

Regulatory region of the gene for the ompA protein, a major outer membrane protein of *Escherichia coli*

(Pribnow box/RNA polymerase/cyclic AMP/ribosome binding site/secondary structure)

N. RAO MOVVA, KENZO NAKAMURA, AND MASAYORI INOUE

Department of Biochemistry, State University of New York at Stony Brook, Stony Brook, New York 11794

Communicated by M. J. Osborn, April 3, 1980

ABSTRACT The ompA protein, an outer membrane protein required for conjugation, is one of the most abundant proteins in *Escherichia coli*. The structural gene for the ompA protein cloned in a plasmid vector, pMF21, conferred sensitivity to ompA protein-specific phages. We have determined the DNA sequence of a fragment of 533 base pairs encompassing the regulatory region of the ompA gene: the promoter region, the 5'-untranslated region, and the region corresponding to the signal peptide for this secretory protein. The promoter region has a sequence that is remarkably homologous with the *lac* and *gal* promoters. Particularly, both the ompA and gal promoters have the same octanucleotide sequence, T-C-A-C-A-C-T, in their RNA polymerase recognition site, which has been shown to be involved in the binding of cyclic AMP receptor protein to the gal promoter. Analogous with the observations in the gal operon, a specific RNA transcript was produced only when glycerol, a DNA-destabilizing agent, was added to a cell-free system directed by a DNA fragment of the ompA gene. These data indicate that the ompA mRNA has an untranslated region at the 5' end of about 140 nucleotides. In this region there are two additional initiation codons (II and III) besides the initiation codon (I) for the pro-ompA protein. AUG-III is located 30 bases upstream from AUG-I and accompanies a ribosome-binding site. Therefore, AUG-III is likely to begin the synthesis of a pentapeptide. The termination codon for the peptide overlaps with AUG-II, so that the ribosomes could reinitiate from AUG-II without being released from the mRNA. This reinitiation leads to the synthesis of a heptapeptide. The termination codon for this peptide also overlaps with AUG-I, which initiates the production of the pro-ompA protein. Because AUG-I also has an adjacent ribosome-binding site, the tandem repeat of initiation codons and ribosome-binding sites may be an important mechanism for facilitating the rate of initiation of translation. Extensive secondary structures exist in the 5' end as well as in the coding region of the ompA mRNA, which may also play a role in the function of the mRNA.

The *Escherichia coli* outer membrane contains a group of "major" proteins that are abundant in the cells (for reviews, see refs. 1 and 2). Among them are the lipoprotein, about 7×10^5 molecules per cell, and the matrix proteins and the ompA (tolG) protein, about $2-3 \times 10^5$ molecules per cell. Analysis of the genetic organization of these proteins leads to an understanding of regulatory features responsible for their efficient expression. Determination of the complete nucleotide sequence of lipoprotein mRNA (3) and of its structural gene (4) has revealed many unique features concerning its expression.

In this paper, we have investigated the mechanism of ompA gene expression. The ompA gene was cloned in the plasmid cloning vehicle, pMF21, and the nucleotide sequence of a DNA fragment of 533 base pairs encompassing the regulatory region

of the ompA gene was determined. This sequence revealed several unique features, including two additional translation initiation codons in the 5' end of the mRNA, which may lead to the synthesis of two oligopeptides. In addition, a series of base pairs corresponding to a binding site for cyclic AMP (cAMP) receptor protein were found.

MATERIALS AND METHODS

Materials. [γ - 32 P]ATP (9,000 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels), used for 5'-end labeling of DNA fragments, was prepared by the method of Johnson and Walseth (5) from carrier-free [32 P]orthophosphate (New England Nuclear) and ADP (Sigma). Restriction enzymes *Tac* I and *Alu* I were obtained from Bethesda Research Laboratories (Rockville, MD); other enzymes were from New England BioLabs. Bacterial strains were constructed and DNA was prepared as described (6).

DNA Sequence Determination. DNA sequence was determined by the method of Maxam and Gilbert (7). The cleaved products were separated by using the thin sequencing gel system ($0.04 \times 20 \times 40$ cm) (8) with 20%, 10%, or 8% polyacrylamide in 7 M urea.

In Vitro Transcription. Transcription was carried out by the method of Fuller and Pratt (9) with some modification: each reaction was performed in a final volume of 10 μ l which contained 30 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 4 mM Mg(CH₃COO)₂, 0.1 mM dithiothreitol, 100 mM KCl, 200 μ M ATP, 200 μ M GTP, 200 μ M CTP, 10 μ M [α - 32 P]UTP (40 Ci/mmol; Amersham), 5% (vol/vol) glycerol, and 0.04 pmol of DNA template. When we examined for glycerol stimulation, glycerol was added to a final concentration of 25%. The reaction was started by adding 0.67 μ g of *E. coli* RNA polymerase (New England BioLabs) and the reaction mixture was incubated at 37°C for 30 min. The reaction was terminated by adding 200 μ l of Tris/EDTA (10 mM Tris-HCl, pH 7.5/1 mM EDTA) containing 15 μ g of yeast tRNA (P-L Biochemicals). The final mixture was extracted with phenol and washed with ether. The RNA transcripts were precipitated by ethanol and dissolved in 30 μ l of solubilizing solution containing 6 M urea, 0.1 mM EDTA, and 0.05% xylene cyanol and bromphenol blue. They were then analyzed by polyacrylamide gel electrophoresis on 5% TBE/urea gels as described by Maniatis *et al.* (10).

RESULTS AND DISCUSSION

Restriction Enzyme Mapping of Regulatory Region of ompA Gene. The ompA gene was cloned from F' 106 DNA by using pMF21 as a cloning vehicle as described (6). The plasmid, pRM111, thus isolated was able to confer to bacteria sensitivity

Abbreviation: cAMP, cyclic AMP.

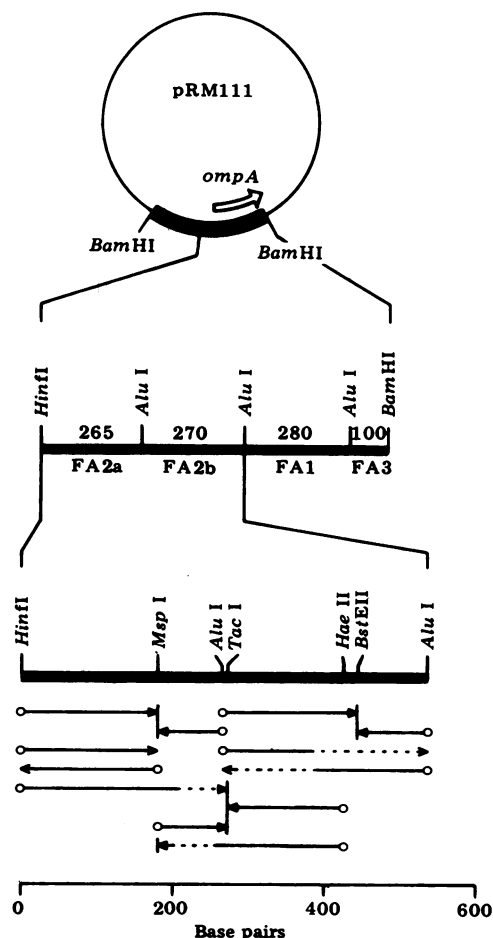


FIG. 1. Restriction enzyme cleavage sites around the *ompA* gene. Shown are: pRM111, which carries a 1.7-kilobase *Bam*HI fragment in pMF21 (6); the map of a 950-base-pair *Hin*fI/*Bam*HI fragment; and the map of FA2a-FA2b and sequencing strategy of this portion. Open circles (O) indicate the position of 32 P label at the 5' end. The singly labeled fragments were obtained either by secondary restriction enzyme cleavage at the sites indicated by the vertical line (\rightarrow) or by strand separation indicated by open arrows (\rightarrow). Broken regions of the arrows indicate that the sequences of these regions were not determined.

to phages K3 and TuII*, but lacked a section of the *ompA* gene that covers a carboxyl-terminal portion of the *ompA* protein (6). The clone carries a 1.7-kilobase-pair *Bam*HI fragment; its restriction enzyme cleavage map is shown in Fig. 1.

DNA Sequence of Regulatory Region of *ompA* Gene. In our preliminary study (6), we showed that the left half of fragment FA2b (see Fig. 1) codes for the amino-terminal portion of the *ompA* protein and the direction of the transcription of the *ompA* gene is from left to right. Therefore, the regulatory region of the *ompA* gene should be in the FA2a fragment. The DNA sequence of the region covered by FA2a and FA2b (total of 533 base pairs) was thus determined. Fig. 1 also shows restriction enzyme fragments and sequence strategies used in the present paper.

Fig. 2 shows the DNA sequence thus determined and the amino acid sequence of the amino-terminal portion of the pro-*ompA* protein (a secretory precursor of the *ompA* protein) deduced from the DNA sequence. The amino acid sequence underlined in Fig. 2 agrees with the amino-terminal sequence of the *ompA* protein previously determined (11). The amino acid sequence also shows the existence of a signal peptide consisting of 21 amino acid residues, as predicted (12). Its unique features have been discussed (6).

Promoter and Transcription Initiation. In the region about 140 base pairs upstream from the initiation codon, there are remarkable homologies with the *lac* promoter as well as the *gal* promoters. As shown in Fig. 3, T¹⁰¹-A-c-G-T-T-G¹⁰⁷ of the *ompA* gene is almost identical with the "Pribnow box" sequence of the *lac* gene, T-A-t-G-T-T-G (13). (The nucleotides in lowercase letters are the only difference between the two sequences.) Twelve nucleotides upstream from the Pribnow box of the *ompA* gene, a "RNA polymerase recognition site" generalized as T-G-T-T-G-A-C-A-A-T-T-T (13, 15, 16), can be found in a⁷⁷-G-T-T-c-A-C-A-c-T-T-g⁸⁸. On the other hand, the RNA polymerase recognition sites for the *lac* and *gal* P₁ and P₂ promoters are located 14, 17, and 12 nucleotides upstream from their Pribnow boxes, respectively (Fig. 2). Their sequences, G-C-T-T-T-A-C-A-C-T-T-T for the *lac* promoter (13) and A-T-G-T-C-A-C-A-C-T-T-T for the *gal* P₁ and P₂ promoters (14) are surprisingly homologous to that of the *ompA* gene (Fig. 3). The underlined nucleotides show homologies with the *ompA* gene. The distributions of A-T base pairs upstream from the RNA polymerase recognition sites are also similar in these genes; a short section of 9, 10, and 2 nucleotides (for *ompA*, *lac*, and *gal*, respectively) immediately upstream from the RNA polymerase recognition site has a very low A-T content (22%, 20%, and 0%, respectively). This section is then followed by a long section of 43, 27, and 34 nucleotides with a very high A-T content (81%, 70%, and 76%, respectively). Among the homologies described above, the most striking one is the octanucleotide sequence, $\begin{matrix} \text{TCACACTT} \\ \text{AGTGTGAA} \end{matrix}$, found in the recognition sites of the *ompA* and *gal* genes. This sequence is suggested to be involved in the binding of cAMP receptor protein to the *gal* promoter (14). The underlined sequence seems to be most important for recognition (14). This binding stimulates the *gal* P₁ promoter but inhibits the *gal* P₂ promoter (14). The latter promoter, however, is stimulated by DNA helix-destabilizing agents such as glycerol *in vitro* in the absence of cAMP and cAMP receptor protein (14, 17).

We found that glycerol has a dramatic effect on the production of a specific RNA transcript in a cell-free system directed by DNA fragments derived from the *ompA* gene. As shown in Fig. 4, only when glycerol was added to a cell-free system with *Hin*fI/*Hae* II DNA fragment (positions 1-425) as template was a specific RNA transcript of about 310 nucleotides long (band b) produced. The size of this RNA agrees with that of the RNA (309 nucleotides) if its synthesis is assumed to be initiated at a site five nucleotides downstream from the Pribnow box (position 113) described above. The RNA transcript produced both in the absence and presence of glycerol (band a) is about 450 nucleotides long, which possibly corresponds to the size of the entire DNA fragment. This RNA is most likely produced as a result of nonspecific transcription from either end of the DNA fragment.

In order to confirm the origin of the transcript described above, we used another DNA fragment, which covers only the first half of the previous DNA fragment (*Hin*fI/*Msp* I fragment, positions 1-179). One can predict that this DNA fragment directs the production of the RNA transcript consisting of 60-70 nucleotides if the transcription starts at around position 113 and moves forwards to the right. As shown in Fig. 4 (lanes 3 and 4), this was found to be the case; band c is about 60 nucleotides long and its production was also markedly enhanced in the presence of glycerol. These results strongly indicate that the *ompA* promoter is located as shown in Fig. 3 and that the transcription initiation site is at around position 113. Our results also suggest that the *ompA* protein is influenced by cAMP. Effects of cAMP on the biosynthesis of *E. coli* outer membrane proteins have been investigated (18, 19), but no significant effects of cAMP on the production of the *ompA* protein have been observed.

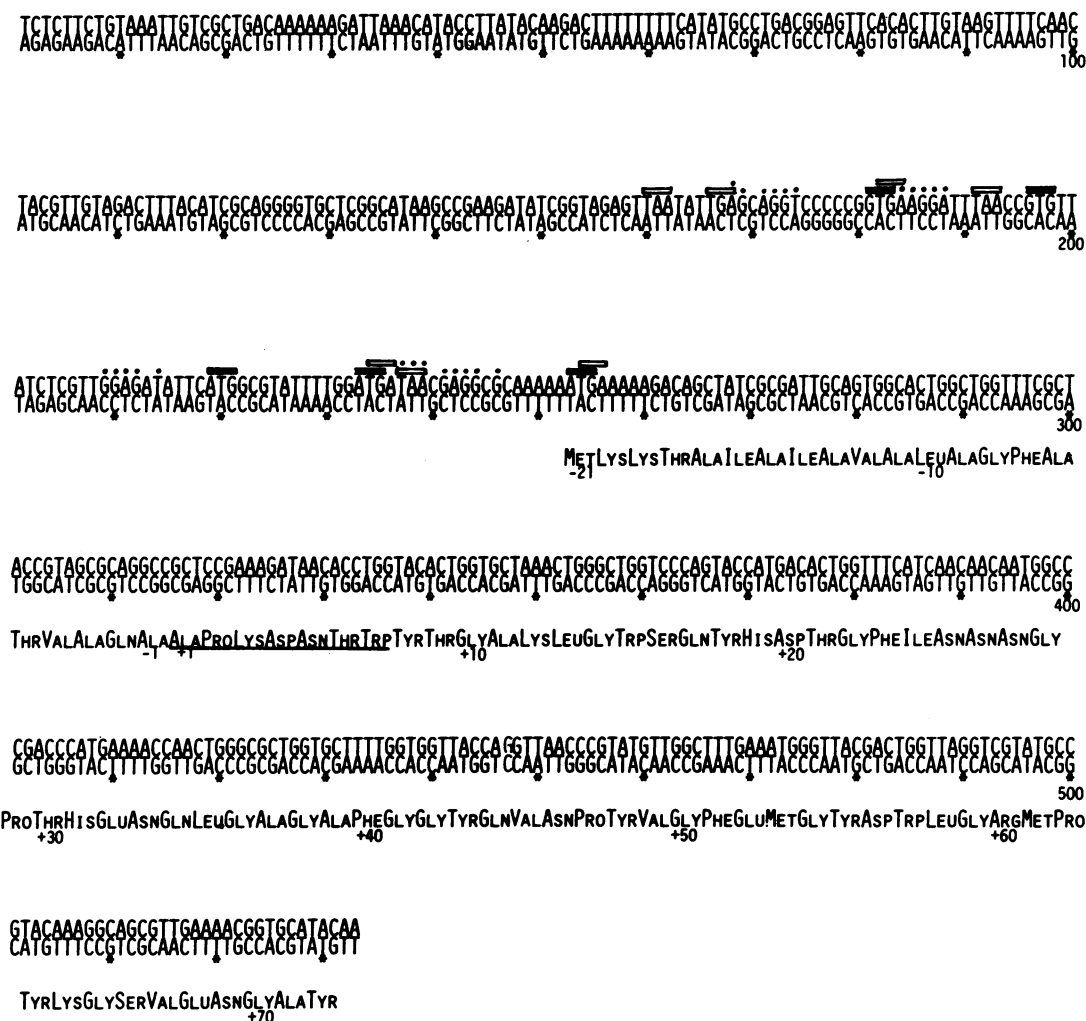


FIG. 2. DNA sequence of the promoter region of the *ompA* gene. The amino acid sequence deduced from the DNA sequence is also shown. The underlined amino acid sequence is identical with the amino-terminal structure of the *ompA* protein determined by Endermann *et al.* (11). In the section containing nucleotide residues 101–300, potential initiation codons and termination codons are indicated by solid and open bars, respectively. Shine–Dalgarno sequences in this section are shown by dots over base pairs.

Besides the Pribnow box (I) described above, there are four other possible Pribnow boxes between the Pribnow box I and the initiation codon, ²⁵³ATC²⁵⁵: (II) ¹⁶⁰T-A-a-t-A-T-t⁶⁶, (III) ¹⁹⁷T-g-T-t-A-T-c²⁰³, (IV) ²²⁵T-A-T-t-t-T-C²³¹, and (V) ²³²g-A-T-G-A-T-A²³⁸, where nucleotides in lowercase letters indicate differences from the generalized sequence, T-A-T-Pu-A-T-Pu (13, 15). There are possible RNA polymerase binding sites for all these Pribnow boxes (not listed), and it is possible that some of them may be functional as promoters *in vivo*.

	RNA polymerase recognition site	Pribnow box
<i>ompA</i> :	AGTTCACACTTG - TAAGTTTTCAAC	- TACGTTG - TAGACTTTACA
<i>lac</i> :	GCIITACACIIT - ATGCTCCGGCTCG	- IATGIIIG - TGTGGA
<i>gal P₁</i> :	ATGTCACACIIT - TCGCATCTTTGTTATGC-	IATGGIT - ATTTCA
<i>gal P₂</i> :	ΔTGTACACIIT - TCGCATCTTTGT	- IATGCTA - TGGTTA

FIG. 3. Comparison of a DNA sequence of the *ompA* gene from positions 77 to 118 with those of the *lac* (13) and *gal* (14) promoters. Underlined nucleotides show homologies with the *ompA* gene. Arrows indicate transcription initiation sites. The sequence double-underlined in the *gal* promoter is required for binding of cAMP receptor protein (14).

Ribosome Binding Site. From the results described above, *ompA* mRNA appears to have a leader sequence (so-called “untranslated region”) of about 140 nucleotides at its 5’ end. Six nucleotides upstream of the initiation codon for the pro-*ompA* protein, ²⁵³ATG²⁵⁵, there is a Shine–Dalgarno sequence, ²³⁷T-A-A-c-G-A-G-G-c-G²⁴⁶, which has a structure complementary to the 3’ end of 16S ribosomal RNA (20). This structure, as well as the existence of a termination codon TAA or TGA (but not TAG) in this region, appears to be very important for the recognition for ribosomal initiation (21). In the *ompA* gene, TAA is found in the ribosome binding site, and there is no TAG codon between this TAA codon and the initiation codon (see Fig. 2).

There are two more initiation codons (II and III) upstream from the initiation codon (I) for the pro-*ompA* protein (Fig. 2); I for ²⁵³ATG²⁵⁵, II for ²³³ATG²³⁵, and III for ²¹⁹ATG²²¹. Initiation codon III also accompanies a Shine–Dalgarno sequence ²⁰⁹G-G-A-G-a-T²¹⁴. The closest termination codon upstream from ATG-III is again TAA but not TAG, which satisfies the Atkins rule (21). Therefore, ATG-III is considered to be a functional initiation codon. As a result, a pentapeptide, Met-Ala-Ile-Phe-Gly, can possibly be synthesized. Because the termination codon, TGA, for this peptide overlaps with ATG-II (note that there is another termination codon, TAA, after TGA),

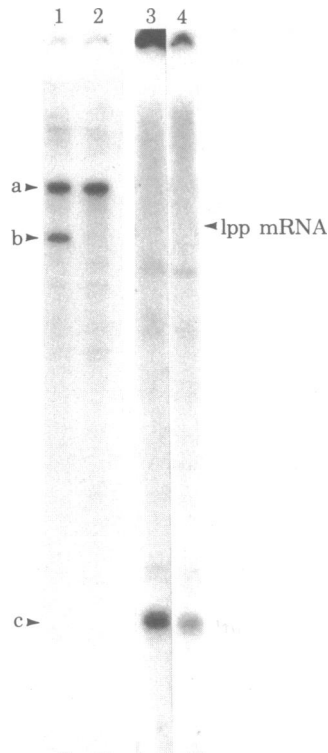


FIG. 4. Analysis of RNA transcripts produced in a cell-free system directed by the *HinfI/Hae II* fragment (positions 1–425) and *HinfI/Msp I* fragment (positions 1–179; see Fig. 1) of the *ompA* gene. The position of the lipoprotein mRNA (322 nucleotides; ref. 3) is shown. Lanes 1 and 2, with the *HinfI/Hae II* fragment in the presence and the absence of glycerol, respectively. Lanes 3 and 4, with the *HinfI/Msp I* fragment in the presence and the absence of glycerol, respectively. RNA products were analyzed by a 5% polyacrylamide gel.

it is likely that the ribosome that completes the synthesis of the pentapeptide immediately restarts from the initiation codon ATG-II by shifting one nucleotide to the left without being released from the mRNA. Therefore, ATG-II would start the production of a heptapeptide, Met-Ile-Thr-Arg-Arg-Lys-Lys. The termination codon, TGA, of this heptapeptide again

overlaps with the initiation codon, ATG-I, in exactly the same fashion as does ATG-II. Therefore, the ribosome can reinitiate from ATG-I without being released from the mRNA. Such reinitiation events from an initiation codon located close to a termination codon have been well documented (see reviews in refs. 22 and 23 for overlapping of AUG with a termination codon).

Tandem repeats of initiation codons coupled with ribosome binding sites in the *ompA* mRNA may play an important role in the efficient translation of the mRNA by increasing the rate of initiation. It is also possible that small peptides are produced that may have some functions in cellular activities. The overlapping of an initiation codon with a termination codon, TGA, found at ATG-I and -II has been observed in the lipoprotein gene (3) as well as in many other prokaryotic genes (28 out of 63 listed in table 1 in ref. 21). The existence of an initiation codon coupled with a termination codon in the untranslated region of the 5' end has also been reported in the mRNA from the lipoprotein (24), where a hexapeptide, Met-Glu-Ile-Asn-Ser-Ile, would be synthesized. However, in this case, there is no ribosome binding site at an appropriate distance from the initiation codon.

There are two more possible initiation codons, $^{196}\text{GTG}^{198}\text{-I}$, and $^{181}\text{GTG}^{183}\text{-II}$. Both accompany ribosome binding sites, $^{184}\text{A-A-G-G-A}^{188}$ for GTG-I and $^{168}\text{A-G-c-A-G-G-T}^{174}$ for GTG-II, and do not violate the Atkins rule (21). Both initiation codons are in the same reading phase as ATG-I. Therefore, if these codons are actually used for translation, the signal peptide of the *ompA* protein would have an extra peptide extension at its amino terminus consisting of 19 or 24 amino acid residues for GTG-I or -II, respectively.

Secondary Structure. At the 5' end of the 140-base leader sequence of *ompA* mRNA, a very stable secondary structure (I) can be constructed as shown in Fig. 5. The ΔG value for this structure is calculated to be -17.3 kcal according to Tinoco and

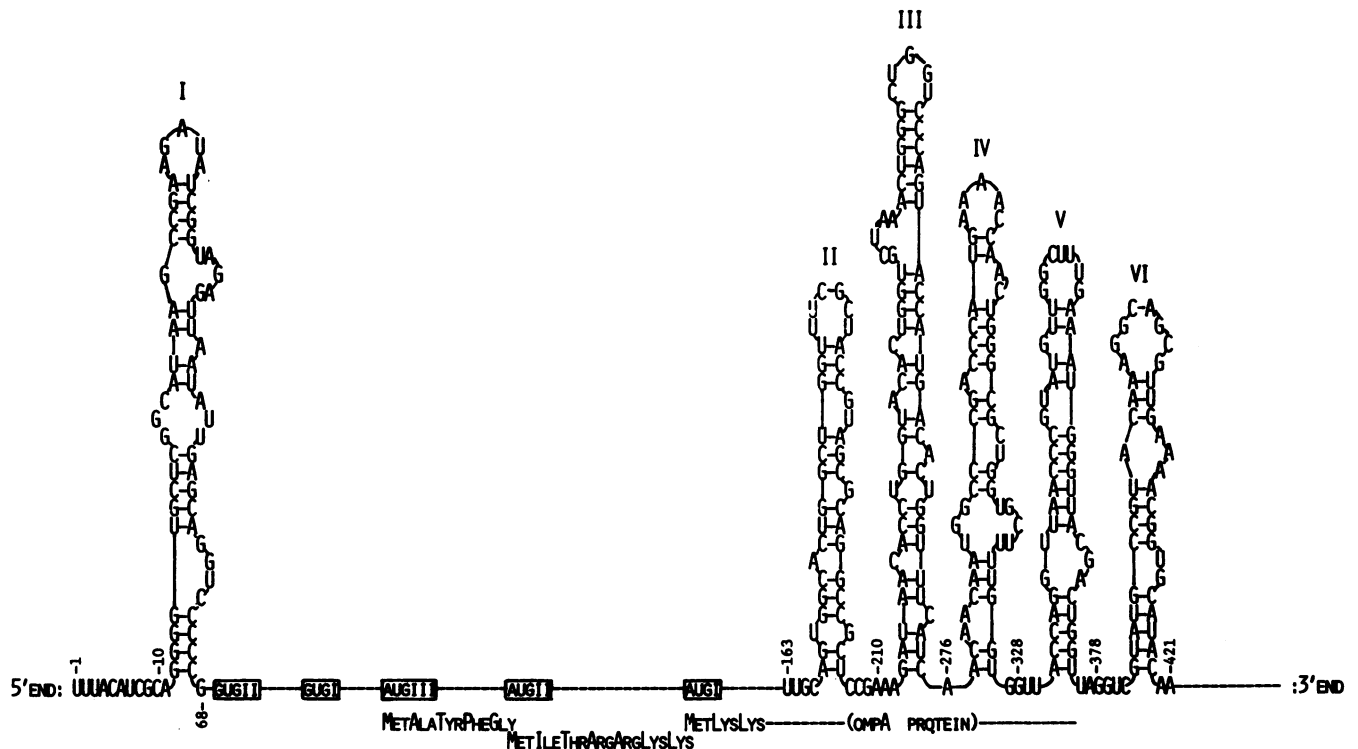


FIG. 5. Possible secondary structures in the *ompA* mRNA. The *ompA* mRNA is assumed to start at position 113 in Fig. 2, as discussed in the text. Five possible initiation codons are shown between structures I and II. Two oligopeptides, which can be initiated from AUG-III and AUG-II, respectively, are also illustrated.

coworkers (25, 26). Furthermore, in the coding region of the mRNA there seem to be extensive secondary structures as shown in Fig. 5. The ΔG values for these structures are -22.4 , -24.7 , -19.5 , -16.7 , and -11.9 kcal for structures II, III, IV, V, and VI, respectively. These secondary structures may have some relevance to the stability of the *ompA* mRNA (27) as well as to the regulatory mechanism of *ompA* gene expression. For several operons such as *trp*, *phe*, *his*, *thr*, and *leu*, secondary structures in the 5'-end leader sequences appear to play an important role in the control of their gene expression (28, 29).

There is no secondary structure in the 99-base sequence between structures I and II, where all potential initiation codons and ribosome binding sites are located (Fig. 5). This may facilitate ribosome binding to the *ompA* mRNA. Similar features are also found in the lipoprotein mRNA, which has no secondary structure in the first 64 bases from the 5' end even though it has extensive secondary structures after position 65 (3).

Note Added in Proof. We have recently completed the entire nucleotide sequence of the structural gene for the *ompA* protein and found that: (i) pro-*ompA* protein consists of 346 amino acid residues; (ii) *ompA* mRNA has 46 bases at the 3'-end untranslated region and has extensive secondary structure; and (iii) 96 amino acid residues at the COOH-terminal part of the *ompA* protein are not required for specific phage recognition.

We thank Dr. M. Freundlich for helpful discussions. This work is supported by grants from the National Institutes of Health (GM19043-08) and the American Cancer Society (BC-67G).

- Inouye, M. (1979) in *Bacterial Outer Membranes; Biogenesis and Functions*, ed. Inouye, M. (Wiley, New York), pp. 1-12.
- DiRienzo, J. M., Nakamura, K. & Inouye, M. (1978) *Annu. Rev. Biochem.* **47**, 481-532.
- Nakamura, K., Pirtle, R. M., Pirtle, I. L., Takeishi, K. & Inouye, M. (1980) *J. Biol. Chem.* **255**, 210-216.
- Nakamura, K. & Inouye, M. (1979) *Cell* **18**, 1109-1117.
- Johnson, R. A. & Walseth, T. F. (1979) *Adv. Cyclic Nucleotide Res.* **10**, 135-167.
- Movva, N. R., Nakamura, K. & Inouye, M. (1980) *J. Biol. Chem.* **255**, 27-29.
- Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560-564.
- Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107-110.
- Fuller, R. S. & Pratt, T. (1978) *Nucleic Acids Res.* **5**, 4613-4623.
- Maniatis, T., Jeffrey, A. & Van de Sande, H. (1975) *Biochemistry* **14**, 3787-3794.
- Endermann, R., Hindennach, I. & Henning, U. (1978) *FEBS Lett.* **88**, 71-74.
- Sekizawa, J., Inouye, S., Halegoua, S. & Inouye, M. (1977) *Biochem. Biophys. Res. Commun.* **77**, 1126-1133.
- Reznikoff, W. S. & Abelson, J. N. (1978) in *The Operon*, eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 221-243.
- Taniguchi, T., O'Neill, M. & de Crombrughe, B. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5090-5094.
- Gilbert, W. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 193-205.
- Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319-353.
- Nakanishi, S., Adhya, S., Gottesman, M. & Pastan, I. (1974) *J. Biol. Chem.* **249**, 4050-4056.
- Aono, R., Yamasaki, M. & Tamura, G. (1978) *J. Bacteriol.* **136**, 812-814.
- Mallick, U. & Herrlich, P. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5520-5523.
- Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **74**, 4734-4738.
- Atkins, J. F. (1979) *Nucleic Acids Res.* **7**, 1035-1041.
- Miller, J. H. (1974) *Cell* **1**, 73-76.
- Platt, T. & Yanofsky, C. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2399-2403.
- Pirtle, R. M., Pirtle, I. L. & Inouye, M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2190-2194.
- Tinoco, I., Borer, P., Dengler, B., Levine, M., Uhlenbeck, O., Grothers, D. & Gralla, J. (1973) *Nature (London) New Biol.* **246**, 40-41.
- Borer, P. N., Dengler, B., Tinoco, I. & Uhlenbeck, O. C. (1974) *J. Mol. Biol.* **86**, 843-853.
- Hirashima, A., Childs, G. & Inouye, M. (1973) *J. Mol. Biol.* **79**, 373-389.
- Lee, F. & Yanofsky, C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4365-4369.
- Keller, E. B. & Calvo, J. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6186-6190.