Proc. Natl. Acad. Sci. USA
Vol. 74, No. 3, pp. 1004–1008, March 1977
Biochemistry

Amino acid sequence for the peptide extension on the prolipoprotein of the *Escherichia coli* outer membrane

(mRNA/*in vitro* translation/lipoprotein/Edman degradation)

SUMIKO INOUYE, SANSAN WANG, JUN SEKIZAWA, SIMON HALEGOUA, AND MASAYORI INOUYE

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

Communicated by Seymour S. Cohen, January 6, 1977

ABSTRACT **The messenger RNA for the lipoprotein of the *E. coli* outer membrane was found to code for a putative precursor, prolipoprotein, which has 20 additional amino acid residues extending from the amino terminus of the lipoprotein. Using the prolipoprotein synthesized in an *E. coli* cell-free system directed by purified messenger RNA for the lipoprotein, the complete amino acid sequence of the amino-terminal precursor region was determined to be as follows: Met-Lys-Ala-Thr-Lys-Leu-Val-Leu-Gly-Ala-Val-Ile-Leu-Gly-Ser-Thr-Leu-Leu-Ala-Gly-. It was also found that the prolipoprotein that accumulates in toluene-treated cells has the same sequence. The significance of the amino acid sequence is discussed in terms of the mechanism of biosynthesis and assembly of the lipoprotein in the *E. coli* outer membrane.**

Escherichia coli contains two distinct membranes, the cytoplasmic membrane and the outer membrane. The outer membrane contains a unique structural protein, a lipoprotein of molecular weight 7200, which is one of the most extensively investigated membrane proteins (see review refs. 1 and 2). The lipoprotein provides an ideal model system for the study of the mechanism of biosynthesis and assembly of membrane proteins. Earlier studies have included analysis of the unusual resistance to puromycin of the biosynthesis of the lipoprotein as well as the biosynthesis on isolated *E. coli* polyribosomes (3, 4). The messenger RNA for the lipoprotein has been shown to have extraordinary stability (5, 6) and has been purified (7, 8). The cell-free products directed by the purified mRNA have been precipitated with antilipoprotein serum and found to migrate slower than the lipoprotein upon sodium dodecyl sulfate gel electrophoresis (7, 8). The messenger RNA for the lipoprotein was found to consist of about 360 nucleotides, providing a much longer sequence than is required to code for the lipoprotein of only 58 amino acid residues (9). Recently we have found that the lipoprotein is probably produced from a precursor which accumulates to toluene-treated cells (3, 10, 11). This precursor contains a peptide extension at the amino terminus of the lipoprotein.

In the present paper, we analyze the product synthesized in a cell-free system directed by the purified mRNA. We have found that the cell-free product has an extension of 20 additional amino acid residues at the amino terminus of the lipoprotein. The complete amino acid sequence of the peptide extension of the product, prolipoprotein, was determined. The unique features of this sequence have led us to propose plausible functions for this region of the prolipoprotein.

MATERIALS AND METHODS

Cell-Free Protein Synthesis. Cell-free protein synthesis was carried out as previously described (7). The mRNA for the lipoprotein was purified by Sephadex G-200 filtrations as previously described (7), and in most cases the SF-2 fraction was used unless otherwise mentioned. The cell-free protein synthesis was carried out in a 300 μl reaction mixture with 30 μCi of a radioactive amino acid (7). After 30 min incubation at 37°, the reaction mixture was put on ice and 1.5 ml of 3 mM magnesium acetate was added. The mixture was then centrifuged at 150,000 × *g* for 30 min. By this centrifugation about 30% of nonradioactive proteins were removed as a pellet but about 80% of cell-free products remained in the supernatant. To the supernatant was added 1.5 ml of 10% trichloroacetic acid and the mixture was incubated for 30 min in a boiling-water bath. The resultant precipitate was collected by centrifugation and washed three times with 2 ml of ether/ethanol (1:1) and two times with 2 ml of ether. The final precipitate was dried with an air jet. In some cases, the cell-free products were precipitated with antisera as described previously (7). The immunoprecipitate was suspended in 2 ml of 5% trichloroacetic acid, incubated in a boiling-water bath for 30 min, and washed as described above. Protein synthesis in toluene-treated cells was carried out as described previously (3, 10), and the prolipoprotein from the toluene-treated cells was purified by immunoprecipitation followed by sodium dodecyl sulfate/gel electrophoresis as described previously (11). The prolipoprotein was extracted from the gel with water and precipitated with 67% acetone.

Determination of Amino Acid Sequence. Trichloroacetic acid-insoluble materials from the cell-free system or acetone-precipitated prolipoprotein from the toluene-treated cells were dissolved in 0.8 ml of hexafluoroacetone and the solution was applied to a JEOL Sequence Analyzer JAS-47K. The 1-chlorobutane extract from each step was dried with an air jet and the radioactivity in each extract was then measured in a liquid scintillation counter.

Polyacrylamide Gel Electrophoresis. Sodium dodecyl sulfate/polyacrylamide gel electrophoresis was carried out as described previously (12).

Other Materials Used. The following radioactive amino acids were used: L-[^{3}H]alanine (41 Ci/mmol, Schwarz/Mann), L-[^{14}C]arginine (0.3 Ci/mmol, Schwarz/Mann), L-[^{14}C]asparagine (0.13 Ci/mmol, Amersham/Searle), L-[^{3}H]aspartic acid (15.8 Ci/mmol, New England Nuclear), L-[^{3}H]glutamic acid (26 Ci/mmol, Amersham/Searle), [^{3}H]glycine (43 Ci/mmol, Amersham/Searle), L-[^{3}H]isoleucine (30 Ci/mmol, Schwarz/Mann), L-[^{3}H]leucine (60 Ci/mmol, Schwarz/Mann), L-[^{3}H]lysine (21 Ci/mmol, Amersham/Searle), L-[^{35}S]methionine (111 Ci/mmol, New England Nuclear), L-[^{3}H]proline (58 Ci/mmol, Amersham/Searle), L-[^{3}H]serine (17 Ci/mmol, Amersham/Searle), L-[^{3}H]threonine (2 Ci/mmol, New England Nuclear), L-[^{3}H]tyrosine (43 Ci/mmol, Amersham/Searle), and L-[^{3}H]valine (17 Ci/mmol, Amersham/Searle).

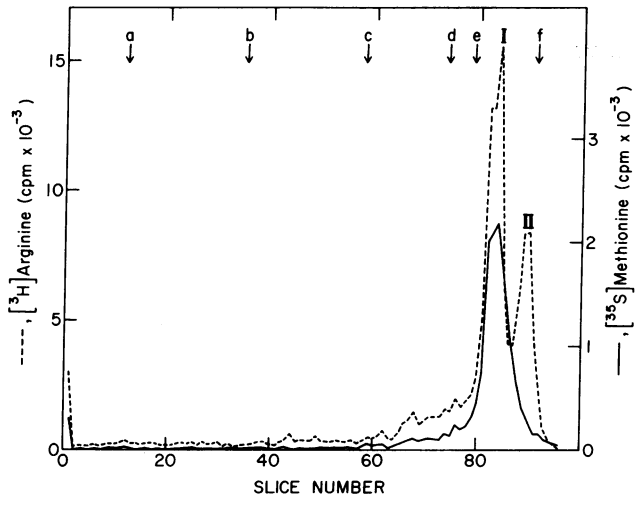

FIG. 1. Sodium dodecyl sulfate/polyacrylamide gel electrophoresis of the cell-free products and the immunoprecipitate of the products from toluene-treated cells. The cell-free product was labeled with [^{35}S]methionine in an *E. coli* cell-free system directed by the purified mRNA as described previously (7). The products for toluene-treated cells were labeled with [^{3}H]arginine, and immunoprecipitated from the envelope as described previously (3, 10). Peaks labeled I and II indicate the new form of the lipoprotein and the lipoprotein, respectively, from toluene-treated cells (11). Arrows with letters indicate the positions of the internal molecular weight standards (14): a, dimer and b, monomer of 5-dimethylaminonaphthalene-1-sulfonyl-(Dns-) bovine serum albumin; c, dimer and d, monomer of Dns-egg white lysozyme; e, cytochrome *c*; f, Dns-insulin.

RESULTS

Identification of the Cell-Free Product. In the previous paper (7), the cell-free product directed by the purified mRNA for the lipoprotein was shown to be crossreactive with antiserum against the lipoprotein and to have the same carboxyl terminal structure as the lipoprotein produced *in vivo*. However, the protein produced *in vitro* seemed to be slightly larger than the *in vivo* product. In Fig. 1, the cell-free product labeled with [^{35}S]methionine was subjected to sodium dodecyl sulfate/polyacrylamide gel electrophoresis together with immunoprecipitated products from toluene-treated cells labeled with [^{3}H]arginine. The products from the toluene-treated cells are separated into two peaks; one peak migrates to the normal lipoprotein position (peak II) and the other to a position corresponding to an apparent molecular weight of 10,000–15,000 (peak I; putative precursor, prolipoprotein). On the other hand, the cell-free product appears as a single peak, comigrating with the higher molecular weight product of the toluene-treated cells.

Determination of the Length of the Extended Region of the Cell-Free Product. From the above results, it was thought that the cell-free product should have a peptide extension at the amino terminus of the lipoprotein. In order to determine the length of the extended region, the cell-free product was labeled with [^{3}H]leucine and subjected to 46 consecutive Edman degradations in a sequenator. Fig. 2 shows the release of radioactivity at each cycle of the Edman degradation. As can be seen from Fig. 2, radioactivity peaks appeared at cycles 6, 8, 13, 17, 18, 30, 37, and 44. As expected from sequenator analysis, the recovery of radioactivity decreases with increasing cycles, and there is always some trailing of radioactivity from the previous cycle because of incomplete degradation at the previous cycle. Therefore, the radioactivity at cycle 18 is slightly higher than that at cycle 17 because of the trailing of the radioactivity from

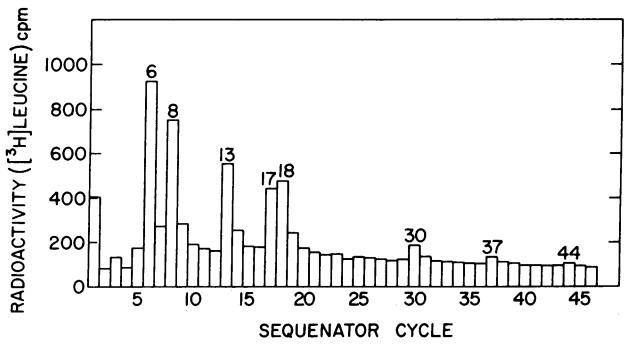

FIG. 2. Forty-six consecutive Edman degradations of the cell-free product labeled with [^{3}H]leucine. The preparation of the product and the Edman degradation were carried out as described in *Materials and Methods*. The total radioactivity applied to the sequenator was 3.6×10^4 cpm.

cycle 17. The reason for the appearance of some radioactivity at cycle 1 is unknown. However, in other experiments with the immunoprecipitate of the cell-free product or toluene-treated cell product (see Fig. 6) labeled with [^{3}H]leucine no significant radioactivity at cycle 1 was detected.

The complete amino acid sequence of the lipoprotein has been determined (13), and it has leucine residues at the 10th, 17th, 24th, and 49th positions from its amino terminus (see Fig. 3). Thus, one can notice that the distribution of leucine residues in the cell-free product at the 30th, 37th, and 44th positions coincides very well with that of the 10th, 17th, and 24th positions of the lipoprotein (Fig. 3). It is also important to point out that there are no leucine residues between the 21st and 29th positions in Fig. 2. Therefore, it appears that the cell-free product must have a peptide extension of 20 amino acid residues at the amino terminal end of the lipoprotein. Fig. 2 therefore also indicates that there are five leucine residues in the extended region at the 6th, 8th, 13th, 17th, and 18th positions.

If the extended region has 20 amino acid residues, one can predict the appearance of amino acids after the 21st cycle of the Edman degradation, because the amino acid sequence of the lipoprotein is known. For example, serine residues are known to be at the 2nd, 3rd, 11th, 12th, 25th, and 33rd positions of the lipoprotein (13) (see Fig. 3). Thus, one can predict that serine residues should appear at the 22nd, 23rd, 31st, 32nd, 45th,

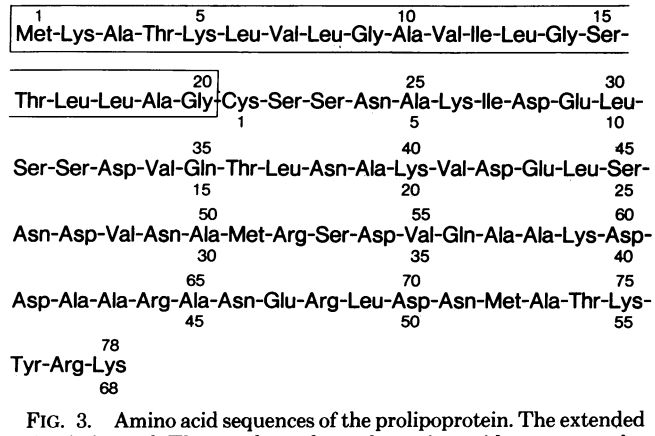

FIG. 3. Amino acid sequences of the prolipoprotein. The extended region is framed. The numbers above the amino acids represent the positions of the prolipoprotein amino acid residues counted in sequences from the amino terminus. The numbers under the amino acids represent the positions counted in sequence from the lipoprotein amino terminus. The amino acid sequence of the lipoprotein is according to Braun and Bosch (13).

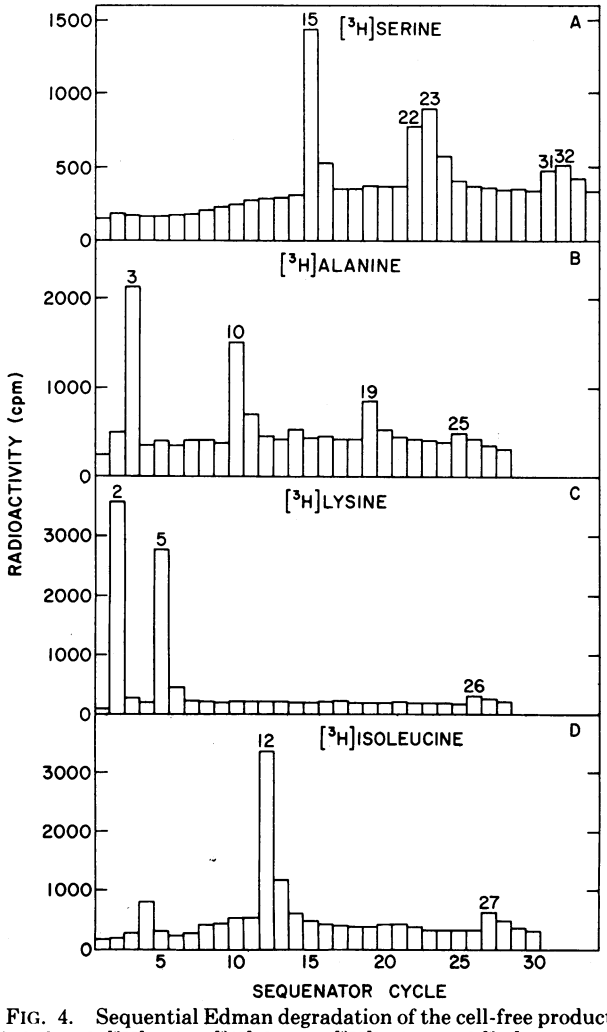

FIG. 4. Sequential Edman degradation of the cell-free products labeled with [³H]serine, [³H]alanine, [³H]lysine, and [³H]isoleucine. The preparation of the cell-free products and the Edman degradation were carried out as described in *Materials and Methods*. For the experiment with [³H]serine and [³H]lysine, the immunoprecipitates of the cell-free product were used. The radioactivities applied to the sequenator were 1.1×10^5, 1.1×10^5, 7.4×10^4, and 1.3×10^5 cpm for [³H]serine (*A*), [³H]alanine (*B*), [³H]lysine (*C*), and [³H]isoleucine (*D*), respectively.

and 53rd cycles of the Edman degradation of the cell-free product. Fig. 4*A* shows Edman degradation of the cell-free product labeled with [³H]serine up to the 34th cycle. The radioactivities clearly appeared at the 22nd, 23rd, 31st, and 32nd cycles as predicted. Fig. 4*A* also shows that there is one serine residue at the 15th position of the extended region. Fig. 4*B*, *C*, and *D* provides further confirmation of the length of the extended region. Alanine residues appeared at the 25th cycle of Edman degradation as predicted (see Fig. 3) as well as at the 3rd, 10th, and 19th positions (Fig. 4*B*). Lysine residues were detected at the 26th position as predicted (see Fig. 3) as well as at the 2nd and 5th positions of the extended region (Fig. 4*C*). Furthermore, isoleucine residues appeared at the 27th cycle of Edman degradation as predicted as well as at the 12th position of the extended region of the cell-free product (Fig. 4*D*).

Determination of the Amino Acid Sequence of the Extended Region. In order to determine the complete amino acid sequence of the extended region, Edman degradation was carried out with the cell-free product labeled with other amino acids. Fig. 5 shows the results from using [³H]valine, [³H]

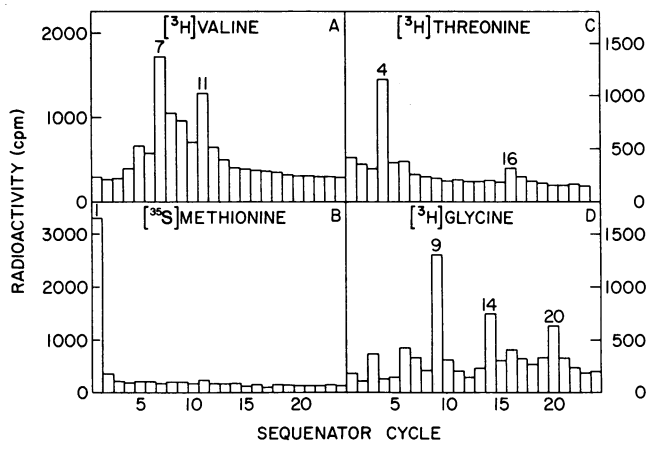

FIG. 5. Sequential Edman degradation of the cell-free products labeled with [³H]valine, [³⁵S]methionine, [³H]threonine, and [³H]-glycine. The preparation of the cell-free products and the Edman degradation were carried out as described in *Materials and Methods*. The radioactivities applied to the sequenator were 1.7×10^5, 2.3×10^4, 5.1×10^4, and 1.1×10^5 cpm for [³H]valine (*A*), [³⁵S]methionine (*B*), [³H]threonine (*C*), and [³H]glycine (*D*), respectively.

threonine, [³⁵S]methionine, and [³H]glycine. Two valine residues were found at the 7th and 11th positions (Fig. 5*A*). Two threonine residues were at the 4th and 16th positions (Fig. 5*B*); a methionine residue was at the 1st position; (Fig. 5*C*); and three glycine residues were at the 9th, 14th, and 20th positions (Fig. 5*D*). The appearance of glycine residues in the extended region is especially interesting because the lipoprotein lacks glycine residues. The results with [³H]glycine (Fig. 5*D*) were also seen when the [³H]glycine-labeled cell-free product precipitated with antiserum against the lipoprotein was sequenced.

Other amino acids ([¹⁴C]arginine, [³H]tyrosine, [³H]glutamic acid, [³H]glutamine, [³H]aspartic acid, [¹⁴C]aspargine, and [³H]proline) were also tested, but they were not detected in the extended region. Tryptophan, cysteine, phenylalanine, and histidine were not examined because the higher molecular weight product of the toluene-treated cells does not contain these amino acid residues (11) and they were hardly incorporated into the cell-free product of the present system. From these results the amino acid sequence of the peptide extension of the prolipoprotein was determined as shown in Fig. 3.

Identification of Peak I Protein of Toluene-Treated Cells. The protein at peak I produced in toluene-treated cells (Fig. 1) has been proposed to be a precursor of the lipoprotein (prolipoprotein) (11). It has the same carboxyl-terminal structure as the lipoprotein and has at least 18 to 19 extra amino acids extending from the amino terminus of the lipoprotein. It has a methionine residue at its amino terminus. Fig. 6 shows a further characterization of the peak I protein by sequential Edman degradation of the protein labeled with [³H]leucine. The result was exactly the same as that obtained for the cell-free product (see Fig. 2). The peak I protein also has leucine residues at the 6th, 8th, 13th, 17th, and 18th positions. From these results it was concluded that the cell-free product is identical to the peak I protein accumulated in toluene-treated cells.

DISCUSSION

In the present paper, we have determined the complete amino acid sequence of the peptide extension of the putative lipoprotein precursor, prolipoprotein, as Met-Lys-Ala-Thr-Lys-Leu-Val-Leu-Gly-Ala-Val-Ile-Leu-Gly-Ser-Thr-Leu-Leu-Ala-Gly- (see Fig. 3). This sequence has several unusual features:

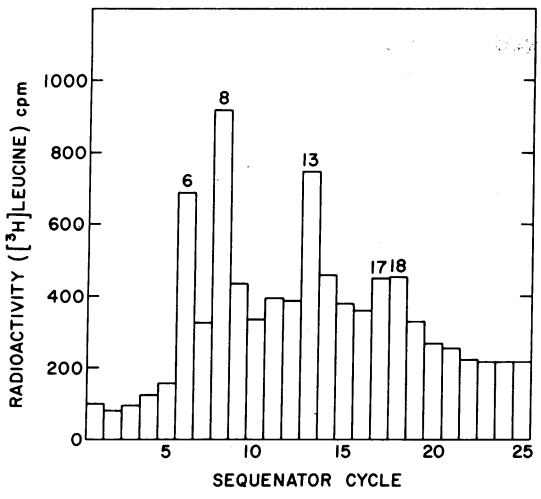

FIG. 6. Sequential Edman degradation of the peak I protein produced in toluene-treated cells. The peak I protein was labeled with [^{3}H]leucine as described previously (3, 7) and purified by sodium dodecyl sulfate/polyacrylamide gel electrophoresis. The peak I protein was extracted with water and precipitated with 67% acetone. The precipitate sample (7.5×10^4 cpm) was dissolved in hexafluoroacetone and then applied to the sequenator.

(*a*) The extended region is basic and positively charged at neutral pH because it contains two lysine but no acidic amino acid residues. (*b*) This region contains three glycine residues which are not present in the lipoprotein; it is particularly interesting that the last amino acid residue of the extended region is glycine, where a specific enzyme must process the prolipoprotein to produce the lipoprotein. (*c*) Sixty percent of the amino acid residues in the extended region are hydrophobic, in contrast to 38% in the lipoprotein. (*d*) The distribution of these hydrophobic amino acids along the peptide chain is completely different from their periodical distribution in the lipoprotein (15).

The extended region can be divided into four separate sections on the basis of the amino acid arrangement: The first section (section S-1), consisting of five amino acid residues, is hydrophilic and positively charged because of two lysine and one threonine residues. The second section (Section I-1) is formed by nine hydrophobic amino acid residues and contains two very similar hydrophobic sequences Leu-Val-Leu-Gly and Val-Ile-Leu-Gly. The third section (section S-2) is the second hydrophilic part, consisting of one serine and one threonine residue. The fourth section (section I-2) is the second hydrophobic segment and is formed by four amino acids; Leu-Leu-Ala-Gly, which is a hydrophobic sequence similar to the sequences found in section I-1. This section is followed by the lipoprotein. Each section of the extra region may play an important role in translocating the lipoprotein produced in the cytoplasm to the outer membrane. For instance, section S-1 is probably involved in leading the initial attachment of the prolipoprotein to the membrane by forming ionic interactions between the positively charged section S-1 and the negatively charged surface of the membrane. Following section S-1, a long stretch of hydrophobic section I-1 is produced and is most likely inserted inside the membrane. The following hydrophilic section S-2 may play a role in keeping the carboxyl-terminal end of section I-1 on the surface of the membrane, or may be inserted into the membrane together with the following hydrophobic section I-2. Fig. 7 shows a schematic diagram of the possible first step in prolipoprotein attachment to the cytoplasmic membrane, where section I-1 is bent at its center and

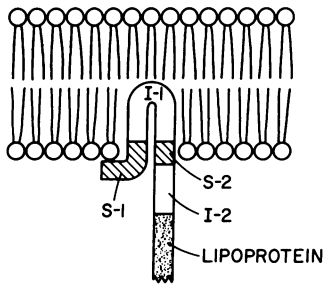

FIG. 7. A possible arrangement of the extended region in the lipid bilayer. S-1, Met-Lys-Ala-Thr-Lys; I-1, Leu-Val-Leu-Gly-Ala-Val-Ile-Leu-Gly; S-2, Ser-Thr; I-2, Leu-Leu-Ala-Gly.

penetrates only half way through the membrane. It is difficult at present to formulate the mechanism by which the peptide extension promotes the attachment of the prolipoprotein to the surface of the cytoplasmic membrane and the translocation across the cytoplasmic membrane into the outer membrane. One hypothetical model for protein translocation across a membrane (signal hypothesis) has been proposed (16). It is possible that many enzyme reactions are coupled with the lipoprotein translocation process; these may include the processing enzyme that cleaves the extended region of the prolipoprotein, the enzyme that links a fatty acid to the amino terminus of the lipoprotein (17), the enzyme required for the formation of glycerylcysteine (17), and the enzyme for the conversion of the lipoprotein free form to the bound form (18).

The amino acid composition of the extended region of the prolipoprotein has been determined using the peak I protein from toluene-treated cells (11). The amino acid composition is in a very good agreement with that of the present cell-free product. Furthermore, the peak I protein was found to have exactly the same sequence as the prolipoprotein produced in the cell-free system. This indicates that the treatment with toluene causes inhibition or inactivation of the processing enzyme(s) for the prolipoprotein.

The amino terminus of the cell-free product was found to be methionine, a major fraction of which appears to be formylated, because without trichloroacetic acid treatment of the product the recovery of [^{3}H]lysine at the second position was only 5%, in contrast to 31% with trichloroacetic acid treatment for 30 min. When the product was treated with trichloroacetic acid for 1 hr, the recovery increased to 38%. In the case of the peak I protein from toluene-treated cells a part of the product appeared to be unformylated, because the results in Fig. 6 were obtained without trichloroacetic acid treatment.

It is of interest to examine whether the existence of precursors for membrane proteins is a general feature. In this regard it should be noticed that some secretory proteins in eukaryotic systems are also produced from their precursors, which have 16 to 20 additional amino acid residues at their amino termini (16, 19, 20).

This work was supported by grants from the U.S. Public Health Service (GM 19043) and the American Cancer Society (BC-67C).

1. Braun, B. (1975) *Biochim. Biophys. Acta* **415**, 335–337.
2. Inouye, M. (1975) in *Membrane Biogenesis*, ed. Tzagoloff, A. (Plenum Publ. Corp., New York), pp. 351–391.
3. Halegoua, S., Hirashima, A. & Inouye, M. (1976) *J. Bacteriol.* **126**, 183–191.
4. Hirashima, A. & Inouye, M. (1975) *Eur. J. Biochem.* **60**, 395–398.
5. Hirashima, A. & Inouye, M. (1973) *Nature* **242**, 405–407.

6. Hirashima, A., Childs, G. & Inouye, M. (1973) *J. Mol. Biol.* **119,** 373–389.
7. Hirashima, A., Wang, S. & Inouye, M. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 4149–4153.
8. Wang, S., Marcu, K. B. & Inouye, M. (1976) *Biochem. Biophys. Res. Commun.* **68,** 1194–1200.
9. Takeishi, K., Yasumura, M., Pirtle, R. & Inouye, M. (1976) *J. Biol. Chem.* **251,** 6259–6266.
10. Halegoua, S., Hirashima, A., Sekizawa, J. & Inouye, M. (1976) *Eur. J. Biochem.* **69,** 163–167.
11. Halegoua, S., Sekizawa, J. & Inouye, M. (1977) *J. Biol. Chem.* **252,** in press.
12. Inouye, M. & Guthrie, J. P. (1960) *Proc. Natl. Acad. Sci. USA* **64,** 957–961.
13. Braun, V. & Bosch, V. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 970–974.
14. Inouye, M. (1971) *J. Biol. Chem.* **246,** 4834–4838.
15. Inouye, M. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 2396–2400.
16. Blobel, G. & Dobberstein, R. (1975) *J. Cell Biol.* **67,** 835–851.
17. Hantke, K. & Braun, V. (1973) *Eur. J. Biochem.* **34,** 284–296.
18. Inouye, M., Shaw, J. & Shen, C. (1972) *J. Biol. Chem.* **247,** 8154–8159.
19. Devillers-Thiery, A., Kindt, T., Scheele, G. & Blobel, G. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 5016–5020.
20. Schechter, I., McKean, D. J., Guyer, R. & Terry, W. (1975) *Science* **188,** 160–162.