

Widespread Antisense Transcription in *Escherichia coli*

James E. Dornenburg,^a Anne M. DeVita,^a Michael J. Palumbo,^a and Joseph T. Wade^{a,b}

Wadsworth Center, New York State Department of Health, Albany, New York, USA^a and Department of Biomedical Sciences, School of Public Health, University at Albany, Albany, New York, USA^b

J.E.D. and A.M.D. contributed equally to this article.

ABSTRACT The vast majority of annotated transcripts in bacteria are mRNAs. Here we identify ~1,000 antisense transcripts in the model bacterium *Escherichia coli*. We propose that these transcripts are generated by promiscuous transcription initiation within genes and that many of them regulate expression of the overlapping gene.

IMPORTANCE The vast majority of known genes in bacteria are protein coding, and there are very few known antisense transcripts within these genes, i.e., RNAs that are encoded opposite the gene. Here we demonstrate the existence of ~1,000 antisense RNAs in the model bacterium *Escherichia coli*. Given the high potential for these RNAs to base pair with mRNA of the overlapping gene and the likelihood of clashes between transcription complexes of antisense and sense transcripts, we propose that antisense RNAs represent an important but overlooked class of regulatory molecule.

Received 1 February 2010 Accepted 12 March 2010 Published 18 May 2010

Citation Dornenburg, J. E., A. M. DeVita, M. J. Palumbo, and J. T. Wade. 2010. Widespread antisense transcription in *Escherichia coli*. *mBio* 1(1):e00024-10. doi:10.1128/mBio.00024-10.

Editor Stanley Maloy, San Diego State University

Copyright © 2010 Dornenburg et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Joseph T. Wade, jwade@wadsworth.org.

Recent high-throughput sequencing analyses of RNA in eukaryotes have revealed a far more complex network of RNAs than previously appreciated, including thousands of RNAs antisense to protein-coding genes (aRNAs) (1). In contrast, relatively few aRNAs have been identified in bacteria (2). Studies of individual plasmid-encoded and chromosomally encoded aRNAs in a variety of bacterial species have demonstrated that aRNAs can regulate expression of the overlapping gene at the level of translation, mRNA stability, or transcription (3–11). Several studies have hinted at the existence of many more aRNAs, in multiple bacterial species, than those currently described (5, 8, 10, 12–18), suggesting that aRNAs have a widespread regulatory function in bacteria.

We sought to identify novel aRNAs in *Escherichia coli*. We generated a cDNA library by extracting RNA from rapidly growing cells (wild-type strain MG1655 grown with aeration in LB to an optical density at 600 nm [OD₆₀₀] of 0.7), treating the RNA with tobacco acid pyrophosphatase to convert 5'-triphosphate groups to monophosphates, ligating an RNA oligonucleotide (5'-ACACUCUUUCCCUACACGACGCUCUCCGAUCU-3') to the RNA 5' ends, reverse transcribing with a primer in which the nine 3'-end proximal bases are random (5'-GTTTCCCAGTCACGATCNNNNNNNNN-3'), and amplifying by PCR. Using Solexa sequencing, we identified unique RNA 5' ends. The mapped RNA 5'-end locations include many known transcription start sites: 24% of sequences of published transcription start sites are matched exactly by a sequence from our library, and 41% of those sequences are ≤2 bp away from a sequence from our library (19). The exact matches include the majority of known aRNAs (GadY, RyjB, RdlA, RdlD, RyeA, SokB, and SokC). The RNA 5'-end locations also include 1,005 locations that map antisense to protein-coding genes (see Table S1 in the supplemental material), suggest-

ing the existence of many more aRNAs. These putative aRNA 5' ends were each sequenced between 1 and 5,488 times. An additional 385 ends map antisense to known and predicted 5' and 3' untranslated regions (UTR) (see Table S1 in the supplemental material) (20).

The housekeeping σ factor σ^{70} binds a bipartite DNA sequence at *E. coli* promoters during transcription initiation. The downstream recognition site, the -10 hexamer, has the consensus sequence TATAAT and is typically positioned 7 or 8 bp upstream of the transcription start site (21). For the set of 471 published transcription start sites (19), the -10 hexamers match the consensus, on average, 3.28 times out of 6 (-10 match score) (base distribution shown in Fig. 1A). In contrast, 1,000 randomly selected sequences antisense to genes match the consensus only 2.00 times out of 6 (control match score) (base distribution shown in Fig. 1B). This difference is highly significant (Mann-Whitney *U* test, *P* of $8.9e^{-70}$). Furthermore, 46% of the RNAs with published start sites initiate with "A," significantly more than expected by chance ($P < 1e^{-22}$) (Fig. 1A and B). The -10 hexamer sequences for the 1,005 putative aRNAs identified in this work have a -10 match score of 3.27, significantly higher than the control match score (Mann-Whitney *U* test, *P* of $8.8e^{-102}$) (base distribution shown in Fig. 1C). This holds true even for the 141 aRNA 5' ends that were sequenced only once (score of 3.12; Mann-Whitney *U* test, *P* of $2.8e^{-21}$). The -10 match score for the 1,005 aRNAs is not significantly different from that for the set of published start sites (Mann-Whitney *U* test, *P* = 0.49). Moreover, 48% of the putative aRNAs initiate with "A," significantly more than expected by chance ($P < 1e^{-50}$) (Fig. 1B and C) but not significantly different from the set of published start sites (Fisher's exact test, *P* of 0.40) (Fig. 1A and C). Thus, the promoters and

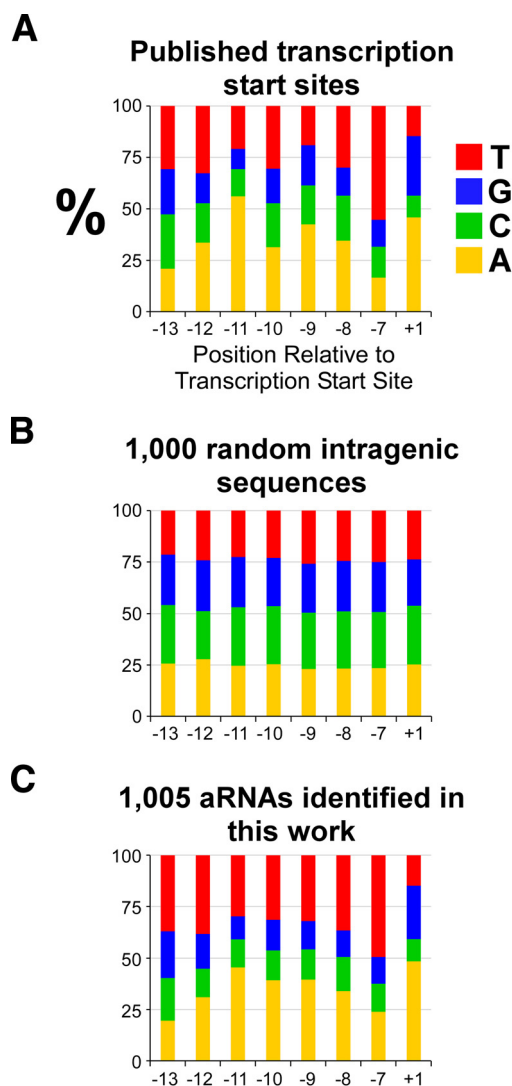


FIG 1 (A) Distribution of nucleotides at the transcription start site (+1) and positions upstream for transcripts with published start sites. Equivalent distributions are shown for 1,000 random intragenic sequences (B) and the 1,005 putative aRNAs identified in this work (C).

transcription start sites of the 1,005 putative aRNAs have DNA sequence properties that are indistinguishable from those of characterized transcripts.

To experimentally validate the putative aRNAs, we fused the promoter regions (up to 200 bp upstream of the putative transcription start site) of 10 aRNAs to a *lacZ* reporter gene and measured expression levels in a β -galactosidase assay. In 9 out of 10 cases tested, we detected *lacZ* expression that was significantly reduced by mutation of the -10 hexamer (Fig. 2A). We conclude that the large majority of putative aRNAs are genuine and that our transcription start site assignments are highly accurate.

We selected two mRNAs, *rplJ* and *yrda*, that each overlap a putative aRNA. We translationally fused the mRNAs in frame to *lacZ*, under control of the natural mRNA promoter, and compared the expression levels of *lacZ* for a wild-type construct and a construct containing a mutated -10 hexamer and +1 nucleotide for the aRNA (+1 nucleotide not mutated for *yrda*). Expression

of *lacZ* increased significantly upon mutation of the aRNA promoter for *rplJ* but not for *yrda* (Fig. 2B). This strongly suggests that the aRNA overlapping *rplJ* represses expression of the mRNA.

Our data demonstrate that (i) antisense transcription is widespread in *E. coli* and (ii) aRNAs can regulate expression of the overlapping gene. Regulation by aRNAs is likely to be widespread, since all previously characterized bacterial aRNAs regulate expression of the overlapping gene (3–11). The majority of aRNAs are likely to be noncoding due to constraints imposed by the overlapping protein-coding sequence. A small fraction of aRNAs may be mRNAs for which the 5'-end UTR is antisense to another gene; however, this is unlikely in most cases, since only 21% of aRNAs initiate ≤ 500 bp upstream of a known translation start site on the same strand. Since they are likely to be noncoding, aRNAs are also likely to be substrates for Rho-dependent termination, which occurs within the first few hundred nucleotides of transcription (14). We conclude that the majority of aRNAs are short (< 500 -nucleotide), noncoding transcripts.

We speculate that most of the novel aRNAs are generated by promiscuous transcription initiation within genes, as has been suggested for eukaryotic genomes (22). This hypothesis is consistent with the presence of many transcription factor and σ binding sites within genes (15, 18, 23–26), the low information sequence requirements required to promote transcription in bacteria (21), and the absence of inhibitory chromatin structure within bacterial genes (26). aRNAs are likely to have a major impact on bacterial gene expression due to the high potential for base pairing with an mRNA and the high likelihood of transcriptional interference resulting from the overlap of aRNA and mRNA transcription units. Given that aRNAs have been identified in a wide range of bacterial species, we propose that aRNAs are important regulators of gene expression in all bacteria.

NCBI short read archive accession number. Raw sequencing data are available under Accession Number SRA012168.4.

ACKNOWLEDGMENTS

We thank Steve Hanes, Marlene Belfort, Randy Morse, Todd Gray, Chris Karch, Michael Keogh, David Grainger, Zarmik Moqtaderi, and Keith Derbyshire for helpful discussions. We thank the Computational Biology and Statistics and Applied Genomic Technologies Core Facilities at the Wadsworth Center, New York State Department of Health, for expert technical assistance.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/content/1/1/e00024-10.full#SUPPLEMENTAL>.

Table S1, XLS file, 0.26 MB.

REFERENCES

- Berretta, J., and A. Morillon. 2009. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* 10:973–982.
- Waters, L. S., and G. Storz. 2009. Regulatory RNAs in bacteria. *Cell* 136:615–628.
- André, G., S. Even, H. Putzer, P. Burguière, C. Croux, A. Danchin, I. Martin-Verstraete, and O. Soutourina. 2008. S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.* 36:5955–5969.
- Brantl, S. 2007. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.* 10:102–109.
- D'Alia, D., K. Nieselt, S. Steigle, J. Müller, I. Verburg, and E. Takano. 2010. Noncoding RNA of glutamine synthetase I modulates antibiotic production in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 192:1160–1164.
- Eiamphungporn, W., and J. D. Helmann. 2009. Extracytoplasmic func-

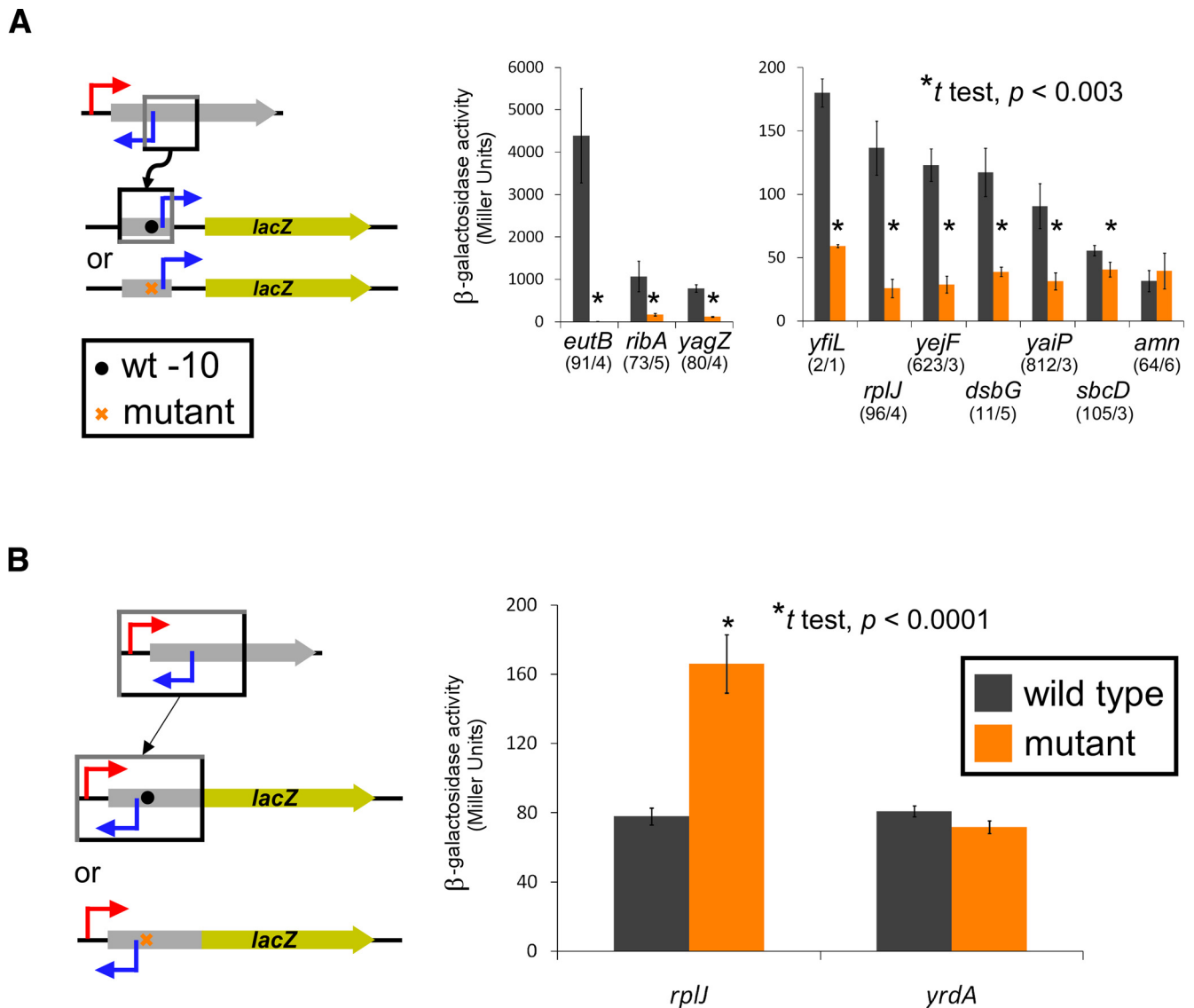


FIG 2 (A) Expression of a *lacZ* reporter gene fused to putative aRNA promoters. Wild-type (gray, right) or mutant (orange, right; -10 hexamers replaced by GGGCCC) aRNA promoter regions (200 bp upstream to 10 bp downstream of +1) were transcriptionally fused to *lacZ* on a single-copy plasmid (a derivative of pBAC-BA-*lacZ*, Addgene plasmid 13423, in which the HindIII-NotI fragment was replaced with an *E. coli* rRNA transcription terminator). β -Galactosidase assays were performed using *E. coli* MG1655 Δ *lacZ*. Gene names indicate the overlapping protein-coding genes. Numbers in parentheses indicate the number of times the aRNA 5' end was sequenced/the number of base matches to the -10 hexamer consensus. Note that one promoter tested (*eutB*) is located in an untranslated region between the *eutB* and *eutC* genes (transcribed within an operon), but the putative RNA overlaps the *eutB* gene. There is no correlation between the number of sequence reads and promoter strength. We speculate that this is due to a combination of differential aRNA stability, introduction of bias by the PCR step of library construction, and the known sequence bias of RNA ligase T4 Rnl1 (27). wt, wild type. (B) Expression of a *lacZ* reporter translationally fused to *rplJ* or *yrdA*, including the natural *rplJ* or *yrdA* protein-coding gene promoter, on a single-copy plasmid (described above). Expression levels were measured for wild-type (gray, right) and mutant (orange, right) aRNA -10 hexamers/+1 transcription start sites (mutations did not alter the protein-coding sequence of the mRNA and did not substantially alter the codon bias; the *rplJ* aRNA -10 hexamer mutated from TACAGT to GACGGT, and the +1 transcription start site mutated from A to G; the *yrdA* aRNA -10 hexamer mutated from CATAAT to CGTAGT, while the +1 transcription start site was unchanged [boldface shows change]). Expression of *rplJ::lacZ* and *yrdA::lacZ* was measured using MG1655 Δ *lacZ* and MG1655 Δ *lacZ* Δ *yrdA*, respectively.

- tion sigma factors regulate expression of the *Bacillus subtilis* *yabE* gene via a cis-acting antisense RNA. *J. Bacteriol.* 191:1101–1105.
- Fozo, E. M., M. Kawano, F. Fontaine, Y. Kaya, K. S. Mendieta, K. L. Jones, A. Ocampo, K. E. Rudd, and G. Storz. 2008. Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.* 70:1076–1093.
 - Georg, J., B. Voss, I. Scholz, J. Mitschke, A. Wilde, and W. R. Hess. 2009. Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol. Syst. Biol.* 5:305.
 - Kawano, M., L. Aravind, and G. Storz. 2007. An antisense RNA controls

synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.* 64:738–754.

- Liu, J. M., J. Livny, M. S. Lawrence, M. D. Kimball, M. K. Waldor, and A. Camilli. 2009. Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.* 37:e46.
- Stork, M., M. Di Lorenzo, T. J. Welch, and J. H. Crosa. 2007. Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. *J. Bacteriol.* 189: 3479–3488.

12. Güell, M., V. van Noort, E. Yus, W. H. Chen, J. Leigh-Bell, K. Michalodimitrakis, T. Yamada, M. Arumugam, T. Doerks, S. Kühner, M. Rode, M. Suyama, S. Schmidt, A. C. Gavin, P. Bork, and L. Serrano. 2009. Transcriptome complexity in a genome-reduced bacterium. *Science* 326:1268–1271.
13. Kawano, M., G. Storz, B. S. Rao, J. L. Rosner, and R. G. Martin. 2005. Detection of low-level promoter activity within open reading frame sequences of *Escherichia coli*. *Nucleic Acids Res.* 33:6268–6276.
14. Peters, J. M., R. A. Mooney, P. F. Kuan, J. L. Rowland, S. Keles, and R. Landick. 2009. Rho directs widespread termination of intragenic and stable RNA transcription. *Proc. Natl. Acad. Sci. U. S. A.* 106:15406–15411.
15. Reppas, N. B., J. T. Wade, G. Church, and K. Struhl. 2006. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate-limiting. *Mol. Cell* 24:747–757.
16. Selinger, D. W., K. J. Cheung, R. Mei, E. M. Johansson, C. S. Richmond, F. R. Blattner, D. J. Lockhart, and G. M. Church. 2000. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat. Biotechnol.* 18:1262–1268.
17. Sittka, A., S. Lucchini, K. Papenfort, C. M. Sharma, K. Rolle, T. T. Binnewies, J. C. Hinton, and J. Vogel. 2008. Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* 4(8). doi:10.1371/journal.pgen.1000163.
18. Wade, J. T., D. C. Roa, D. C. Grainger, D. Hurd, S. J. W. Busby, K. Struhl, and E. Nudler. 2006. Extensive functional overlap between sigma factors in *Escherichia coli*. *Nat. Struct. Mol. Biol.* 13:806–814.
19. Gama-Castro, S., V. Jiménez-Jacinto, M. Peralta-Gil, A. Santos-Zavaleta, M. I. Peñaloza-Spinola, B. Contreras-Moreira, J. Segura-Salazar, L. Muñoz-Rascado, I. Martínez-Flores, H. Salgado, C. Bonavides-Martínez, C. Abreu-Goodger, C. Rodríguez-Penagos, J. Miranda-Ríos, E. Morett, E. Merino, A. M. Huerta, L. Treviño-Quintanilla, and J. Collado-Vides. 2008. RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.* 36:D120–D124.
20. Bockhorst, J., Y. Qiu, J. Glasner, M. Liu, F. Blattner, and M. Craven. 2003. Predicting bacterial transcription units using sequence and expression data. *Bioinformatics* 19:34–43.
21. Gross, C. A., C. Chan, A. Dombroski, T. Gruber, M. Sharp, J. Tupy, and B. Young. 1998. The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb. Symp. Quant. Biol.* 63:141–155.
22. Struhl, K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* 14:103–105.
23. Grainger, D. C., D. Hurd, M. Harrison, J. Holdstock, and S. J. Busby. 2005. Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc. Natl. Acad. Sci. U. S. A.* 102:17693–17698.
24. Herring, C. D., M. Rafaele, T. E. Allen, E. I. Kanin, R. Landick, A. Z. Ansari, and B. O. Palsson. 2005. Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J. Bacteriol.* 187:6166–6174.
25. Shimada, T., A. Ishihama, S. J. Busby, and D. C. Grainger. 2008. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res.* 36:3950–3955.
26. Wade, J. T., N. B. Reppas, G. M. Church, and K. Struhl. 2005. Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites. *Genes Dev.* 19:2619–2630.
27. Romaniuk, E., L. W. McLaughlin, T. Neilson, and P. J. Romaniuk. 1982. The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. *Eur. J. Biochem.* 125:639–643.