

E. COLI MAP†

Two New Genes Located between 2758 and 2761 Kilobase Pairs on the *Escherichia coli* Genome

ANDRAS MICZAK,* ANIL K. CHAUHAN, AND DAVID APIRION

Department of Molecular Microbiology, Box 8230, Washington University School of Medicine, St. Louis, Missouri 63110

In the course of studying a gene, *ssrA*, that codes for a small stable RNA in *Escherichia coli* (2, 7, 13), we discovered that a 2.8-kb *AvaI-HindIII* fragment that carries the *ssrA* gene contains in addition two genes that code for small proteins of about 6 and 15 kDa; we refer to these genes as *smpA* and *smpB* for small proteins. The 2.8-kb *AvaI-HindIII* fragment hybridizes to DNA from λ phage clones 439 and 440 from the collection of Kohara et al. (4, 5). When the 2.8-kb DNA is expressed in vivo in a T7 polymerase/promoter system (14), these two proteins are synthesized.

The sequence of the first 15 amino acids of the larger protein (about 15 kDa in size) is as follows: X, Lys, Lys, Lys, Ala, X, Lys, Pro, Gly, Ser, Ala, Thr, Ile, Ala, Leu (where X indicates an unidentified residue). This was determined by automated sequencing of the purified protein. No match for this sequence was found in a protein sequence query system.

A shorter *AvaI-EcoRV* fragment (0.6 kb) codes for the smaller polypeptide. The nucleotide sequence of this fragment has been determined. It contains two overlapping open reading frames reading in the same direction; they can code for polypeptides of about 7.0 and 5.5 kDa. The polypeptide expressed, by use of a T7 polymerase/promoter system, is encoded by the smaller open reading frame as determined by sequencing of the first few amino acids of the protein. No match for the protein-coding sequence was found in a DNA data bank.

Deletion analysis of the 2.8-kb *AvaI-HindIII* fragment and expression of the deleted plasmids indicate that the order of the genes in this region of the *E. coli* chromosome is *smpA smpB ssrA*.

The 2.8-kb *AvaI-HindIII* fragment was part of a larger *HindIII-HindIII* fragment of 5.6 kb (13) that has been independently cloned by Rostas et al. (11). They reported that the *recN* gene is contained in this fragment, and they sequenced 2.2 kb of DNA containing the *recN* gene (11). The right-hand side (with respect to the *E. coli* map) of their sequence overlaps with 55 nucleotides from our sequence. The overlap region is not a part of an open reading frame. Thus, we can present here a detailed map of a 5.6-kb fragment of the *E. coli* chromosome that contains four genes, three coding for proteins and one coding for RNA (Fig. 1).

In the *E. coli* map to the left of this cluster are the genes for *pheA*, *tyrA*, and *aroF* (they are at present in a single cluster [1]).

Quite a few proteins with molecular masses around 15 kDa

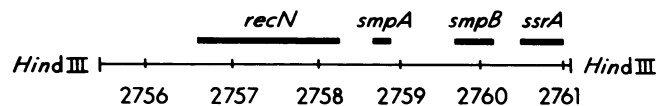


FIG. 1. Map of a 5.6-kb *HindIII* fragment of the *E. coli* chromosome. The numbers refer to the positions (in kilobase pairs) on the physical map (4). See reference 7 for a detailed restriction map of this region.

have been reported, e.g., ribosomal proteins S6A, (14.7 kDa), S6B (14.8 kDa) (10), bacterioferritin (14.3 kDa) (12), and HTP protein, encoded by *htpE* (14.7 kDa) (6). There are, however, very few proteins known with molecular masses around 7 kDa, e.g., ribosomal proteins L29, L30, and L32 (8, 9). While we do not know the functions of the proteins identified here, it is of interest that many very small *E. coli* proteins whose functions are known bind either to RNA or to DNA. For example, ribosomal proteins bind to RNA, and the small histonelike proteins of *E. coli* bind to DNA (3). Consequently, a possible function of the smaller protein reported here could be interaction with nucleic acids.

Nucleotide sequence accession number. The nucleotide sequence of the 0.6-kb *AvaI-EcoRV* fragment has been assigned EMBL/GenBank accession number X52620.

ACKNOWLEDGMENTS

We thank J. Fleischman for critical reading of the manuscript. This work was supported by Public Health Service grant GM19821 from the National Institutes of Health.

REFERENCES

1. Bachmann, B. J. 1990. Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol. Rev.* **54**:130–197.
2. Chauhan, A. K., and D. Apirion. 1989. The gene for a small stable RNA (10Sa RNA) of *Escherichia coli*. *Mol. Microbiol.* **3**:1481–1485.
3. Drlica, K., and J. Rouviere-Yaniv. 1987. Histonelike proteins of bacteria. *Microbiol. Rev.* **51**:301–319.
4. Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**:495–508.
5. Komine, Y., T. Adachi, H. Inokuchi, and H. Ozeki. 1990. Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. *J. Mol. Biol.* **212**:579–598.
6. Neidhardt, F. C., R. A. VanBogelen, and V. Vaughn. 1984. The genetics and regulation of heat shock proteins. *Annu. Rev. Genet.* **18**:295–329.
7. Oh, B.-K., A. K. Chauhan, K. Isono, and D. Apirion. 1990. Location of a gene (*ssrA*) for a small, stable RNA (10Sa RNA) in the *Escherichia coli* chromosome. *J. Bacteriol.* **172**:4708–4709.

* Corresponding author.

† For information about this section, see the January 1990 issue of *ASM News* (56:6–7).

8. **Phillips, T. A., P. L. Bloch, and F. C. Neidhardt.** 1980. Protein identifications on O'Farrell two-dimensional gels: locations of 55 additional *Escherichia coli* proteins. *J. Bacteriol.* **144**:1024–1033.
9. **Phillips, T. A., V. Vaughn, P. L. Bloch, and F. C. Neidhardt.** 1987. Gene-protein index of *Escherichia coli* K-12, edition 2, p. 919–966. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
10. **Reeh, S., and S. Pederson.** 1979. Post transcriptional modification of *Escherichia coli* ribosomal protein S6. *Mol. Gen. Genet.* **173**:183–187.
11. **Rostas, K., S. J. Morton, S. M. Picksley, and R. G. Lloyd.** 1987. Nucleotide sequence and LexA regulation of the *Escherichia coli recN* gene. *Nucleic Acids Res.* **15**:5041–5049.
12. **Smith, J. M., G. C. Ford, P. M. Harrison, J. Yariv, and A. J. Kalb.** 1989. Molecular size and symmetry of bacterioferritin of *Escherichia coli*: X-ray crystallographic characterization of four crystal forms. *J. Mol. Biol.* **205**:465–467.
13. **Subbarao, M. N., and D. Apirion.** 1989. A precursor for a small stable RNA (10Sa RNA) of *Escherichia coli*. *Mol. Gen. Genet.* **217**:499–504.
14. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci. USA* **82**:1074–1078.