene on the untreated control arrays is equal to 1 and expression in treated cultures is displayed as a ratio of treated to untreated. Yellow blocks represent upregulation by BCM and blue represents downregulation. (B) Response to BCM of genes present in K-12 *E. coli* but absent from enterohemorrhagic *E. coli*, displayed as in (A). (C) Hierarchical cluster analysis of the response of orthologous genes in enterohemorrhagic *E. coli* O157:H7 strain EDL933. (D) Response to BCM treatment of genes present in enterohemorrhagic *E. coli* but absent from K-12. (E) Expression of ORFs in response to BCM displayed as a scatterplot of probe intensity in the control array (X-axis) and BCM-treated array (Y-axis) from a representative pair of arrays. The diagonal line represents equal probe hybridization intensity between both arrays; points above the diagonal are genes upregulated by treatment with BCM and points below the diagonal are

downregulated. The red lines located at 100 intensity units represent the threshold below which probe-level analysis is 90% likely to call the probe absent. Therefore, probes in the upper left quadrant are ORFs whose expression was induced *de novo*. Grey points are orthologous genes and violet points are K-12-specific genes. **(F)** Scatterplot of probe intensity for intergenic (IG) regions of MG1655 after treatment by BCM.

Fig. 2. Proteomic response to Rho inhibition as detected by difference gel electrophoresis (DIGE)
Two-dimensional electrophoresis gel of protein extracted from BCM-treated and control cultures. Control protein is pseudocolored green and BCM-treated protein as red. Differentially-expressed proteins (indicated by name) were identified by mass spectrometry.

Fig. 3. Reduced-genome *E. coli* is resistant to Rho inhibition and deletion of elongation factors NusA and NusG

(A) Hierarchical cluster analysis of ORF gene expression in strain MDS42, BCM-treated MDS42, MDS42 $\Delta nusA$, and MDS42 $\Delta nusG$. Probe intensity is normalized to the untreated MDS42 strain. (B) Efficiency of colony formation assay of the indicated strains. Cultures at dilutions of 10^{-2} , 10^{-4} , and 10^{-6} were spotted onto a control plate or a plate containing BCM at 25 $\mu\text{g/ml}$. *int-kilR* and *int-ydaE* are fragments of the *rac* prophage that were deleted.

Figure 4. Effect of BCM on the leftward operons of *rac* and λ phages

(A) Map of the leftward operon of the *rac* prophage. Gray arrows: genes; open-headed arrows: PCR primers; brackets: deletions of *intR-ydaE* and *intR-kil*; bent arrow: the operon's promoter, (P_{RM}). (B) Map of the homologous operon of λ phage. Dashed lines show proposed transcripts produced. (C) RT-PCR using primer pairs indicated on the map in panel A shows that BCM treatment yields an elongated transcript. RT: reverse transcriptase. (D) Average β -galactosidase activity (*lacZ* expression) from the phage strain shown in panel B. The standard error is less than 5%.