Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 1111–1115, February 1982
Biochemistry

Overlap between *ampC* and *frd* operons on the *Escherichia coli* chromosome

(gene organization/DNA sequence/protein sequence/regulation of termination/structural gene–promoter overlap)

THOMAS GRUNDSTRÖM AND BENGTÅKE JAURIN

Department of Microbiology, University of Umeå, S-901 87 Umeå, Sweden

Communicated by Stanley N. Cohen, November 16, 1981

ABSTRACT **The promoter for the *Escherichia coli ampC* β-lactamase gene is shown to be located within the last gene of the fumarate reductase (*frd*) operon. Evidence is presented that the *ampC* attenuator serves as the terminator for transcription of this preceding operon. The nucleotide sequence was determined for two proteins that were shown to be encoded by a DNA segment preceding the *ampC* gene. The two proteins consisted of 131 and 119 triplets and had molecular weights of 15,000 and 13,100, respectively. The twelve COOH-terminal amino acids of the 13,100 molecular weight protein were found to overlap the *ampC* promoter. Accordingly, a G·C insertion in the promoter gave both increased transcription of *ampC* and a frameshift in this overlapping gene, resulting in readthrough proteins. Thus, we describe a type of very compact genetic organization of operons in prokaryotes.**

Sequence analysis of the DNA of many bacteriophages have revealed a very compact organization of their genetic information. For both phages fd and φX174, operons overlap in such a way that transcripts start at the terminator of the preceding operon (1, 2). A segment of phage λ including the *cro* and *cII* genes is transcribed in both directions (3). Furthermore, φX174 contains overlapping cistrons (2, 4, 5).

In the *Escherichia coli* chromosome there is also a close relationship between operons (6, 7). It has been demonstrated that a short protein from the *cheA* gene of *E. coli* is encoded by the same DNA segment as the COOH-terminus of a longer protein from the same locus (8). However, as of yet there has been no report of the overlap between promoters and structural genes, or that an attenuator for one operon can be used as the terminator for the preceding operon.

E. coli K-12 encodes chromosomally for a β-lactamase. Its structural gene, *ampC*, is located at 93.8 min, very close to *frdA*, which is the structural gene for the larger subunit of fumarate reductase (9). The gene for a M_r 27,000 protein has been localized to between *frdA* and *ampC* (10). This protein has recently been shown to be encoded by the *frdB* gene and to represent the smaller subunit of fumarate reductase (unpublished data). Herein, we present evidence that a M_r 15,000 and a M_r 13,100 protein are encoded between the *frdB* and *ampC* genes and are within the *frd* operon.

We have recently reported the sequence of the *ampC* β-lactamase gene of *Escherichia coli* K-12 (11, 12). The start points for both transcription and translation were established. It was found that the *ampC* leader sequence contained an attenuator that mediated a growth-rate-dependent regulation of the gene. Here we show that the gene preceding *ampC* overlaps the *ampC* promoter and that the *ampC* attenuator is the terminator for the *frd* operon.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

The construction of the plasmids pNU5, pNU6, pNU27, pNU28, and pNU35 has been described (10, 13). The chromosomal DNA segment carried in the different plasmids is shown in Fig. 2*A*. Plasmid pNU79 is a derivative of pNU6 with the DNA segment between the *Sac* II_1 and *Sac* II_3 sites deleted. The plasmid was constructed by O. Olsson of this laboratory.

DNA Techniques. A crude plasmid DNA preparation was made from chloramphenicol-treated cells (13) and further purified by two consecutive CsCl gradient centrifugations of 6 hr each in a Beckman VTi65 rotor. Restriction enzyme fragments were prepared from polyacrylamide gels (14) and 5′-end-labeled as described (11). The labeled DNA fragments were recleaved with a second restriction enzyme. The nucleotide sequence of the DNA was determined by the method of Maxam and Gilbert (15). The strategy for the DNA sequence determination is outlined in Fig. 2*B*. The sequence was analyzed by using previously described computer programs (12).

Protein Analysis. The introduction of plasmids into the minicell-producing strain M2141 by transformation and the purification and labeling of minicells were as described in ref. 10. The radioactive samples were analyzed by $NaDodSO_4$ gel electrophoresis, using a discontinuous buffer system (16). The running gel consisted of 25% acrylamide and 0.45% *N,N′*-methylenebisacrylamide. Standard proteins ranging in molecular weight between 3400 and 94,000 were run in parallel. After fixation, staining, and destaining (10), the gel was fluorographed by using EN^{3}HANCE (New England Nuclear). The *in vitro* protein synthesis was performed as described in ref. 11.

Edman Degradation of [^{35}S]Methionine-Labeled Proteins. Proteins synthesized in minicells containing plasmid pNU6 (10) were separated by electrophoresis as described above. The two proteins with molecular weights of approximately 14,000 were cut out of the gel and dissolved in 2 ml 0.1% $NaDodSO_4$. The gel pieces were crushed and incubated 12 hr at 4°C. The polyacrylamide was removed by filtering the solution through glass wool followed by a 15-min spin in an Eppendorf centrifuge. The proteins were dialyzed against H_2O for 24 hr and lyophilized. Edman degradation of the labeled proteins was as described by Wiman *et al.* (17). The radioactivity of each fraction was measured in a liquid scintillation counter.

RESULTS AND DISCUSSION

Localization and Polarity of Proteins Encoded Near *ampC*. Plasmids pNU5 and pNU6 both possess the chromosomally encoded *ampC* β-lactamase of *E. coli* K-12. They consist of *Eco*RI/*Pst* I-digested chromosomal DNA and the largest *Eco*RI/*Pst* I fragment of plasmid pBR322 (13). The chromosomal DNA inserts of the two plasmids are in opposite orientation relative to the vector DNA (Fig. 2*A*). The expression of proteins from these plasmids was tested in both a coupled *in vitro* trans-

cription–translation system and in minicells (Fig. 1 *A* and *B*). In addition to the M_r 39,600 β-lactamase and the M_r 27,000 smaller subunit of fumarate reductase, two other proteins with apparent M_r of about 14,000 were synthesized from both plasmids.

The location of the promoters of pBR322 and their relative strengths *in vitro* have been determined (18). Transcripts from a promoter located at the beginning of the tetracycline resistance gene of pBR322 are synthesized towards the *Eco*RI site of the plasmid. The terminator for a short transcript of the origin region of plasmid pBR322 appears to be overrun by a small fraction of the RNA (18). This leads to some readthrough transcription towards the *Pst* I site of the plasmid. Thus, *Eco*RI/*Pst* I digested DNA inserted into plasmid pBR322 is subject to a high degree of externally initiated transcription from the *Eco*RI site and, although to a lower degree, also from the *Pst* I site. A restriction map of the relevent parts of the chromosome and the chromosomal DNA inserts in the plasmids is shown in Fig. 2*A*. The figure shows that the major external transcription proceeds in the same direction as the transcription from the *ampC* promoter in plasmid pNU6 but in the opposite direction in plasmid pNU5. The two proteins with M_r of approximately 14,000 are both expressed at a much greater extent from pNU6 than from pNU5 (Fig. 1*B*). This indicates that the two proteins are translated in the same direction as *ampC* (Fig. 2*A*). Nevertheless, the amounts of β-lactamase made *in vivo* from strains carrying pNU5 and pNU6 differ by less than 25% (data not shown). This, together with the much lower expression of *ampC* relative to that of the two proteins even in pNU5, argues for the existence of a terminator of transcription between the genes for these proteins and *ampC*.

To more precisely locate the genes for the two proteins, the plasmid pNU79, which lacks the DNA between *Sac* II_1 and *Sac* II_3 sites (Fig. 2*A*), was expressed both in minicells and in a coupled *in vitro* transcription–translation system. As shown in Fig. 1 *A* and *B*, both proteins are synthesized from this plasmid. This locates the proteins between the *Sac* II_3 and the *Pst* I_4 sites (Fig. 2*A*). The exact location of the entire *ampC* gene is known from DNA sequence data (11, 12). No reading frame other than that of the *ampC* gene is open for more than 87 codons between the beginning of *ampC* and the *Pst* I_4 site (12). This maps the two proteins between the *Sac* II_3 site and the start of *ampC*.

Nucleotide Sequence of the Two Genes. The nucleotide sequence of the *frdB* gene has recently been determined (unpublished data). Here we present the nucleotide sequence between *frdB* and *ampC*. The strategy for the sequence analysis is shown in Fig. 2*B*. The sequences of both DNA strands in this stretch were determined, and the result is displayed in Fig. 3.

At two positions, 29 to 35 and 435 to 441, there are seven bases with complete base pairing homology to the seven 3′-terminal bases of 16S rRNA (19). At both positions a methionine start codon, ATG, appears six bases after the putative ribosome-binding site. After 131 and 119 codons of open frame, respectively, TAA stop codons follow. These frames would yield proteins with molecular weights of 15,000 and 13,100, respectively. The molecular weights agree well with those of the protein bands seen in the minicells and in the coupled *in vitro* transcription–translation experiments. This indicates that the proteins are indeed encoded by these DNA segments.

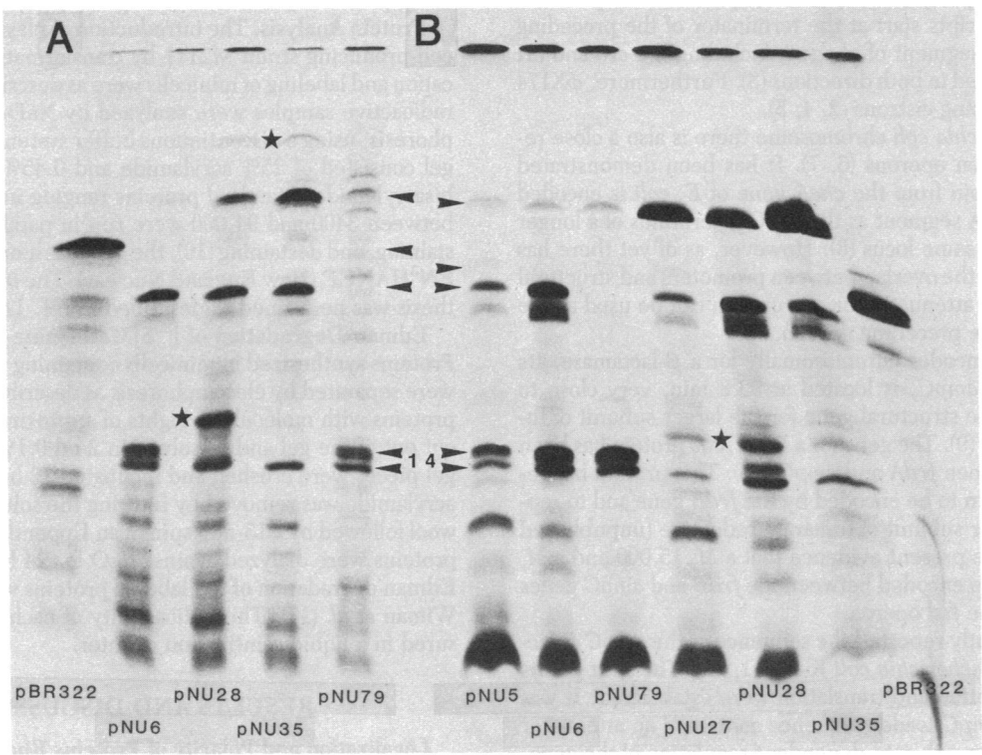

FIG. 1. Expression of proteins in a coupled *in vitro* transcription–translation system (*A*) and in minicells (*B*). The polypeptides were labeled with [^{35}S]methionine as described (10, 11). Autoradiogram of $NaDodSO_4$/polyacrylamide gel of ^{35}S-labeled protein from the indicated plasmids. The plasmids pNU27 and pNU28 are derivatives of pNU5 and pNU6, respectively, containing the *ampP15G16* mutation (11). Plasmid pNU35 is a derivative of pNU6 that in addition to the promoter mutation *ampP15G16* also contains the *ampL35A* mutation in the *ampC* attenuator (11). The arrowheads indicate the positions for, from the top of the gel, the precursor and mature *ampC* β-lactamase, precursor and mature TEM-1 β-lactamase, the M_r 27,000 *frdB* protein, and the two M_r about 14,000 proteins, respectively. The stars indicate the position of the new run-off frameshift peptide and the M_r 56,000 protein synthesized from plasmids with the frameshift mutation *ampP15G16*.

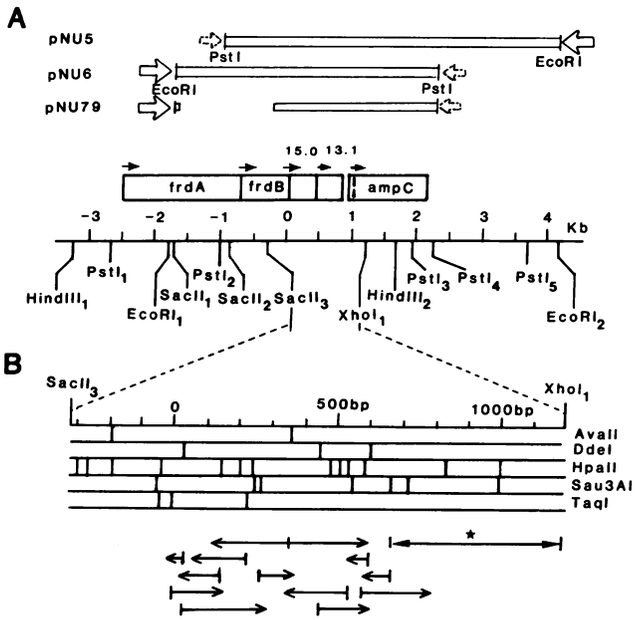

FIG. 2. (*A*) Restriction enzyme map of a segment of the *E. coli* chromosome surrounding the *ampC* gene. The chromosomal DNA inserts of the plasmids pNU5, pNU6, and pNU79 are displayed. The *Eco*RI and *Pst* I sites used in their construction are indicated. The big arrows and the small dotted arrows at the ends of the inserts represent the direction of strong and weak externally initiated transcription, respectively. The boxes show the location of *ampC* and *frdB* based on previous sequence analysis (refs. 11 and 12; unpublished data) and the location of the *frdA* protein from DNA sequence analysis by S. Cole (personal communication). The arrows over the boxes represent the direction of translation of *frdA*, *frdB*, the M_r 15,000 and 13,100 proteins, and *ampC* (refs. 11 and 25 and this paper). (*B*) Strategy for the DNA sequence analysis. Each horizontal strip represents the cleavage map for a different restriction enzyme, which is indicated at the end of the strip. Arrows represent sequencing readings from 5'-labeled ends. The segment marked with a star indicates published sequence readings of *ampC* (11, 12). bp, Base pairs.

[^{35}S]Methionine label was released from the lower of the two bands with mobilities corresponding to a molecular weight of about 14,000 after the 11th round of successive Edman degradation from the NH_2 terminus (data not shown). Assuming that the NH_2-terminal formylmethionine residue is processed away, this agrees well with the location of a methionine codon at position 12 in the 131-codon gene. No significant release of label was obtained during 20 rounds of Edman degradation of the upper of the two roughly 14,000 M_r bands (data not shown). This result is in agreement with the sequence of the 119-codon gene.

The M_r 13,100 Protein Gene Overlaps the *ampC* Promoter. A very compact organization is found in the segment whose sequence has been determined. First, the TAA stop codon of the *frdB* gene is within the ribosome-binding site of the M_r 15,000 protein gene and the stop codon of this gene is within the ribosome-binding site of the M_r 13,100 protein gene. Even more interesting is the finding that the stop codon of the M_r 13,100 protein gene is located at the fourth to sixth bases before the start base for the *ampC* mRNA (Fig. 3). At least the last 12 amino acids of the M_r 13,100 protein are encoded within the *ampC* promoter sequence.

To test whether the M_r 13,100 protein is indeed encoded by the open reading frame in the *ampC* promoter, proteins were synthesized *in vitro* from plasmid pNU28 (Fig. 1*A*) and in minicells from plasmids pNU27 and pNU28 (Fig. 1*B*). Plasmids pNU27 and pNU28 are derivatives of plasmids pNU5 and pNU6, respectively, carrying the *ampP15G16* mutation (11). This up-promoter mutation is an insertion of a G·C base pair between the 15th and 16th base pairs preceding the transcriptional start base of the *ampC* operon. The frameshift in this mutant leads to an open frame from the M_r 13,100 protein into the frame of the *ampC* protein. A protein with an apparent M_r of 56,000 which is not present in the wild type, is synthesized, although in a low amount, *in vitro* (Fig. 1*A*) by the promoter mutant. The size agrees with the 519 codons of open frame from the beginning of the M_r 13,100 protein gene to the end of the *ampC* structural gene. The synthesis of this M_r 56,000 protein is further increased off of plasmid pNU35 (Fig. 1*A*). This plasmid has, in addition to the frameshift mutation, a second mutation (*ampL35A*) leading to a less stable *ampC* attenuator stem (11).

The *ampP15G16* mutation adversely affected the synthesis of the M_r 13,100 protein both *in vitro* and in minicells. This shows that the open frame covering the *ampC* promoter encodes this protein. In the gel system used in Fig. 1, the M_r 13,100 protein migrated slightly slower than the M_r 15,000 protein. However, when the ratio of acrylamide to bisacrylamide was increased, the M_r 13,100 protein affected by the *ampP15G16* mutation migrated faster than the other protein. Both *in vitro* and in minicells from plasmids with the *ampP15G16* mutation, a new protein with a slightly slower mobility than the M_r 13,100 protein appeared. A run-off frameshift peptide from the M_r 13,100 protein gene ending at the *ampC* terminator/attenuator would have a molecular weight roughly corresponding to this mobility. The synthesis of this peptide without the appearance of the M_r 56,000 protein in minicells indicates a more efficient termination in the minicell system than in the coupled *in vitro* transcription–translation system. Moreover, plasmid pNU35, which in addition to the promoter mutation carries a second mutation (*ampL35A*) that decreases the stability of the attenuator/terminator stem (11), leads to decreased synthesis of this putative run-off frameshift peptide (Fig. 1 *A* and *B*). This shows that the synthesis of this peptide is dependent on a functional terminator stem.

Transcription from the Preceding Operon Terminates at the *ampC* Attenuator. No stretch of the sequence shown in Fig. 3, except for the *ampC* attenuator, shows the features commonly found among terminators for the *E. coli* RNA polymerase (20). In agreement with this was the finding that the higher degree of externally initiated transcription of the inserted DNA in plasmid pNU6 increased transcription of the M_r 15,000 and 13,100 proteins but not of the *ampC* β-lactamase compared to pNU5 (see above). This argues for the lack of a functional *in vivo* terminator from the *Eco*RI site in *frdA* to the end of the gene for the M_r 13,100 protein, and the existence of a terminator between the M_r 13,100 protein and the *ampC* gene. The stop codon of the gene for the 13,100 protein is located 19 nucleotides before the start of the dyad symmetry of the *ampC* attenuator. Therefore, the first terminator available for any transcript of the M_r 13,100 protein gene is the *ampC* attenuator. No other terminator-like sequence occurs prior to the end of the *ampC* structural gene. The synthesis of a peptide only slightly larger than the M_r 13,100 protein and not of a M_r 56,000 protein in minicells with the frameshift mutant supports the view that the *ampC* attenuator functions also as a terminator for transcripts initiated before the *ampC* promoter.

The absence of any difference between plasmids pNU5 and pNU6 in β-lactamase synthesis, despite the much higher expression of the M_r 15,000 and 13,100 proteins from pNU6, argues for efficient termination *in vivo* of transcription starting from promoters preceding the *ampC* promoter. We have proposed (11) that the termination of transcripts from the *ampC*

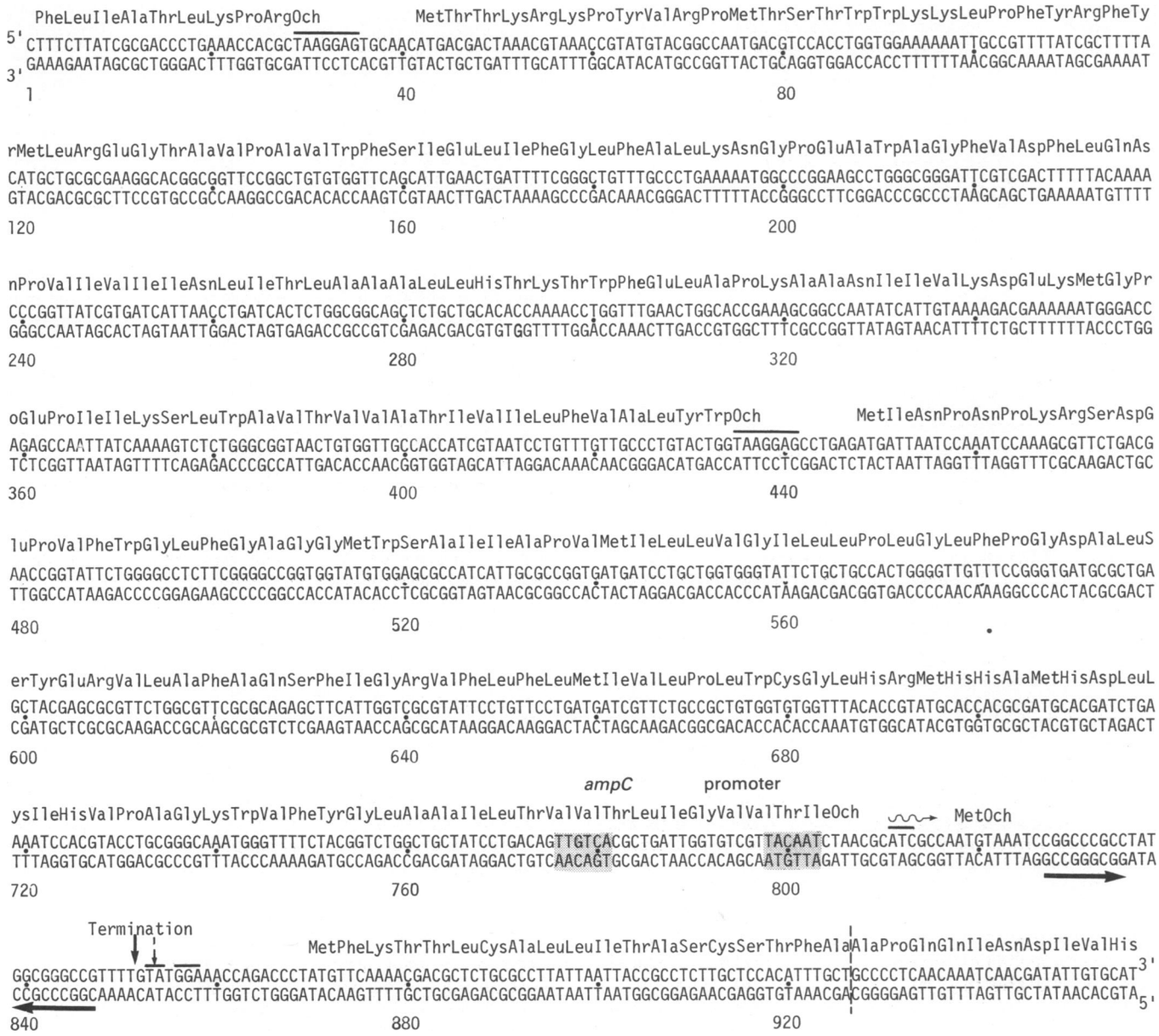

FIG. 3. DNA sequence of the genes for a M_r 15,000 and a M_r 13,100 protein from *E. coli* K-12. Every 20th base pair is marked with a dot between the strands. The count is written below. The three-letter abbreviations for amino acids appear directly over their three-base codons beginning with the methionine start codon. Och, ochre termination codon. The last nine codons of *frdB* (unpublished data) are also shown. The start of transcription from the *ampC* promoter (11) is marked by the wavy arrow. The major and minor termination points of the *ampC* attenuator (11) are indicated by vertical solid and dashed arrows, respectively. The regions of dyad symmetry in the attenuator/terminator are marked by horizontal arrows. Solid lines designate possible ribosome binding sites (19). The boundary between the signal peptide and the mature β-lactamase is marked by a vertical dashed line. The −35 and −10 regions of the β-lactamase promoter are marked by stippled boxes.

promoter is suppressed by the initiation of translation at the ATG codon at positions 818 to 820 (Fig. 3).

There are several possible explanations for a less efficient antitermination of transcripts from promoters located upstream. The proposed weak ribosome-binding site at positions 811 to 813 will not be exposed as readily as when it is the first three bases of the *ampC* β-lactamase transcript. A ribosome at the stop codon for the M_r 13,100 protein would shield the proposed ribosome-binding site of the attenuator. In addition, the earlier-started mRNAs have the possibility of internally base pairing between positions 787 and 819 and could thus form a hairpin-like stem and loop structure with a free energy (ΔG) of −10.6 kcal (−44.4 kJ)/mol (21). This stem would cover the postulated translational start sequence at positions 811 to 820. Such mRNA structures have been shown to impair initiation of translation (22, 23).

The function of the M_r 15,000 and 13,100 proteins is not known. No mutation has been mapped to this region of the *E. coli* chromosome. We have not detected termination of transcription between *frdA*, *frdB*, and these proteins. This indicates a polycistronic operon. The *frdA* and *frdB* proteins are bound to the inner membrane (24). The M_r 15,000 and 13,100 proteins contain very long stretches of hydrophobic amino acids. This indicates that these two proteins also may interact with the inner membrane.

In summary, we have found overlaps between the *ampC* gene and the preceding *frd* operon on the *E. coli* chromosome. The last gene of the *frd* operon covers the *ampC* promoter, and the

ampC attenuator is the terminator for this operon. A short genome is very favorable for all unicellular organisms in which each individual cell mainly competes for the fastest possible growth. Thus, we have to expect that overlapping genes as demonstrated herein will represent quite a common genomic organization in *E. coli* and related organisms.

We are indebted to Professor Staffan Normark for his encouragement and support of this study. Lars Rask is acknowledged for his help with the amino acid sequence determination. We thank Monica Persson for technical assistance, and Christopher Korch for carefully reading the manuscript. We acknowledge the collaboration of S. Cole and J. Weiner in collecting the cited unpublished data. This work was supported by grants from the Swedish Natural Science Research Council (Dnr 3373) and from the Swedish Medical Research Council (Dnr 5438).

1. Sugimoto, K., Sugisaki, H., Okamoto, T. & Takanami, M. (1977) *J. Mol. Biol.* **110,** 487–507.
2. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., III., Slocombe, P. M. & Smith, M. (1977) *Nature (London)* **265,** 687–695.
3. Honigman, A., Hu, S.-L., Chase, R. & Szybalski, W. (1976) *Nature (London)* **262,** 112–116.
4. Barrell, B. G., Air, G. M. & Hutchison, C. A., III (1976) *Nature (London)* **264,** 34–40.
5. Linney, E. & Hayashi, M. (1974) *Nature (London)* **249,** 345–348.
6. Post, L. E., Arfsten, A. E., Reusser, F. & Nomura, M. (1978) *Cell* **15,** 215–229.
7. Post, L. E., Strycharz, G. U., Nomura, M., Lewis, H. & Dennis, P. P. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1697–1701.
8. Smith, R. A. & Parkinson, J. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5370–5374.
9. Guest, J. & Nice, H. (1978) *J. Gen. Microbiol.* **109,** 329–333.
10. Grundström, T., Jaurin, B., Edlund, T. & Normark, S. (1980) *J. Bacteriol.* **143,** 1127–1134.
11. Jaurin, B., Grundström, T., Edlund, T. & Normark, S. (1981) *Nature (London)* **290,** 221–225.
12. Jaurin, B. & Grundström, T. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 4897–4901.
13. Edlund, T., Grundström, T. & Normark, S. (1979) *Mol. Gen. Genet.* **173,** 115–125.
14. Akusjärvi, G. & Pettersson, U. (1978) *Virology* **91,** 477–480.
15. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
16. Laemmli, U. K. & Favre, M. (1973) *J. Mol. Biol.* **80,** 575–599.
17. Wiman, K., Trägårdh, L., Rask, L. & Peterson, P. A. (1979) *Eur. J. Biochem.* **95,** 265–273.
18. Stüber, D. & Bujard, H. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 167–171.
19. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1342–1346.
20. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13,** 319–353.
21. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nat. New Biol.* **246,** 40–41.
22. Roberts, T. M., Kacich, R. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 760–764.
23. Iserentant, D. & Fiers, W. (1980) *Gene* **9,** 1–12.
24. Spencer, M. E. & Guest, J. R. (1974) *J. Bacteriol.* **117,** 954–959.
25. Cole, S. T. & Guest, J. R. (1980) *Mol. Gen. Genet.* **179,** 377–385.