

## Repetitive Sequences in the Murein-Lipoprotein of the Cell Wall of *Escherichia coli*

(covalent lipid/peptidoglycan/mucopeptide)

V. BRAUN\* AND V. BOSCH\*

Max-Planck-Institut für Biologie, Tübingen, West Germany

Communicated by Eugene P. Kennedy, February 3, 1972

**ABSTRACT** The amino-acid sequence of the murein-lipoprotein of the *Escherichia coli* cell wall is presented. This protein is covalently bound to a lipid component as well as to the murein (peptidoglycan, mucopeptide). The sequence is also highly repetitive. At the N-terminal portion, there are three adjacent almost identical sequences, indicating repeated duplication of a gene coding originally for 15 amino acids. The C-terminal part of the polypeptide chain is more variable but still shows striking homology when certain sequence gaps are introduced. The lipid is bound to the N-terminal serine of the dipeptide (Ser-Ser) that extends from the repetitive sequence. At the C-terminal end where the murein is bound, a tripeptide extends from the repetitive portion. Here there are several basic amino acids and the only aromatic amino acid in the lipoprotein. The sequence is Lys-Tyr-Arg-Lys. The linkage to the murein is formed between the  $\epsilon$ -amino group of the C-terminal lysine and the carboxyl group of the optical L-center of meso-diaminopimelic acid. The polypeptide chain is composed of 57 amino acids and lacks glycine, proline, cysteine, phenylalanine, histidine, and tryptophan. 63% of the amino acids are hydrophilic, but because of the covalently linked lipid this structural membrane protein has very hydrophobic properties.

The cell wall of *Escherichia coli* is a favorable system for obtaining insight into the molecular organization of biological membranes by chemical means. It contains the so-called rigid layer, which is linked together by covalent bonds. The structure of this layer is therefore preserved even under rigorous isolation conditions during which other membranes are dissociated into their individual components with irreversible loss of their *in vivo* arrangement.

The basic structure of the rigid layer is the murein (peptidoglycan, mucopeptide), which consists of polysaccharide chains crosslinked by short peptide bridges (1). It forms a net that encloses the whole cell. From our previous studies (2-5) we concluded that, depending upon the size of the cell, about  $10^5$  lipoprotein molecules are covalently linked to the murein of one cell. We further hypothesized that these lipoprotein molecules may represent the covalently anchored core of subunits of which the cell wall is built. The other components of the cell wall could be organized around the lipoprotein so that the pattern of lipoprotein molecules on the murein would reflect the overall organization of the cell wall. To support this model, we studied the structure of the

lipoprotein further. Of special interest were the attachment sites of the covalently linked lipid and the murein to the polypeptide chain. We report here the complete amino-acid sequence of the protein, the attachment site of the lipid, and the mode of binding of the lipoprotein on the murein. Experimental details and the complete list of sequenced peptides will be published elsewhere.

### METHODS

#### Experimental procedure

The following preparations and techniques have been described (2-4): isolation of murein-lipoprotein complex from *E. coli* B, grown to  $5 \times 10^8$  cells/ml in glucose-minimal medium (6) or Difco M3; digestion of murein-lipoprotein with hen egg white lysozyme (EC 3.2.1.17) and isolation of lipoprotein either by precipitation at pH 3.7 or chromatography on Sephadex G-75; preparation of murein from a digest of murein-lipoprotein complex with Pronase (EC 3.4.4), preparation of a lipopeptide (7) from such a digest, partial acid hydrolysis of murein, isolation of a murein repeating unit or a portion of the murein peptide side chain containing in addition 1 mol of lysine and arginine (3, 4); amino-acid and amino-sugar analysis on a Unichrom amino-acid analyzer (Beckman, Munich).

#### Preparation of peptides

The peptides used for determination of the amino-acid sequence were obtained from digests of the murein-lipoprotein complex with Pronase, thermolysin, subtilisin (EC 3.4.4.16), or trypsin (EC 3.4.4.4). For example, 3.2 g murein-lipoprotein complex in 300 ml of 0.01 M Tris-HCl buffer were digested for 2 hr at 60° with Pronase (pH 7.5) or thermolysin (pH 8). Insoluble murein was removed by a 2-hr centrifugation at  $37,000 \times g$ , and then lipopeptide was precipitated by adjusting the pH of the clear supernatant to 3.2. The soluble peptides were chromatographed.

Special treatments were necessary to cleave the insoluble lipoprotein with trypsin. Lipoprotein still bound to the murein, lysozyme-released lipoprotein dispersed by sonification, lipoprotein treated with maleic anhydride (8), or lipoprotein dispersed in 8 M urea and then diluted to 0.1 M urea were incubated with trypsin [protein:enzyme ratio 50:1 (w/w)] for 15 hr at 37° at pH 8-9. Hydrolysis with trypsin was incomplete but fortunately yielded different peptide spectra. The soluble peptides were separated into groups by column chromatography on Dowex 50  $\times$  2 with the following logarithmic gradient: starting with 0.1 M

Abbreviations: Dpm, meso-2,6-diaminopimelic acid; MurNAc, N-acetylmuramic acid.

\* Present address: Max-Planck-Institut für Molekulare Genetik, 1000 Berlin 33, Ihnestr. 63-73, Germany.

pyridine adjusted to pH 2 with formic acid, the molarity and pH of the pyridine were raised to final values of 2 M and pH 5.0. The peptides were further purified on phosphocellulose with a gradient as used on Dowex 50 × 2, or on Dowex 1 × 2 (9) with an *N*-ethylmorpholine-pyridine-acetate gradient. The purity of the peptides was checked by paper chromatography with butanol-acetic acid-water (4:1:5, upper phase), or, for basic peptides, methanol-water-pyridine-12 N HCl (80:17.5:10:2.5). When only a single spot was observed, the purity was evaluated by determination of the amino-acid composition with the amino-acid analyzer. Peptides were usually hydrolyzed in 5.7 N HCl for 20 hr at 105°. However, peptides containing amino sugars of the murein were hydrolyzed for 15 hr in 4 N HCl at 105°. The amino-acid sequence of peptides was mostly determined by the subtractive Edman procedure (10, 11), but the sequences of some were, in addition, directly determined from the cleaved phenylthiohydantoin derivatives of the N-terminal amino acids by chromatography on starch-paper (12) or on silica gel thin-layer plates (13). In addition, sequential digestions of peptides by carboxypeptidases A and B (14) (EC 3.4.2.1; EC 3.4.2.2) and aminopeptidase M (15) were used as a further tool for sequence determination. Asparagine and glutamine were mostly determined in this way but if these amino acids were C-terminal they were also determined by amino-acid analysis after the preceding step of Edman degradation without acid hydrolysis. The free amino groups at the murein-lipoprotein attachment site were also determined by dinitrophenylation (16) of suitable degradation products (see *Results*). The ether-soluble and water-soluble dinitrophenyl derivatives were chromatographed on silica gel thin-layer plates with *n*-propanol-35% ammonia (70:30).

## RESULTS AND DISCUSSION

### Repetitive amino-acid sequence

The lipoprotein, both when bound to the murein and when released by degradation of murein-lipoprotein by lysozyme, is insoluble and therefore difficult to digest. Unspecific proteases such as Pronase, thermolysin, or subtilisin had to be used. These enzymes yielded many overlapping peptides so that every peptide bond was established in at least two but in most cases in four to five different peptides. However, these overlaps proved to be important since we could then determine with certainty the repetitive design of the sequence

as shown in Fig. 1. The sequence is presented to show maximal homology between peptides *within* the polypeptide chain.

Such a degree of repetition has not previously been observed (see, e.g., discussion in refs. 18 and 19). Starting from the third amino acid, a sequence of 14 amino acids is nearly identically repeated by the next sequence of 15 amino acids followed again by 7 amino acids. The structural gene probably coded originally for 15 amino acids; it doubled and the triplet coding for lysine or arginine at the C-terminal end was deleted. Then, only half of the original gene was fused with the doubled gene. Now a more variable part of the protein starts whose descendance from the original gene is obvious when all the following peptides are aligned so that they end either with lysine or arginine. The observed homologies are still striking, but several amino-acid deletions (17) have to be introduced. The repetitive N-terminal portion of the polypeptide chain, which has remained essentially constant, may reflect a high selection pressure during evolution that allowed only few, and then very conservative, amino acid replacements (Ile/Val; Gln/Asn/Ser; Ala/Ser; Ala/Thr; Lys/Arg). It may also mean, since the lipoprotein almost certainly plays mainly a structural role, that the protein did not have to adapt considerably to changing environmental conditions in order to fulfill its function. Otherwise many amino-acid replacements in the original part and the duplicated portions would have occurred, and no such clear-cut duplications within one gene could be observed today. The sequence from Asn (3) to Arg (56) could be established by ordinary sequence analysis. However, determination of the rest of the sequence involved elucidation of the attachment sites of the lipid and the murein.

### Lipid attachment site

Both murein and lipid are covalently bound to the protein of the murein-lipoprotein complex. Lipid is found with the protein after the release of the protein from the murein by lysozyme degradation. This lipid could not be removed upon purification by gel chromatography in 2% sodium dodecyl sulfate (SDS), gel electrophoresis in 1% SDS, or by extraction with mixtures of various organic solvents of different polarities. Lipopeptides, released by Pronase digestion of the lipoprotein, always contained all the lipid originally present in the lipoprotein. They were soluble in methanol and could be precipitated by addition of water. By combination of several techniques, mainly partial alkali and acid hydrolyses

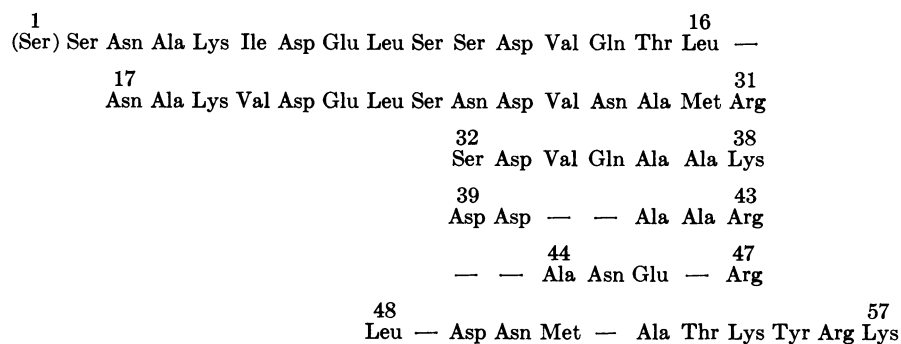


FIG. 1. Amino-acid sequence of the murein-lipoprotein of the *E. coli* cell wall. Starting at the N-terminal end, the numbers denote the positions in the sequence. The polypeptide chain is composed of 57 amino acids. *Dashes* represent hypothetical sequence gaps (17) introduced in order to obtain maximal homology between the peptides. The N-terminal serine in parentheses has not been proven by sequence determination.



nized. Since similar amounts of arginine were released from both peptides, the length of the amino-acid side chain at the  $\epsilon$ -amino group of lysine is apparently without major influence.

In summary, the attachment of the polypeptide chain at the N-terminal end to a lipid and at the C-terminal end to the murein is a unique feature that is manifested in the structural peculiarities of the lipoprotein. The two serine residues at the N-terminal end extend from the repetitive part of the polypeptide chain and the same is true for the striking concentration of basic amino acids at the C-terminal end. The only aromatic amino acid in the whole polypeptide is also situated there. The sequence Lys-Tyr-Arg-Lys could thus serve as a recognition site for the enzyme that attaches the lipoprotein molecules to the murein. A schematic drawing of the lipid and murein attachment sites is presented in Fig. 2.

There is another interesting aspect of this sequence. According to Perutz, Kendrew, and Watson (22), in hemoglobins and myoglobin, nonpolar sites tend to repeat along  $\alpha$ -helical segments at regular intervals of about 3.6 residues, making one side of the helix nonpolar which is at the interior of the globin chain. Starting at the third amino acid (Asn) of the lipoprotein (see Fig. 1), where the repetitive sequence also commences, nonpolar residues occur constantly along the whole sequence at intervals of 4 and 3 residues, respectively, in a consistently alternating set. If an  $\alpha$ -helical conformation for the lipoprotein could be proven, and our preliminary measurements of the circular dichroism show indeed a very high  $\alpha$ -helical content, one side of the helix would be polar and the other nonpolar. If this would be a single helix about 50 amino acids long, then the protein would also have two differently polar faces. On the other hand if the protein is folded, it would be of considerable interest to know whether the hydrophobic side chains are buried in the interior of the protein, as is the case with soluble proteins, or whether they interact primarily with the surrounding, presumably mostly hydrophobic, components of the cell wall. A relatively high  $\alpha$ -helical content was claimed for other proteins in various membranes [see review (23) and a critical evaluation (24)].

The polypeptide chain is composed mainly of hydrophilic amino acids. However, with the lipid attached the protein becomes distinctly hydrophobic. The lipid at the N-terminal end of the lipoprotein could extend quite far into the cell wall, perhaps even into a lipid bilayer, and could interact with other lipids, hydrophobic regions of proteins, or with the lipid A of lipopolysaccharide (25). When the polypeptide chain forms one single  $\alpha$ -helix it could have a length of 85 Å, and in a more extended structure the protein could span the width of the cell wall (thickness, 120–140 Å) (26, 27). Immunological studies with wild-type cells and mutants lacking surface structures to a different extent (e.g., rough strains) yielded promising results for a more detailed localization of the lipoprotein molecules in the cell wall (Mayer, H. and Braun, V., unpublished).

The evidence is now conclusive that the protein found together with the murein (28) is a constitutive part of the rigid layer and is not a contaminant from various cell-wall proteins. Our previous concept of individual lipoprotein molecules of identical structure containing covalently bound lipid and being linked by a covalent bond on average to every 10th–12th repeating unit of the murein is thus confirmed.

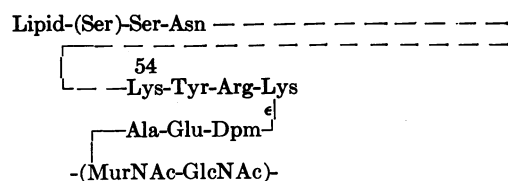


FIG. 2. Lipid and murein attachment sites on the polypeptide chain. The dashed line represents the amino-acid sequence between the attachment sites that is omitted in the schematic illustration. Dpm, meso-2,6-diaminopimelic acid; MurNAc, N-acetylmuramic acid; GlcNAc, N-acetylglucosamine.

We express our deep gratitude to Dr. U. Henning for his encouraging and generous support. Part of this work was supported by the Deutsche Forschungsgemeinschaft.

- Weidel, W. & Pelzer, H. (1964) "Bagshaped macromolecules—a new outlook on bacterial cell walls," *Advan. Enzymol.* **26**, 193–232.
- Braun, V. & Rehn, K. (1969) "Chemical characterisation, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the *E. coli* cell wall. The specific effect of trypsin on the membrane structure," *Eur. J. Biochem.* **10**, 426–438.
- Braun, V. & Sieglin, U. (1970) "The covalent murein-lipoprotein structure of the *Escherichia coli* cell wall. The attachment site of the lipoprotein on the murein," *Eur. J. Biochem.* **13**, 336–347.
- Braun, V. & Wolff, H. (1970) "The murein-lipoprotein linkage in the cell wall of *Escherichia coli*," *Eur. J. Biochem.* **14**, 387–391.
- Braun, V., Rehn, K. & Wolff, H. (1970) "Supramolecular structure of the rigid layer of the cell wall of *Salmonella*, *Serratia*, *Proteus*, and *Pseudomonas fluorescens*. Number of lipoprotein molecules in a membrane layer," *Biochemistry* **9**, 5041–5049.
- Vogel, H. J. & Bonner, D. M. (1956) "Acetylornithase of *Escherichia coli*, partial purification and some properties," *J. Biol. Chem.* **218**, 97–106.
- Braun, V. & Hantke, K. (1972) "Evidence for a covalent bond between lipid and a membrane protein (rigid layer of the cell wall of *E. coli*)," in 19th Proceedings *Protides in Biological Fluids*, ed. Peeters, H. (Academic Press, New York and London), pp. 221–224.
- Butler, P. J. G., Harris, J. I., Hartley, B. S. & Leberman, R. (1967) "Reversible blocking of peptide amino groups by maleic anhydride," *Biochem. J.* **103**, 781–791.
- Rudloff, V. & Braunitzer, G. (1961) "Ein Verfahren zur präparativen Gewinnung natürlicher Peptide. Die Isolierung der tryptischen Spaltprodukte von Humanhämoglobin A über Dowex 1x2 unter Verwendung ninhydrin-negativer flüchtiger Puffer," *Z. Physiol. Chem.* **323**, 129–144.
- Edman, P. & Begg, G. (1967) "A protein sequenator," *Eur. J. Biochem.* **1**, 80–91.
- Konigsberg, W. (1967) "Subtractive Edman degradation" in *Methods in Enzymology*, ed. Hirs, C. H. W. (Academic Press, New York and London), Vol. XI, pp. 461–469.
- Schroeder, W. A. (1967) "Degradation of peptides by the Edman method with direct identification of the PTH-amino acid," in *Methods in Enzymology*, ed. Hirs, C. H. W. (Academic Press, New York and London), Vol. XI, pp. 445–461.
- Sjöquist, J. (1967) "Thin layer chromatography of PTH-amino acids," *Anal. Biochem.* **18**, 264–269.
- Ambler, R. P. (1967) "Carboxypeptidases A and B," in *Methods in Enzymology*, ed. Hirs, C. H. W. (Academic Press, New York and London), Vol. XI, pp. 436–445.
- Pfeiderer, G. & Celliers, P. G. (1963) "Isolierung einer Aminopeptidase aus Nierenpartikeln," *Biochem. Z.* **339**, 186–189.
- Sanger, F. (1952) "The arrangement of amino acids in proteins," in *Advances in Protein Chemistry*, eds. Anson,

- M. L., Bailey, K. & Edsall, J. T. (Academic Press, New York), Vol. 7, pp. 1-67.
17. Braunitzer, G., Hilschmann, N., Rudloff, V., Hilse, K., Liebold, B. & Müller, R. (1961) "The haemoglobin particles. Chemical and genetic aspects of their structure," *Nature* **190**, 480-482.
  18. Dickerson, R. E. (1971) "Sequence and structure homologies in bacterial and mammalian type cytochromes," *J. Mol. Biol.* **57**, 1-15.
  19. Smith, E. L. (1970) "Evolution of enzymes," in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York and London), Vol. I, pp. 267-339.
  20. Hantke, K. (1971) "Evidence for a covalent bond between lipid and protein in the murein-lipoprotein of *E. coli*," *Abstract in Lunteren Lectures on Molecular Genetics, The Bacterial Cell Envelope*, Section 1, pp. 1-2.
  21. Folk, J. E., "Carboxypeptidase B," in *The Enzymes*, ed. Boyer, P. D. (Academic Press, New York and London), Vol. 3, pp. 57-79.
  22. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) "Structure and function of haemoglobin. Some Relations between polypeptide chain configuration and amino acid sequence," *J. Mol. Biol.* **13**, 669-678.
  23. Wallach, D. F. H. & Gordon, A. (1968) "Lipid protein interactions in cellular membranes," *Fed. Proc.* **27**, 1263-1267.
  24. Gordon, D. J. & Holzwarth, G. (1971) "Optical activity of membrane suspensions: calculation of artifacts by Mie scattering theory," *Proc. Nat. Acad. Sci. USA* **68**, 2365-2369.
  25. Westphal, O. & Lüderitz, O. (1954) "Chemische Erforschung von Lipopolysacchariden gramnegativer Bakterien," *Angew. Chem.* **66**, 407-417.
  26. Murray, R. G. E., Steed, P. & Elson, H. E. (1965) "The location of the mucopeptide in sections of the cell wall of *Escherichia coli* and other gramnegative bacteria," *Can. J. Microbiol.* **11**, 547-560.
  27. de Petris, S. (1967) "Ultrastructure of the cell wall of *Escherichia coli* and chemical nature of its constituent layers," *J. Ultrastruct. Res.* **19**, 45-83.
  28. Martin, H. H. & Frank, H. (1962) "Quantitative Bausteinanalyse der Stützmembran in der Zellwand von *Escherichia coli* B," *Z. Naturforsch. B* **17**, 190-196.