Nucleotide sequence of the gene ompA coding the outer membrane protein II* of Escherichia coli K-12

E.Beck

Mikrobiologie, Universität Heidelberg, Im Neuenheimer Feld 230, D-6900 Heidelberg, and

E.Bremer,

Max-Planck-Institut für Biologie, Correnssstrasse 38, D-7400 Tübingen, GFR

ABSTRACT

A nucleotide sequence of 2271 basepairs has been deter-
mined from cloned E. coli DNA which contains ompA. Within
that sequence, starting at nucleotide 1037, an open trans-
lational reading frame encodes a protein of 367 amino acids
which starting with amino acid 22 agrees with the primary
structure of protein II*. The preceeding 21 amino acids
constitute a typical signal sequence. There is a non-trans-
lated region of 360 nucleotides in front of the translational
start. The insertion point of an IS1 element 110 nucleotides
upstream from the start codon and an amber codon at the
position of amino acid residue 28 have been localized in
the DNA from two ompA mutants.

INTRODUCTION

Protein II* (1, ompA protein, for other such proteins
and other nomenclature see ref. 2) is one of the few
abundant proteins of the outer membrane of the E. coli
cell envelope, and it is present at a concentration of
about 10^5 copies per cell (3). The transmembrane protein (4)
serves non-physiological and physiological functions: it
can act as a phage receptor (5, 6), it is required for the
action of a colicin (7), it plays a role in F-mediated
conjugation (6, 8, 9), it is required in combination with
the outer membrane lipoprotein (10) for the structural
integrity of this membrane (11), and it is responsible for
certain surface properties of the cell (12). Comparable to
the lipoprotein protein II* is produced in large quantities
during growth in a variety of media (unpublished data and
ref. 13), and the gene is likely to be very efficently ex-
pressed.

The primary structure of the protein consisting of 325
amino acid residues, has recently been determined (14). The
aim of this work was to use DNA sequencing analysis of the
ompA gene and of adjacent DNA segments to confirm this
analysis on the nucleotide level and to extend it to those
parts of the genome not coding for the mature protein II[*],
i.e. the sequences coding for the putative signal peptide
used for membrane integration and sequences
controlling the expression of the gene. Knowledge of the
nucleotide sequence should also provide a basis for sequence
specific manipulation of the gene, and of its control elements.

MATERIAL AND METHODS

The plasmids relevant to the present study are listed in
Table 1. The construction of these plasmids is described in
reference (15). All plasmids were maintained in strain UH100

Table 1: Plasmids used in the sequence analysis

Plasmid	Derivation and composition (ref.)
pTU100	wild type *ompA* gene on a 7.5 kb fragment cloned in pSC101 (21)
pTU102	*ompA* mutant (allele 31) in pTU100 (15)
pTU104	insertion 72 in pTU100 (15)
pTU202	EcoRI fragment with insertion 72 from pTU104 cloned in pBR325 (Henning, unpublished)
pTU301	1.83 kb [a)] BamHI fragment ($\approx$ Bam-C, encoding the CO_2H-terminal part of protein II[*]) from pTU102, cloned in pBR322 (15)
pTU302	1.78 kb [a)] BamHI fragment ($\approx$ Bam-N, encoding the NH_2-terminal part of protein II[*]) from pTU102, cloned in pBR322 (15)

[a)] The size of these fragments had previously been determined
by agarose gel electrophoresis only ( ), and was given as
1.76 and 1.67 kb instead of 1.83 and 1.78 kb, respectively.

(a *recA*, *ompA* derivative of strain W620; ref. 15), and where
hybrid plasmids derived from pBR322 or pBR325 were present
cells were grown at 37° C in L-broth containing ampicillin
(20 µg/ml). Plasmid DNA was prepared from cleared lysates (16)
by RNase treatment, phenol extraction and chromatography on
Biogel A 0.5m.

The DNA sequencing procedures are all from Maxam and
Gilbert (17) with the modification of the purine specific
cleavage reaction described by Gray et al. (18).

The restriction endonucleases BamHI, Sau3A, HpaII, HinfI,
HhaI, HaeIII, and AluI were prepared with slight modifications
as described by Roberts (19). Polynucleotide kinase and calf
intestinal phosphatase were from Boehringer Mannheim GmbH,
γ-^{32}P-ATP was prepared as described by Johnson and Walseth (20).

RESULTS

Restriction endonuclease mapping. A 7.5 kb EcoRI DNA frag-
ment from *E. coli* which carries *ompA* has recently been cloned
in the vector plasmid pSC101 (21). Using this plasmid (pTU100)
the structural gene for protein II* has been located at the
junction of two BamHI fragments of very similar size, Bam-N
(1.87 kb) containing the NH_2-terminal and Bam-C (1.83 kb)
containing the CO_2H-terminal segment of the gene (15, see
fig. 1). Subcloning of this DNA segment into a high copy number
plasmid like pBR322 was not possible, most likely because
too high an *ompA* gene dosage effect is lethal (21, 15). There-
fore, DNA carrying *ompA* mutant alleles which produce
protein II* in much reduced amounts (22) were used as a
source to clone the two relevant BamHI fragments separate-
ly in plasmid pBR322 (15, 23). The mutants used were *ompA*-72
and *ompA*-31 (15; an insertion and a nonsense mutant of the
amber type, respectively, see below).

Prior to DNA sequencing cleavage maps for restriction
nucleases to be used were established within each of the
two BamHI fragments. To this end a mixture of the two frag-
ments from plasmid pTU100 was analysed for further cleavage
by the following restriction nucleases: AluI, HhaI, HaeIII,
HinfI, HhaI, HgaI, Sau3A. The size of the resulting sub-
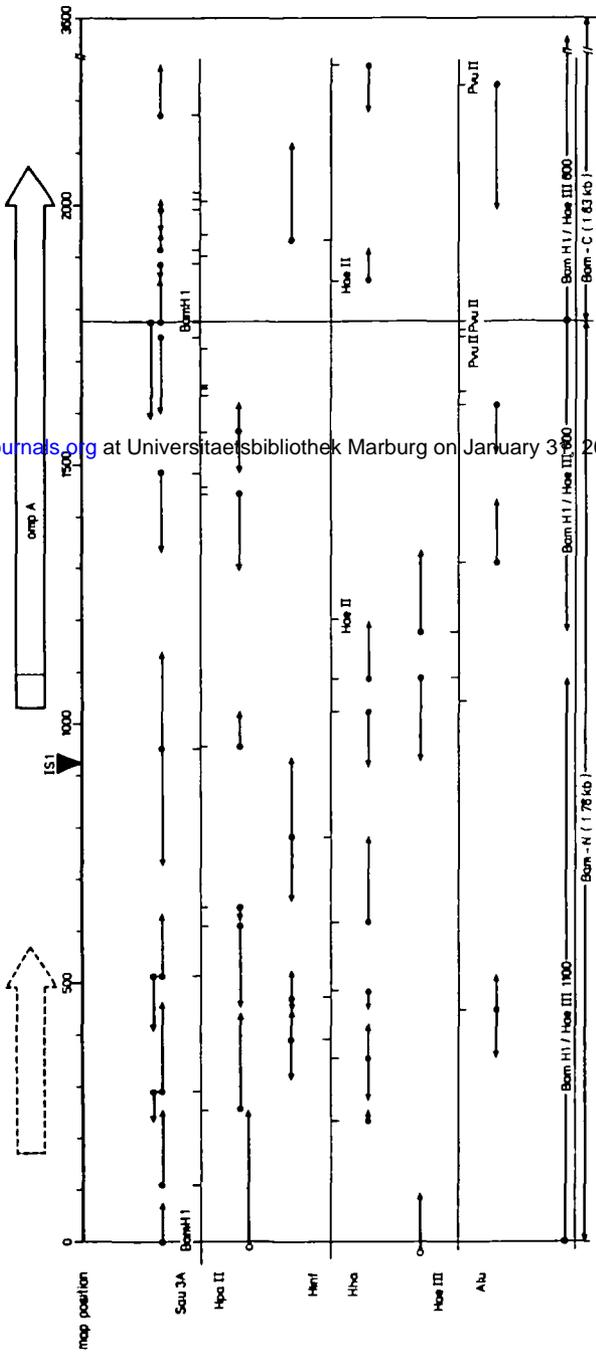
Fig. 1: Scheme of the sequences analysed. The maps of the restriction endonucleases used in the sequence analysis are shown starting at the left end of the 1.78 kb BamHI fragment (Bam-N). The dots and arrows indicate the labeled 5'-ends and the direction and the length of the nucleotide sequences determined. The vertical lines at positions 0, 1780, and 3600 indicate the borders of the two separately cloned BamHI fragments Bam-N and Bam-C. The BamHI/HaeIII fragments used to establish the physical map are also included. The dashed arrow in front of the ompA gene indicates a continuous reading frame for a hypothetical protein. The integration point of the IS1 element in the mutant ompA-72 is shown by the triangle. The dashed line within ompA indicates the transition from the signal sequence to the mature protein.

fragments was determined by electrophoresis on a 6 % polyacry-
lamide gel standardized with restriction fragments of known
length from fd DNA (24). Their order was determined from
the size distribution of cleavage products obtained by sub-
sequent partial cleavage of long terminal endlabeled re-
striction fragments (25), such as the two BamHI/HaeIII frag-
ments of about 1100 and 600 basepairs from fragment Bam-N
and the 800 basepair BamHI/HaeIII fragment from fragment
Bam-C (see fig. 1). From the results obtained a restriction
map was constructed for the Bam-N fragment and for the first
one third of the Bam-C fragment. This map covered completely
the region of the *ompA* gene and sufficed to identify and to
select restriction fragments for the following sequence
analysis.

DNA sequencing. DNA sequencing was carried out using as
a starting material mainly plasmids pTU302 and pTU301 which
contain the subcloned BamHI fragments from the amber mutant
ompA-31. The entire plasmid DNAs were digested with restric-
tion endonucleases Sau3A, HpaII, HaeIII, HinfI, HhaI, and
AluI, and terminally labeled using γ-^{32}P-ATP and polynucleo-
tide kinase. The mixtures of fragments were resolved by
electrophoresis on 6 % acrylamide gels next to reference
samples from vector DNA (pBR322) which had been cleaved by
the same enzymes. All bands appearing in addition to those
from pBR322 on the autoradiographs were eluted from the gel.
After secondary cleavage with an appropriate enzyme subfrag-
ments were separated on a second 6 % polyacrylamide gel,
isolated, and sequenced using the chemical degradative
methods of Maxam and Gilbert (17). Sequences were ordered
by sequence overlaps and according to the cleavage map
described above. For simplicity all parts of fragment Bam-N
were sequenced, whereas from plasmid pTU301 subfragments
from within or close to the *ompA* gene were selected
according to the size and to the secondary cleavage pro-
ducts. Figure 1 outlines the sequencing strategy used and
Figure 2 shows the composite nucleotide sequence obtained.
Alltogether a continuous sequence of 2271 basepairs was
established, of these 1036 precede the *ompA* gene and 197

follow the gene. Some additional sequences from fragment Bam-C could be arranged into a block of about 200 basepairs at the *ompA* distal end of the Bam-C fragment (not shown). As indicated in Figure 2 most of the sequence was determined in both DNA strands or is confirmed by the known amino acid sequence of protein II* (14).

Plasmid pTU104 carries insertion element IS1. The mutant phenotype exhibited by this plasmid, when present in strains with a chromosomal *ompA* mutation abolishing synthesis of protein II*, is a much reduced expression of *ompA*. A preliminary analysis of the plasmid DNA had shown that the 1.78 kb BamHI fragment had aquired an insertion of about 700 basepairs, i.e. instead of this fragment one of about 2.4 kb was found (15). Since it appeared likely that this insertion would impair *ompA* expression control and leave the structural gene intact we have localized the inserted material. To this end the BamHI fragment carrying the insertion was isolated from plasmid pTU202 (Table 1) by preparative electrophoresis on a 4 % polyacrylamide gel and cleaved separately with HinfI and HaeIII, end-labeled and sequenced from the resulting 5'-ends. As a result of this analysis sequences from about 280 basepairs from the new DNA segment were determined and found to be identical with the known nucleotide sequence from insertion element IS1 (26). The other sequences analysed, about 660 basepairs, were homologous to the *ompA*-31 DNA and did not reveal any base exchange. The insertion point of the IS1 element into the *ompA*-31 DNA was best to be seen in base sequences read from two nearby Hinf clevage sites, one at position 781 in the *ompA* DNA and the other at position 758 in IS1. The corresponding DNA sequences are shown in Figure 3. As observed in other integration sites of IS1 (27, 28) there is a duplication of nine basepairs in the recipient DNA (nucleotide 926 - 934). The position of the IS1 insertion is 110 basepairs in front of the start codon of the *ompA* gene.

DISCUSSION

We have determined a nucleotide sequence of 2271 base-

R1
R2
R3

1 GATCCGTTAACTACGAAAATAGGCAACTTATTCTTAAGGGGCAAGATTAATTTATGTTTCCCGTCACCACGACAAAAATTTGCGAGGCTCTTTCGAAA
        ***   ***   ***

R1
R2
R3

101 ATAGGGTTGATCTTTGTTGTCACTGGATGTACTGTCATCCATACAGTATCTCACAGGGGCTGGATTGATATGTACACTTCAGGCTATGCACATCGTTC
                                                                         MetTyrThrSerGlyTyrAlaHisArgSer
        ***   ***   ***   ***

R1
R2
R3

201 TTCGTCGTTCTCATCCGCAGCAAGTAAAATTGCGCGTGTCTCTACGGAAAACACTACAGCCGGGCCTTATAGTGAAGTTGTCTATCCGAAGATCAGCCC
    SerSerPheSerSerAlaSerLysArgIleAlaArgValSerThrGluAsnThrThrAlaGlyLeuIleGlyLeuIleValValTyrArgGluAspGlnPro
        ***   ***   ***   ***

R1
R2
R3

301 ATGATGACGCAACTTCTACTGTTGCCATTGTTACAGCAACTCGGTCAGCAATCGGTCAGCAACTCTGGTAACACCGCACAAAAACTGAGTCGGGAAT
    MetMetThrGlnLeuLeuLeuLeuProLeuLeuGlnGlnLeuGlyGlnSerArgTyrGlnLeuTrpMetThrProGlnAsnGlnLysLeuSerArgGluTrp
    ******   ***   ***   ***

R1
R2
R3

401 GGCTTCAGGCATCTGGGCTACCCTTAACGAAAGTAATGCAGATTAGCCAGCTCTCCCCTTGCCACACTGGGGAGTCAATGGTTCGCCGCTTTACGCACGGG
    ValGlnAlaSerGlyLeuProLeuGlyTrpLysValMetGlnIleLeuSerProGlnHisThrValGlnSerMetValAlaArgAlaLeuArgThrGly
        ***   ***   ***   ***

P1
P2
P3

501 CAATTACAGTGGTGATCGGTTGGTTGGCAGATGATTGACTGAAGAAGAGCATGCTGAACTTGTTGAGCGGCAAATGAAGGTAACGCTATGGGGTTT
    AsnTyrSerValValIleGlyTrpLeuAlaAspAlaAspLeuGluHisGluGluValLysAlaAlaAsnGluAlaLysGlyGlyValAlaMetGlyPhe
        ***   ***   ***   ***   ***

R1
R2
R3

601 ATTATACATTCCGGTAAGCGCATCCTCTCACGCCACGAGACAACTTTCCGGGCTAAAAATTCACTCTAATTTGTATCATTAAGTAAATTTAGGATTAATC
    IleIleHisSerGlyLysArgIleLeuSerArgHisGluThrThrPheArgAlaLysAsnSerLeu***
        ***   ***   ***   ***

R1
R2
R3

701 CTGGAACTTTTTTGTCGCCCAGCCAATGCTTTCAGTCGTGACTAATTTCCTTGCGGAGGCTTGTCTGAGCGGTTTCCGCGATTCTCTTCTGTAAATTGT
                                                                          ***   ***
        ***   ***   ***

R1
R2
R3

801 CGCTGACAAAAGATTAAACATACCTTATACAAGACTTTTTTTTCATATGCCTGACGGAGTTCACACTTGTAAGTTTCAACTACGTTGTAGACTTTAC
                                                                          ***
        ***

```
901  ATCGCCAAGGGTGCTCGGCATAAGCCGAAGATATCGGTAGAGTTAATATTCAGCAGATCCCCGGTGGGATTTAACCGTGTTATCTCGTTGGAGATAT    R1
                         ***              ***    ***            ***            ***                        R2
                                                                                                           R3

1001 TCATGGCGTATTTCGATGATAACGAGGCGCAAAAATGAAAAAGACACCTATCGCGATTGCAGTGCAGTGGCACTGGCTGGTTTCGCTACCGTAGGCGCAGGCCG  R1
                  MetLysLysThrAlaIleAlaIleAlaValaLeuAlaGlyPheAlaThrValAlaGlnAlaAla                               P2
                       ******                                              ***                              R3

1101 CTCCGAAAGATAACACCTGGTACTCTGGTGCTAAACTGGGCTGGTCCCGTACCATGATACTGGTTTCAACAACAATGGCCGACCCATGAAAATCCA          M1
     ProLysAspAsnThrTryTyrThrGlyAlaLysLeuGlyTrySerGlnTyrHisAspThrGlyPheGlyPheIleAsnAsnAsnGlyProThrHisGluAsnGln  R2
              ***                                                    ***                                      R3

1201 ACTGGGCGCTGGTGCTTTTGGTGGTTACCAGGTTAACCCGTATGTTGGCTTTGAAATGGGTTACGACGGTTAGGTCGTATGCCGTACAAAGGCAGCGTT        M1
                                                                                                      ***        M2
     LeuGlyAlaGlyAlaPheGlyGlyTyrGlnValAsnProTyrValGlyPheGluMetGlyTyrAspGlyArgMetProTyrLysGlySerVal              R3
                  ***                                                      ***                               

1301 GAAAACGGTGCCATACAAAGCTCAGGGCCGTTCAACTGACCGCTAAACTGGGTTACCCAATCACTGACGACCTGGACATCCTCTGCTGGCATGG           R1
                                                                                        ***                     R2
     GluAsnGlyAlaIleTyrLysGlyAlaGlnGlnGlyValGlnGluLeuThrAlaLysLeuGlyTyrProIleThrAspAsnLeuAspIleTyrThrArgLeuGlyMetVal  R3
                  ***                                                      ***                               

1401 TATGGCGTGCAGACACTAAATCCAACGTTTATGGTAAAAACCACGACCACCGGCGTTTCTCCGGTCTTGACTGGCGGTGTTGAGTACGCGATCACTCCTGA        M1
                                                                                            ***  ***             M2
     TrvargAlaAspThrLysAsnSerAsnValTyrGlyLysAsnHisAspThrGlyValSerProValPheAlaGlyGlyValGluTyrAlaIleThrProGlu        R3

1501 AATCGGCTACCCGTCTGGAATACCAGTGGACGAACAACATCGGTGACGCACACCGGCACCGTCGTCCGGACAACGGCATGCCTGAGCCTGGGTGTTTCC         M1
                             ***                                                                                 M2
     IleAlaThrArgLeuGluTyrGlnThrTryThrAsnAsnIleGlyAspAlaHisThrIleGlyThrArgProAspAsnGlyMetLeuSerLeuGlyValSer       R3

1601 TACCGTTTCGGTCAGGGCGAGCGAGCCGCCTCCACGTCGTCCCGGCTCCAGCTCCGGCACCGGAAGTACAGACCCAGACCACTTCACTCTGAAGTCTGACGTTC    R1
                                                                                                      ***        R2
     TyrArgPheGlyGlnGlyGlyGluAlaAlaProValAlaAlaProProAlaProAlaProGluValGlnThrGlnThrLysPheHisPheThrLysSerAspValLeu R3

1701 TGTTCAACTTCAACAAAGCAACCCTGAAACCGGAAGGTCAGGTCAGGCTGTCTCTGAATCAGTCTGTACAGCCAGCTGAGCCAACTTGGATCCGAAAGACGGTTCCGT  M1
                                                                                                     ***         R2
     PheAsnPheAsnLysAlaThrLeuLysProGluGlyGlnAlaAlaLeuAlaLeuAsnGlnSerAsnLeuTyrSerGlnLeuSerAsnLeuAspProLysAspGlySerVal  R3
```

1801 AGTGTCTGGGTTACACCGACCGCATCGGTTCTGACGCTTACAACCAGGGTCTGTCCGAGCGCGTGGCAGTCTGTTGTTGATTACCTGATCTCCAAA •

P1
P2
P3

ValValLeuGlyTyrThrAspArgIleGlySerAspAlaTyrAsnGlnGlyLeuSerGluArgGlnArgArgGlnSerValValAspTyrLeuIleSerSerLys

***
***
***

1901 GGTATCCCGGCAGACAAGATCTCCGCACGTGGTATGGGCGAATCCAACCCGGTTACTGGCAACACCTGGACAACAGCGTGAAACAGCGTGCTGCACTGATCG •

P1
P2
P3

GlyIleProAlaAspLysIleSerAlaArgGlyMetGlyGluSerAsnProValThrGlyAsnThrCysAspAsnValLysGlnArgGlnArgGlnAlaLxLeuIleAsp

***
**
***

2001 ACTGCCTGGCTCCGGATCGTCGCGTAGAGATCGAAGTTAAAGGTATCAAAGACGTTGTAACTCAGCCGAGGCCTTAAGTTCTCGTCTGGTAGAAAACGC •

R1
R2
R3

CysLeuAlaProAspArgArgValGluIleGluValLysGlyIleLysAspValValThrGlnProAsnAla***

***
***
***

2101 TGCTGCGGGTTTTTTTTGCCTTAGTAAATTGAACTGACTTTCGTGCAGTTATTCCTTACCCAGCAATCCTGCCAGATCCTGCTTCAGAGAAGACATTTT •

R1
R2
P3

****
***
***

2201 ATTCGCGTATTTCTCTTGTTTTCCGCATCTTCAATCAGCTGAACAATCGTTTCAGAAAGCCGTTTTACCGGC •

P1
P2
P3

***

Fig. 2: Nucleotide sequence of the ompA DNA plus strand. Included in the sequence are the amino acid sequences of the ompA protein (positions 1037 - 2174) and of a hypothetical gene product (positions 172 - 666). Stop codons in the three reading frames are indicated as asterisks. In mutant ompA-31 the TGG codon (Try) in position 1118 is a TAG amber codon.

3019
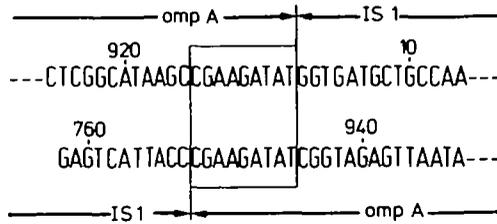
Fig. 3: Integration point of IS1 in the DNA of the mutant *ompA*-72. The sequence of the HinfI fragment starting at position 781 in the *ompA* DNA (upper line) and of the HinfI fragment starting at position 758 in the IS1 DNA (lower line) are aligned with respect to the nine repeated nucleotides. The first five nucleotides in the lower sequence were not read from the gel but inferred from the position of cleavage by restriction endonuclease HinfI.

pairs from cloned *E. coli* DNA which contains *ompA* as demonstrated by genetical and biochemical analysis (15). Within this sequence from base positions 1037 to 2174 an amino acid sequence encompassing 346 residues can be deduced, and this sequence, starting with amino acid 22, agrees with the primary structure of protein II* (14). In the DNA from mutant *ompA*-31 there is a codon of the amber type corresponding to residue 28 which in the wild type protein is tryptophan. Unquestionably this amber codon is responsible for the phenotype of the mutant *ompA*-31; it has now been shown that this gene is not expressed at all in a strain lacking nonsense suppressors, and it is well expressed (about wild type level) when amber suppressors such as *supD* or *supF* are present (15). Originally the nucleotide sequence did not agree at five positions with the amino acid sequence. These discrepancies were resolved at four positions in favor of the DNA sequence and at one position in favor of the protein sequence.

The twentyone amino acid residues preceding the sequence of protein II* constitute a typical signal sequence which was predicted for this protein already some time ago (29). Very recently Movva et al. (30) published a DNA sequence encoding this signal sequence plus the first nineteen amino acid residues of the *ompA* protein. Their sequence completely agrees with ours. These authors have discussed the signifi-

cance of the *ompA* signal sequence particularly regarding the loop model for protein secretion across a membrane (2, 31), and the relevant points will not be repeated here.

The transcription initiation site of the *ompA* gene is not yet known. However, this site in all probability is located not as near the translation initiation as in case of the *lpp* gene (encoding the outer membrane lipoprotein) where mRNA is initiated 38 basepairs upstream of the ATG initiation codon (32). The presence of IS1 110 basepairs upstream severely reduces *ompA* expression (< 10 % of the wild type level). Furthermore, significant homologies to the two characteristic promoter signal sequences at positions -10 ("Pribnow box") and -35 do not exist up to 110 basepairs upstream of translation initiation. The first theoretically possible promoter region is found between base positions 870 and 930 where an RNA polymerase recognition site may be realized with the sequence GTTGTAAGACTTT (positions 887 - 898) which, however, is followed by a very questionable Pribnow heptamer CATAAGC (positions 919 - 925). DNA binding to nitrocellulose filters in the presence of RNA polymerase (33), using various restriction endonuclease fragments, strongly indicates the existence of RNA polymerase binding sites between basepairs 620 and 780 (unpublished data). In this area a potential polymerase recognition sequence CGAGACAACTTT (positions 636 - 647) and "Pribnow box" TCTAATT (positions 665 - 671) can be found suggesting that transcription may be initiated at position 678, as far as 360 basepairs upstream of translation initiation. It should be noted that in this extended region no poly-peptide of reasonable length is encoded in either DNA strand. There are many stretches of self-complementary nucleotide sequences in this potential mRNA leader segment as are in the first half of the *ompA* gene. This, and in analogy to the very stable mRNA of the outer membrane lipo-protein (34), may contribute to a stable secondary structure of *ompA* mRNA and thus be responsible for its high stability (35).

From basepairs position 2091 to 2116, immediately be-

hind *ompA* translation termination, a region exist which
can be folded in a hairpin structure ending with a stretch
of nine T. This structure is comparable to other trans-
criptional stop signals, as e.g. in phage fd and phage λ
(for review see ref. 36).

It may finally be mentioned briefly that from basepairs
172 to 666, a protein with a molecular weight of about
17.000 (165 residues) may be encoded. We have not observed
a candidate for this hypthetical protein when pTU100
directed protein synthesis was examined in minicells (21),
however, only polypeptides incorporated into cell envelopes
were studied and a soluble protein would have escaped de-
tection.

ACKNOWLEDGEMENT

REFERENCES

1 Garten, W., Hindennach, I. and Henning, U. (1975) Eur. J.
  Biochem. 59, 215 - 221

2 DiRienzo, J. M., Nakamura, K. and Inouye, M. (1978) Ann.
  Rev. Biochem. 47, 481 - 532

3 Henning, U., Hohn, B. and Sonntag, I. (1973) Eur. J.
  Biochem. 39, 27 - 36

4 Endermann, R., Krämer, C. and Henning, U. (1978) FEBS
  Letters 86, 21 - 24

5 Datta, D.B., Arden, B. and Henning, U. (1977) J. Bacteriol.
  131, 821 - 829

6 van Alphen, L., Havekes, L. and Lugtenberg, B. (1977) FEBS
  Letters 75, 285 - 290

7 Chai, T. and Foulds, J. (1974) J. Mol. Biol. 85, 465 - 474

8 Skurray, R.A., Hancock, R.E.W. and Reeves, P. (1974) J.
  Bacteriol. 119, 726 - 735

9 Schweizer, M. and Henning, U. (1977) J. Bacteriol. 129,
  1651 - 1652

10 Braun, V. (1975) Biochim. Biophys. Acta 415, 335 - 377

11 Sonntag, I., Schwarz, H. Hirota, Y. and Henning, U. (1978)
   J. Bacteriol. 136, 280 - 285

12 Manning, P.A., Pugsley, A.P. and Reeves, P. (1977) J. Mol. Biol. 116, 285 - 300

13 Lugtenberg, B. Peters, R., Bernheimer, H. and Berendsen, W. (1976) Molec. Gen. Genet. 147, 251 - 262

14 Chen, R., Schmidmayr, W., Krämer, C., Chen-Schmeisser, U. and Henning, U. (submitted for publication)

15 Bremer, E., Beck, E., Hindennach, I., Sonntag, I. and Henning, U. (1980) Molec. gen. Genet., in press

16 Clewell, D.B. (1972) J. Bacteriol. 110, 667 - 676

17 Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560 - 564

18 Gray, C.P., Sommer, R., Polke, C., Beck, E. and Schaller, H. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 50 - 53

19 Roberts, R.J. (1976) CRC Critical Reviews in Biochem. 2, 123 - 164

20 Johnson, R.A. and Walseth, T.F. (1979) in: Adv. Cyclic Nucleotide Res. 10, 135 - 167, Raven Press, New York

21 Henning, U., Royer, H.-D., Teather, R.M., Hindennach, I. and Hollenberg, C.P. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 4360 - 4364

22 Henning, U., Sonntag, I. and Hindennach, I. (1978) Eur. J. Biochem. 92, 491 - 498

23 Bolivar, F., Rodriguez, R., Greene, P.J., Bettlach, M., Heyneker, H.L., Boyer, H.W., Crosa, J and Falkow, S. (1977) Gene 2, 95 - 113

24 Beck, E., Sommer, R., Auerswald, E.A., Kurz, Ch., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M. (1978) Nucleic Acids Res. 5, 4495 - 4504

25 Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3, 2387 - 2397

26 Ohtsubo, H. and Ohtsubo, E. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 615 - 619

27 Calos, M., Johnsrud, L. and Miller, J.H. (1978) Cell 13, 411 - 418

28 Grindley, N.D.F. (1978) Cell 13, 419 - 426

29 Sekizawa, J., Inouye, S., Halegoua, S. and Inouye, M. (1977) Biochem. Biophys. Res. Comm. 77, 1126 - 1133

30 Movva, N.R., Nakamura, K. and Inouye, M. (1980) J. Biol. Chem. 225, 27 - 29

31 Inouye, M.and Halegoua, S. (1980) CRC Critical Review Biochem. (in press)

32 Nakamura, K. and Inouye, M. (1979) Cell, 18, 1109 - 1117

33 Seeburg, P.H. and Schaller, H. (1975) J. Mol. Biol. 92, 261 - 277

34 Nakamura, K., Pirtle, R.M., Pirtle, I.L., Takeishi, K. and Inouye, M. (1980) J. Biol. Chem. 255, 210 - 216

35 Garten, W. and Henning, U. (1976) FEBS Letters 67, 303 - 305

36 Rosenberg, M. and Count, D. (1979) Ann. Rev. Genetics 13, 319 - 353

APPENDIX

Protein-chemical evidence for the signal sequence of the
pro-*ompA* protein.

Konrad Gamon, Waltraud Schmidmayr and Ulf Henning
Max-Planck-Institut für Biologie, D-7400 Tübingen

The *E. coli* outer membrane protein II* (*ompA* protein) is
produced from a precursor (1 - 3) and from DNA sequencing of
the *ompA* gene (see preceding communication) or part of it (4)
it has been inferred that this precursor possesses, at the
NH_2-terminal end, the signal sequence: Met - Lys - Lys - Thr -
Ala - Ile - Ala - Ile - Ala - Val - Ala - Leu - Ala - Gly -
Phe - Ala - Thr - Val - Ala - Gln - Ala.

Protein synthesis in minicells directed by plasmid pTU100
(carrying the wild type *ompA* gene) led to massive accumula-
tion of protein II* and a second cell envelope protein which
most likely represents the pro-*ompA* protein since both proteins
were precipitable with antibody raised against protein II*
(5). In order to obtain direct evidence for the signal sequence
mentioned we have allowed 2×10^{10} minicells harboring pTU100
to synthesize protein (in 1 ml incubation mixture) under con-
ditions described (5) in the presence of 50 μCi ^{35}S-methionine
(1500 Ci/mmol), or 200 μCi ^{14}C-lysine (340 μCi/mmol), or
500 μCi ^{3}H-isoleucine (40 Ci/mmol; all isotopes from Amersham).
Radioactive protein II* and its precursor were separated by
SDS polyacrylamide gel electrophoresis (5) and the latter was
eluted by electrodialysis. Purified non-radioactive protein II*
(6) was added as carrier (5 mg) and the mixtures were subjec-
ted to manual Edman degradation (7). Cold PTH-methionine,
-lysine, or -isoleucine were added, after each step of degra-
dation, to the corresponding radioactive samples. The amino
acid derivatives were separated by thin-layer chromatography
(7), the spots from the three PTH-amino acids were removed
from the chromatograms, and counted in a liquid scintillation
spectrometer.(Absolute yields of radioactivity recovered per
labeled residue have not been measured because of too many

variables that cannot be corrected for without much effort,
e.g., degree of solubilization of a PTH-amino acid from the
silica gel of chromatograms, individual losses during and
after chromatography, counting efficiency of the labeled pro-
tein, etc.) The results are shown in Figure 1. Those obtained
for lysine are somewhat unsatisfactory because they do not
allow to definitely exclude the presence of lysine in position
4 of the signal sequence. Also, and as can be judged from the
radioactivity recovered in PTH-lysine, incorporation of this
amino acid into minicell protein was only 10 % of that found
for the other two amino acids. All other data, however, prove
the signal sequence deduced from the DNA sequence and vice
versa show that the corresponding protein in minicells does
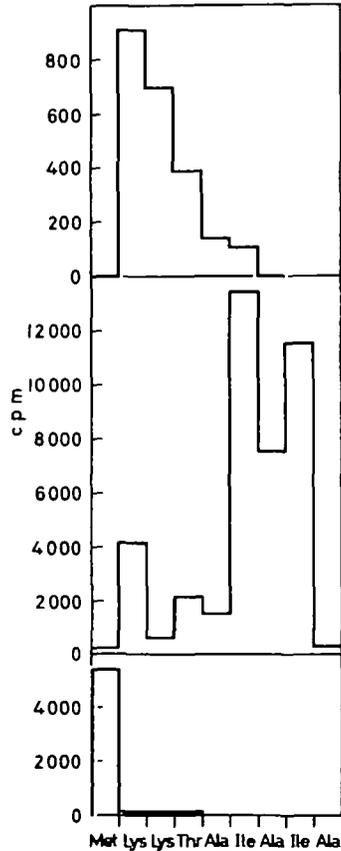
Fig. 1: Radioactivity in PTH-amino
acids released from pro-*ompA* pro-
tein. Bottom panel, labeled with
^{35}S-methionine; middle panel, la-
beled with ^{3}H-isoleucine; top pa-
nel, labeled with ^{14}C-lysine. The
residues shown (reflecting conse-
cutive Edman degradation steps)
are those expected from the signal
sequence.

in fact consist of pro-*ompA* protein.

REFERENCES

1   Sekizawa, J., Inouye, S., Halegoua, S. and Inouye, M. (1977)
    Biochem. Biophys. Res. Comm. 77, 1126 - 1133
2   Halegoua, S. and Inouye, M. (1979)J. Mol. Biol. 130,
    39 - 61
3   DiRienzo, J. M. and Inouye, M. (1979) Cell 17, 155 - 161
4   Movva, N.R., Nakamura, K. and Inouye, M. (1980) J. Biol.
    Chem. 255, 27 -29
5   Henning, U., Royer, H.-D., Teather, R.M., Hindennach, I.
    and Hollenberg, C.P. (1979) Proc. Natl. Acad. Sci. U.S.A.
    76, 4360 - 4364
6   Hindennach, I. and Henning, U. (1975) Eur. J. Biochem. 59,
    207 - 213
7   Edman, P. and Henschen, A. (1975) in: Protein Sequence
    Determination, Needleman, S.B., Ed., pp. 232 - 279, Springer
    Verlag, Berlin - Heidelberg - New York