

Identification of a Small RNA within the *pdh* Gene Cluster of *Mycoplasma pneumoniae* and *Mycoplasma genitalium*

HINRICH W. H. GÖHLMANN, JANUARY III WEINER, ASTRID SCHÖN,† AND RICHARD HERRMANN*

Zentrum für Molekulare Biologie Heidelberg, Mikrobiologie Universität Heidelberg, 69120 Heidelberg, Germany

Received 12 November 1999/Accepted 15 March 2000

A highly abundant and heterogenous small RNA about 205 to 210 bases long named MP200 RNA has been identified in *Mycoplasma pneumoniae*. It was localized on the genome within a 319-bp-long intergenic space of the pyruvate dehydrogenase (*pdh*) gene cluster. A database search at the DNA level revealed the highest similarity to a sequence located within the *pdh* gene cluster of *Mycoplasma genitalium* that was also shown to be transcribed into two abundant, but smaller RNAs than the ones in *Mycoplasma pneumoniae*. The RNAs from both *M. pneumoniae* and *M. genitalium* have the potential to code for cysteine-rich 29- and 23-amino-acid-long peptides, but so far, these peptides have not been identified experimentally in bacterial protein extracts.

Mycoplasma pneumoniae (M129; ATCC 29342), a member of the bacterial class *Mollicutes*, possesses one of the smallest genomes of all known self-replicating organisms, with a size of only 816 kb (8). Based on the complete genome sequence, 677 open reading frames (ORFs) and 39 genes coding for various untranslated RNA species were predicted. These RNAs include 35 tRNA genes (17; J. Brosius and K. Lühn, personal communication), one set of rRNAs (5S, 16S, and 23S), the 4.5S RNA, the 10Sa RNA, and the RNase P RNA (8).

A recent publication on the identification of several small RNAs in *Mycoplasma capricolum* (19) prompted us to search for similar molecules in *M. pneumoniae*. Here, we report the isolation and characterization of a hitherto unknown, 205- to 210-base-long small RNA from *M. pneumoniae*, which is also conserved in a shorter, 170- to 180-base-long version in *Mycoplasma genitalium*. The detailed description of our laboratory procedures can be found on our website (http://www.zmbh.uni-heidelberg.de/M_pneumoniae/MP200).

Identification of a highly abundant RNA and mapping on the genome. *M. pneumoniae* cells grown under standard conditions at 37°C (15) were used for the isolation of total RNA. The RNA was extracted by treating the bacteria with a solution of guanidinium thiocyanate followed by phenol-chloroform extraction and precipitation of the RNA with isopropanol. Separation of total RNA on an 8% polyacrylamide gel containing 7 M urea revealed a reproducible pattern of bands (Fig. 1) after silver staining (2). Several known species of small RNAs, like tRNAs, 4.5S RNA, and 5S rRNA, could be assigned to individual bands. In addition, two RNAs were migrating closely together, but clearly separated just above the 200-base RNA size marker. Therefore, we estimated their sizes to be 205 and 210 bases (Fig. 1). Treatment of a gel after the run with RNase in the presence of 25 mM EDTA confirmed the RNA character, since it caused the disappearance of these two bands and all of the other ones. According to their size, the two newly identified RNA species were named MP200 RNA. To map the MP200 RNA on the genome, the two RNAs were isolated from a denaturing 8% polyacrylamide gel, radioac-

tively labeled at their 5' ends by phage T4 polynucleotide kinase in the presence of [γ - 32 P]ATP, and then used as probes in Southern blot and dot blot experiments (16). The MP200 RNA could be first localized on an *Eco*RI fragment by Southern blotting against a cosmid gene bank covering the complete genome of *M. pneumoniae* (23) and then mapped more precisely within the pyruvate dehydrogenase gene (*pdh*) cluster between *pdhB* and *pdhC*. This cluster consists of the four genes *pdhA*, *pdhB*, *pdhC*, and *pdhD*, which are organized in this order and all transcribed in the same direction (8). Screening with strand-specific oligonucleotides showed that the MP200 RNA was also transcribed in the same direction as these four genes. Primer extension analysis with the oligonucleotides 5'-TTT CTT GTC TTC CAT CTT-3' and 5'-AAG ATG GAA GAC AAG AAA TGC-3' allowed us to map the 5' ends of both RNAs to the same adenine at nucleotide position 553,546 of the *M. pneumoniae* genome (8). Based on these mapping results, the boundaries of the coding region of the MP200 RNA could be predicted in the 319-bp-long intergenic region between *pdhB* and *pdhC*. Upstream of the experimentally defined 5' end of the MP200 RNA, a putative promoter -10 region (TAGAAT) is present, but a conserved -35 region similar to the corresponding regions from *Escherichia coli* and *Bacillus subtilis* has not been found, which is consistent with observations for other *M. pneumoniae* genes (6, 11, 20).

Mapping of the 3' end. Since the 5' ends of the two RNAs were identical, it seemed very likely that the size differences derived from a heterogeneity of the 3' end. Therefore, we attempted to map the 3' ends from both isolated RNAs by nuclease protection assay and two-dimensional RNA mobility shift (3). Both methods failed to define a unique 3' end; instead, we observed a terminal heterogeneity in both cases. The mobility shift and end group determination assay allowed us to define a mixture of A and U for the smaller RNA and U as the last base for the larger one (data not shown).

Since the 3' end of both species could not be unambiguously determined, it was important to define the length of both RNAs. Taking the data from a denaturing polyacrylamide gel (Fig. 1), we estimated the sizes of the two RNAs to be 205 and 210 nucleotides, based on the migration of the RNA size marker. The unique 5' end and the proposed 3' ends of both RNAs are indicated in Fig. 2. The localization of the RNAs within the *pdh* gene cluster provoked the question of whether these RNAs are stable degradation products of a larger polycistronic RNA encoded by the *pdh* cluster. This seemed to be

* Corresponding author. Mailing address: Zentrum für Molekulare Biologie Heidelberg, Mikrobiologie Universität Heidelberg, 69120 Heidelberg, Germany. Phone: (49)-6221-54 68 27. Fax: (49)-6221-54 58 93. E-mail: r.herrmann@mail.zmbh.uni-heidelberg.de.

† Present address: Institut für Biochemie, Bayerische Julius-Maximilians-Universität, Biozentrum, Würzburg, Germany.

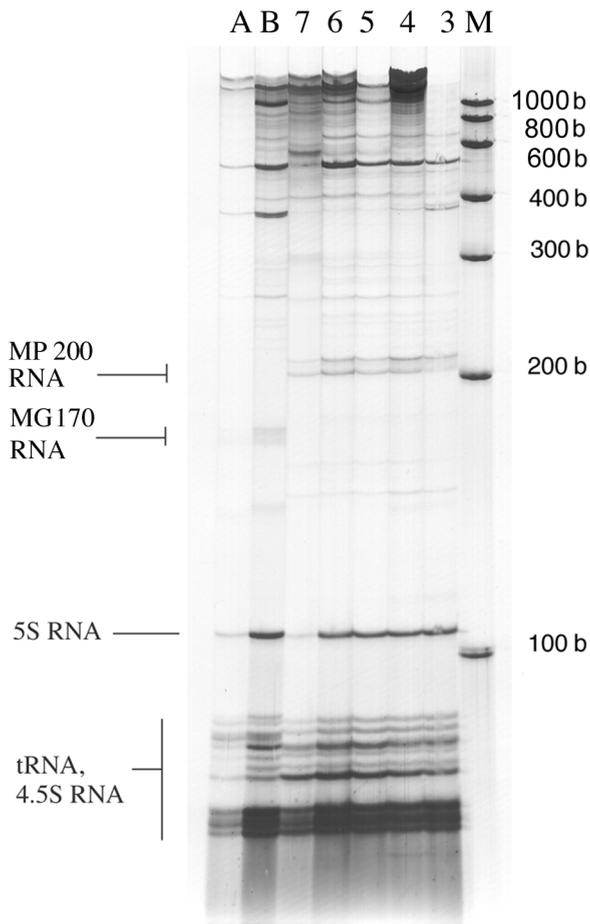


FIG. 1. Separation of different total RNA preparations on an 8% polyacrylamide gel with 7 M urea. The RNA was silver stained as described previously (2). Lanes: A and B, different *M. genitalium* preparations; 3 to 7, *M. pneumoniae* preparations from the 3rd through 7th days of culture growth, respectively; M, RNA size marker (Novagen). Sizes are shown to the right (bases [b]).

unlikely for the following reasons. (i) A promoter structure was identified (Fig. 2) upstream of the defined 5' end of the MP200 RNA, resembling other experimentally defined promoters in *M. pneumoniae* (6, 11, 20; J. Weiner and G. Browning, unpub-

lished data). (ii) Northern blot analyses with total RNA which has been separated on denaturing agarose gels did not reveal any RNA larger than the two small RNAs when we probed with MP200 RNA-specific oligonucleotides (data not shown). (iii) Primer extension analysis allowed us to determine a 5' end of the *pdhC* mRNA downstream of the proposed 3' end of the MP200 RNA with a possible promoter structure upstream of the *pdhC* mRNA 5' end.

Identification of an MP200 RNA homolog in *M. genitalium* named MG170 RNA. Database searches with the program BLASTN (1) showed no significant DNA sequence similarity to any known DNA sequence except to a sequence from the genome of *M. genitalium* (Fig. 2). This sequence was also located in the *pdh* gene cluster in the 253-bp-long intergenic region between the genes *pdhB* and *pdhC* (7). Repetition of the analyses, which has been done for the characterization of the MP200 RNA, revealed that the homologous RNA from *M. genitalium* was about 30 nucleotides shorter than, but otherwise with very similar features to, the corresponding RNA in *M. pneumoniae*. For example, the RNA could be separated by polyacrylamide gel electrophoresis into two bands with sizes of 170 and 180 nucleotides (Fig. 1); therefore, it was named MG170 RNA. The 5' ends of the RNA from both bands were identical at genome position 331214, as determined by primer extension, with the possible -10 region (TAAAT) of a promoter (Fig. 2). Three different complementary oligonucleotides, each of which shared at least 15 contiguous nucleotides, cross-reacted in Northern blotting analyses with the RNAs from both *M. genitalium* and *M. pneumoniae* and therefore stressed their sequence similarities. Comparative analyses of *pdh* clusters of other sequenced bacterial genomes, such as those of *B. subtilis* (12), *Acholeplasma laidlawii* (21), *Escherichia coli* (4), and *Mycobacterium tuberculosis* (5), showed that none had an intergenic space long enough to code for a 200-base-long RNA.

Secondary structure analysis. The RNAs from both *Mycoplasma* species contain palindromic sequences; therefore, we examined possible secondary structures by using the mfold program (version 3.1) (10). In either case, several structures have been found, with minimal free energy values (ΔG) of -77 and -55.3 kcal/mol, respectively (Fig. 3). All structures share two similar stems at the 5' end and two more in the 3' region; furthermore, a stem-loop structure is conserved between bases 80 and 100 throughout all secondary structures examined (Fig. 3). The shorter RNAs (205 and 170 bases) have similar struc-

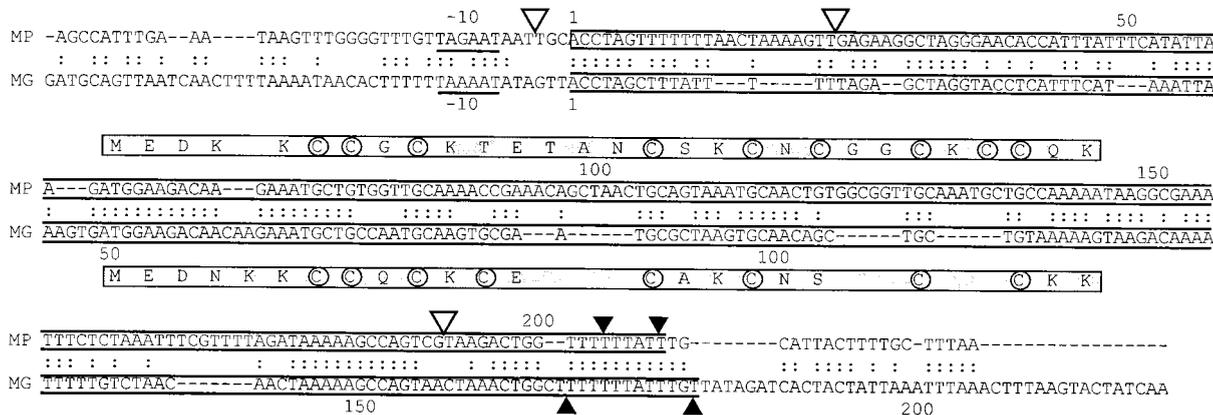


FIG. 2. Comparison of the sequences encoded by MP200 RNA in *M. pneumoniae* and MG170 RNA in *M. genitalium*. The ORFs coding for the cysteine-rich peptides are boxed. The first nucleotide (numbered 1 at the top) at the 5' end corresponds to genome positions 553546 in *M. pneumoniae* and 331214 in *M. genitalium*. Proposed 3' ends are indicated by solid triangles, and the sites of transposon insertion are indicated by open triangles.

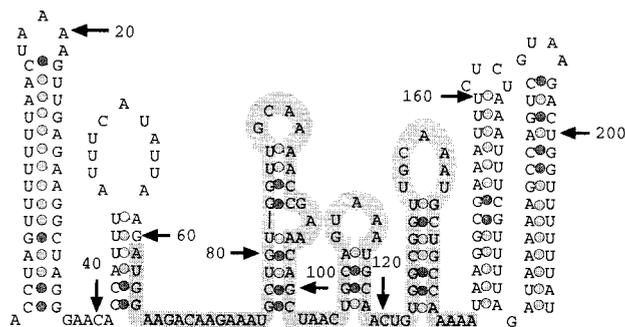
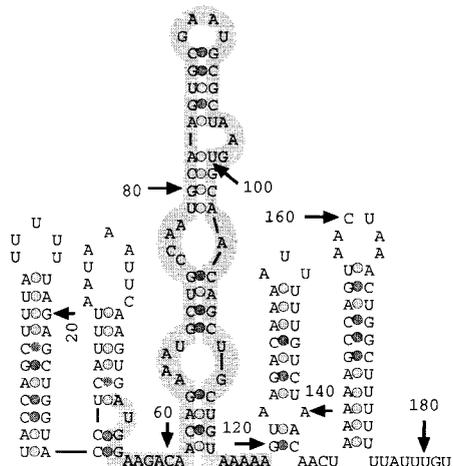
Mycoplasma pneumoniae, $\Delta G = -77$ kcal/mol*Mycoplasma genitalium*, $\Delta G = -55.3$ kcal/mol

FIG. 3. Prediction of secondary structures. The predicted structures and their respective free energy values (ΔG) were determined with the mfold program.

tures, except for slight differences in the 3' terminating stem. The U/A-rich terminating palindromic sequences are characteristic of many transcription terminator sequences (24).

Is the MP200 RNA an mRNA? Translation of the MP200 RNA into all three possible reading frames revealed that it has the potential to code for a 29-amino-acid-long peptide with 9 cysteine residues (Fig. 2). Strikingly, 8 of those 9 cysteines were conserved in a smaller 23-amino-acid-long protein, which could be translated from the homologous RNA in *M. genitalium*. Both RNAs lack a ribosomal binding site or a downstream box, as described for *E. coli* (18, 14) or *M. genitalium* (13), but this does not a priori exclude translation, since the absence of these translation signals has been also observed for other genes in *M. pneumoniae* and *M. genitalium*, which are known to be translated (8; unpublished data). So far, several attempts have failed to prove that the 29-amino-acid-long peptide is synthesized in *M. pneumoniae*. In vivo labeling of the bacterial proteins with [³⁵S]cysteine and subsequent electrophoretic analysis (Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis) did not reveal a peptide of the expected size. Similarly, an immunological approach to detect the peptide by enzyme-linked immunosorbent assay or Western blotting with monospecific antibodies produced in rabbits which were directed against the chemically synthesized peptide was unsuccessful (data not shown). The failure to detect the peptide in protein extracts of *M. pneumoniae* could also mean

that the RNA is synthesized constitutively, but translation is induced only under certain growth conditions, or that the peptide is exported into the medium.

A possible function for the MP200 RNA. There are two reasons to assume that the identified small RNAs are functional. (i) Both small RNAs identified are conserved in both *Mycoplasma* species, although they are of different lengths. (ii) The RNAs are present in high copy numbers.

Data about copy number per cell are not available, but comparison of the intensities of the stained RNA bands (Fig. 1) indicates that the number might be in the range of low-copy-number tRNA species. So far, we have no proof that the RNAs are essential. An exhaustive transposon mutagenesis of *M. pneumoniae* (9; S. N. Peterson, personal communication) did not provide conclusive evidence. Transposon insertions were mapped near the 5' end of the MP200 RNA at genome positions 553543 and 553570 and near the 3' end at genome position 553738 (Fig. 2), but none was in the proposed ORF of the cysteine-rich peptide.

The mechanisms of function can only be speculated about. The small RNAs of *E. coli* were grouped into three categories of mechanisms of action (22): RNAs that act via direct RNA-RNA basepairing, RNAs that act via RNA-protein interaction, and RNAs with intrinsic activities. Since both small RNAs from *M. pneumoniae* and *M. genitalium* have a pronounced secondary structure (Fig. 3), as predicted with the mfold program, it is possible that they act as RNAs and fall into one of these three categories of mechanisms of action. On the other hand, the sequences of the peptides are very unusual, and because of their high conservation between two *Mycoplasma* species, it seems unlikely that they are just random. Such cysteine-rich motifs—for example, MEDKKCCGCKT (see our web page; http://www.zmbh.uni-heidelberg.de/M_pneumoniae/MP200)—can be found in different metalloproteins, although those are, in general, much larger proteins which do not share any other similarities to the *Mycoplasma* peptides. They could also function as reducing agents to protect the bacterium against oxidative stress. At present, we cannot decide whether the small RNAs themselves are the functional gene products or whether they serve as mRNAs. The possibility, that they have multiple functions also cannot be excluded.

We thank Elisabeth Pirkl for excellent technical assistance, Ina Catrein and Frank Schönsiegel for doing some preliminary experiments during a laboratory course, and R. Frank for synthesis of the oligonucleotides.

This research was supported by the Graduiertenkolleg "Pathogene Mikroorganismen: Molekulare Mechanismen und Genome," a grant from the Deutsche Forschungsgemeinschaft (He780/10-1), and the Fonds der Chemischen Industrie.

H.W.H.G. and J. W. contributed equally to this work.

REFERENCES

- Altschul, S., W. Gish, W. Miller, E. Myers, and D. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Bassam, B. J., G. Caetano-Anolles, and P. M. Gresshoff. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196:80-83.
- Beier, H., and H. J. Gross. 1991. Sequence analysis of RNA, p. 221-236. In A. Brown (ed.), *Essential molecular biology*, vol. II. IRL Press, Oxford, United Kingdom.
- Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453-1474.
- Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eglmeier, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, B. G. Barrell et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis*

- from the complete genome sequence. *Nature* **396**:537–544.
6. **Dhandayuthapani, S., W. G. Rasmussen, and J. B. Baseman.** 1998. Identification of mycoplasmal promoters in *Escherichia coli* using a promoter probe vector with Green Fluorescent Protein as reporter system. *Gene* **215**:213–222.
 7. **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley et al.** 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**:397–403.
 8. **Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkel, B.-C. Li, and R. Herrmann.** 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**:4420–4449.
 9. **Hutchison, C. A., S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1999. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**:2165–2169.
 10. **Jaeger, J. A., D. H. Turner, and M. Zuker.** 1990. Predicting optimal and suboptimal secondary structure for RNA. *Methods Enzymol.* **183**:281–306.
 11. **Krause, D. C., T. Proft, C. T. Hedreyda, H. Hilbert, H. Plagens, and R. Herrmann.** 1997. Transposon mutagenesis reinforces the correlation between *Mycoplasma pneumoniae* cytoskeletal protein HMW2 and cytoadherence. *J. Bacteriol.* **179**:2668–2677.
 12. **Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Conner-ton, and A. Danchin.** 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**:249–256.
 13. **Loechel, S., J. M. Inamine, and P. Hu.** 1991. A novel translation initiation region from *Mycoplasma genitalium* that functions in *Escherichia coli*. *Nucleic Acids Res.* **19**:6905–6911.
 14. **Martin-Farmer, J., and G. R. Janssen.** 1999. A downstream CA repeat sequence increases translation from leadered and unleadered mRNA in *Escherichia coli*. *Mol. Microbiol.* **31**:1025–1038.
 15. **Proft, T., and R. Herrmann.** 1994. Identification and characterization of hitherto unknown *Mycoplasma pneumoniae* proteins. *Mol. Microbiol.* **13**:337–348.
 16. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 17. **Simoneau, P., L. Cheng-Ming, S. Loechel, R. Wenzel, R. Herrmann, and P. Hu.** 1993. Codon reading scheme in *Mycoplasma pneumoniae* revealed by the analysis of the complete set of tRNA genes. *Nucleic Acids Res.* **21**:4967–4974.
 18. **Sprengart, M. L., E. Fuchs, and A. G. Porter.** 1996. The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. *EMBO J.* **15**:665–674.
 19. **Ushida, C., and A. Muto.** 1995. Novel small stable RNAs of *Mycoplasma capricolum*. *DNA Res.* **2**:229–230.
 20. **Waldo, R. H., III, P. L. Popham, C. E. Romero-Arroyo, E. A. Mothershed, K. K. Lee, and D. C. Krause.** 1999. Transcriptional analysis of the *hmw* gene cluster of *Mycoplasma pneumoniae*. *J. Bacteriol.* **181**:4978–4985.
 21. **Wallbrandt, P., V. Tegman, B.-H. Jonsson, and Å. Wieslander.** 1992. Identification and analysis of the genes coding for the putative pyruvate dehydrogenase enzyme complex in *Acholeplasma laidlawii*. *J. Bacteriol.* **174**:1388–1396.
 22. **Wassarman, K. M., A. Zhang, and G. Storz.** 1999. Small RNAs in *Escherichia coli*. *Trends Microbiol.* **7**:37–45.
 23. **Wenzel, R., and R. Herrmann.** 1989. Cloning of the complete *Mycoplasma pneumoniae* genome. *Nucleic Acids Res.* **17**:7029–7043.
 24. **Wilson, K. S., and P. H. von Hippel.** 1995. Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc. Natl. Acad. Sci. USA* **92**:8793–8797.