Biochem. J. (1982) **203**, 33–43
Printed in Great Britain

Primary structure of major outer-membrane protein I (*ompF* protein, porin) of *Escherichia coli* B/r

Robert CHEN, Cornelia KRÄMER, Waltraud SCHMIDMAYR, Ursula CHEN-SCHMEISSER and Ulf HENNING
Max-Planck-Institut für Biologie, Correnstrasse 38, D-7400 Tübingen, Federal Republic of Germany

(*Received 15 June 1981/Accepted 17 November 1981*)

In the outer membrane of Gram-negative bacteria hydrophilic pores exist, allowing the diffusion of various low-molecular-weight solutes. These pores are formed by proteins, the porins. In a preliminary communication [Chen, Krämer, Schmidmayr & Henning (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5014–5017] we presented the primary structure of one of these porins, the 340-amino-acid-residue protein I (*ompF* protein) from *Escherichia coli* B/r. In the present paper we give the experimental evidence for this sequence. Two tryptophan positions, one valine position, two aspartic acid positions and nine out of 82 amide determinations have been corrected. To aid further studies on this class of transmembrane proteins, the isolation of most of the constituent peptides is documented.

The outer membrane of Gram-negative bacteria [and apparently also that of mitochondria (Zalman *et al.*, 1980)] constitutes a molecular sieve allowing the diffusion of various low-molecular-weight hydrophilic solutes. Diffusion proceeds through channels composed of proteins called porins (Nakae, 1976; Nikaido, 1979*a*,*b*). In *Escherichia coli* K-12 so far five different polypeptides are known or highly suspected to form such pores: the *lamB* protein (Luckey & Nikaido, 1980; Nakae, 1979), the *tsx* protein (Hantke, 1976) and a group of three proteins [for these and further references see Nikaido (1979*b*), Pugsley & Schnaitman (1978), Pugsley *et al.* (1980) and Tomassen & Lugtenberg (1980)] that are very closely related structurally to each other (Ichihara & Mizushima, 1978; Gamon *et al.*, 1978) and to the protein that is the subject of the present paper. Depending on the type of porin, these channels appear to exhibit a certain degree of specificity for the substances they allow to pass, either directly (Luckey & Nikaido, 1980) or in addition by a gating action mediated by a periplasmic protein (Heuzenroeder & Reeves, 1980). Porins are also known to form non-specific channels and to exclude solutes essentially only on the basis of molecular weight (Nikaido, 1979*a*,*b*; Schindler & Rosenbusch, 1978; Benz *et al.*, 1978, 1979).

Abbreviations used: DABITC, 4-*NN*-dimethylaminoazobenzene 4′-isothiocyanate; DABTH, 4-*NN*-dimethylaminoazobenzene 4′-thiohydantoin; dansyl, 5-dimethylaminonaphthalene-1-sulphonyl.

We have presented, in a preliminary communication, the primary structure of one of the latter porins, the *ompF* protein (protein I) from *E. coli* B/r (Chen *et al.*, 1979), and in the present paper we give the experimental evidence for the sequence. Two tryptophan positions, one valine position, two aspartic acid positions and several amide determinations have been corrected. (These corrections do not invalidate any of the conclusions drawn in the preliminary communication.) The protein undoubtedly will be the subject of further intensive studies, particularly concerning its structure as a pore, its arrangement in the outer membrane and its relatives mentioned above. To aid such studies we document the isolation of most of the constituent peptides.

Experimental

Materials

Citraconic anhydride was purchased from Pierce (Rotterdam, The Netherlands). Fluorescamine and dansyl chloride were products of Serva (Heidelberg, Germany). Trifluoroacetic acid, heptafluorobutyric acid and 4-*NN*-dimethylaminoazobenzene 4′-isothiocyanate were from Fluka (Buchs, Switzerland). Micro polyamide sheets were from Schleicher und Schüll (Dassel, Germany), and cellulose thin-layer plates (Polygram Cel 300) from Macherey und Nagel (Düren, Germany). Trypsin treated with 1-chloro-4-phenyl-3-tosylamidobutan-2-one ('TPCK') was obtained from Merck (Darmstadt,

0306-3275/82/040033-11$01.50/1 © 1982 The Biochemical Society

Germany), thermolysin was from Serva (Heidelberg, Germany) and *Staphylococcus aureus* proteinase SV8 was from Miles (Elkhart, IN, U.S.A.). Protein I was isolated as described by Hindennach & Henning (1975).

Preparation of CNBr-cleavage fragments

Protein I (140 nmol) was treated with CNBr in 70% (v/v) trifluoroacetic acid for 24 h at room temperature in the dark. The mixture was freeze-dried, and the fragments were allowed to react with citraconic anhydride and separated on a Sephadex G-100 column (2.5 cm × 150 cm) in 50 mM-Tris/HCl buffer, pH 8.1, containing 8 M-urea, as described previously (Chen *et al.*, 1978). Deacylation of the fragments was achieved by treatment with 50% (v/v) acetic acid for 4 h at room temperature.

For the cleavage at methionine and tryptophan residues (Ozols & Gerard, 1977) the protein sample was dissolved in 88% (v/v) formic acid/heptafluorobutyric acid (1:1, v/v) and treated with CNBr. The fragments were purified on a Sephadex G-75 (superfine grade) column (2.5 cm × 150 cm) in 50% (v/v) acetic acid and at a flow rate of 4.5 ml/h (Chen *et al.*, 1980).

Partial tryptic digestion of CNBr-cleavage fragments

Citraconylated fragments (50 nmol) were digested with trypsin at an enzyme/substrate ratio of 1:200 (by wt.) in 0.1 M-*N*-methylmorpholine/acetate buffer, pH 8.1, for 2 h at 37°C. The peptides were then deacylated [50% (v/v) acetic acid for 4 h at room temperature] and purified on a Sephadex G-50 (superfine grade) column (2.5 cm × 150 cm) with 50% (v/v) acetic acid as solvent.

Enzymic digestions and isolation of peptides

The CNBr-cleavage fragments (20 nmol) were digested with trypsin, thermolysin or proteinase SV8 in 1 ml of 0.1 M-*N*-methylmorpholine/acetate buffer, pH 8.1, for 20 h at 37°C. The enzyme/substrate ratios of trypsin and thermolysin were 1:100, whereas for proteinase SV8 it was 1:20. If necessary, the peptides were fractionated on a Sephadex G-50 (superfine grade) column (2.5 cm × 150 cm) or a Sephadex G-25 column (1 cm × 150 cm) in 36% (v/v) acetic acid before they were 'fingerprinted' on cellulose thin-layer plates. In the first dimension electrophoresis was in 0.25 M-pyridine and 5% (v/v) acetone adjusted to pH 5.4 with acetic acid. In the second dimension the chromatogram was developed in pyridine/butan-1-ol/acetic acid/water (10:15:3:12, by vol.). In one experiment (Fig. 11) electrophoresis was run in 40 mM-pyridine / acetate buffer, pH 3.5, and chromatography was in butan-1-ol / acetic acid / pyridine/water (1:2:8:8, by vol.). The peptides were eluted from the cellulose with 50% (v/v) acetic acid.

Amino acid analyses and sequence determinations

A 2 nmol sample of peptide was hydrolysed in 100 µl of 6 M-HCl in the presence of 0.02% 2-mercaptoethanol for 20 h at 110°C. The hydrolysate was dried *in vacuo* and was taken up in 50 µl of 0.2 M-sodium citrate buffer, pH 2.2. Portions of the sample were analysed in a Durrum D-500 amino acid analyser at the sensitivity of 0.5 *A* unit by the procedure of Spackman *et al.* (1958). Tryptophan was detected by the reddening of tryptophan-containing peptides on the peptide 'map' when it was treated with Ehrlich's reagent spray.

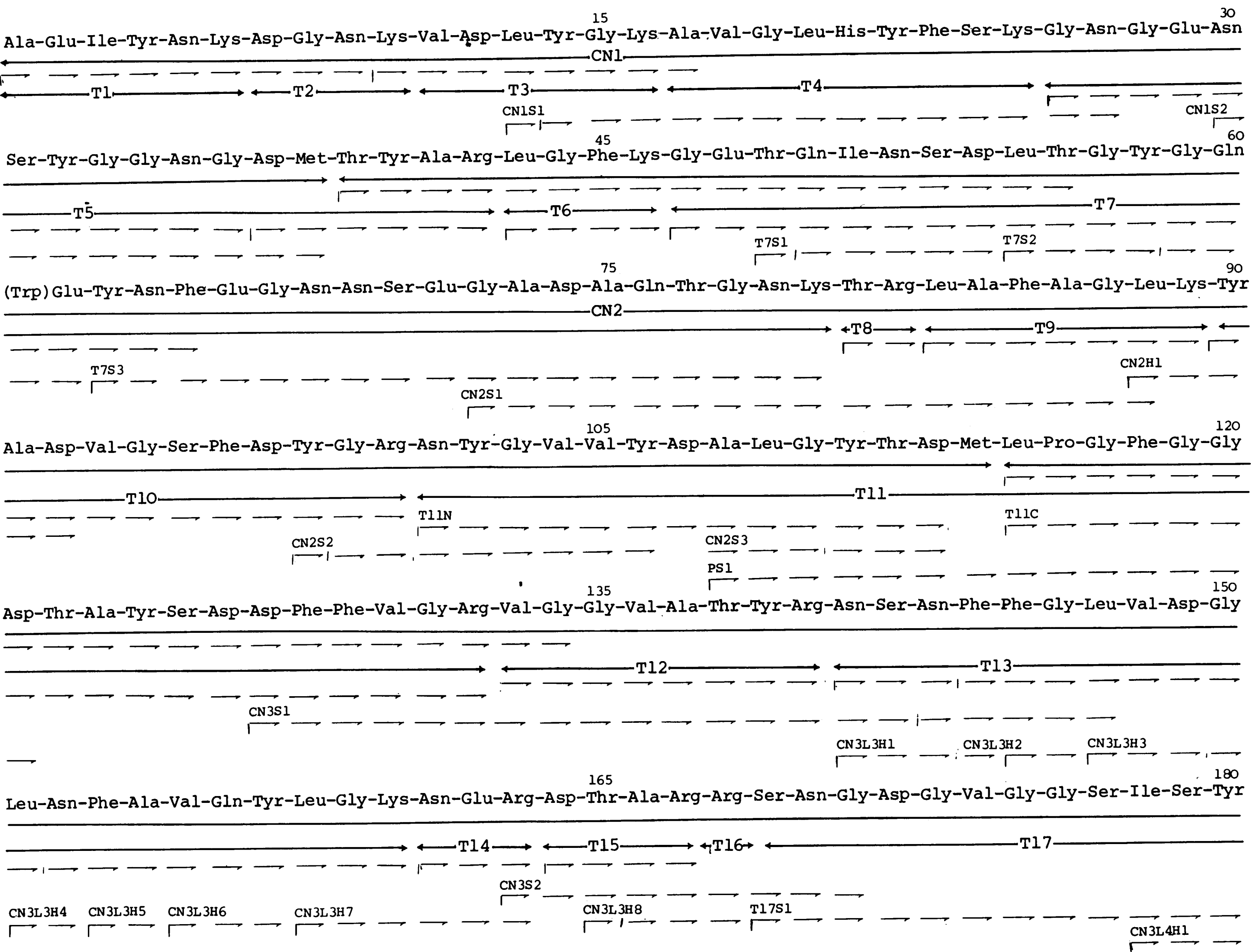
For the determination of the amino acid sequences 2–5 nmol of peptide was degraded stepwise by a modification of the micro dansyl–Edman procedure (Gray & Hartley, 1963; Chen, 1976). In the present work the following additional modifications were introduced. To 2–5 nmol of the peptide, first dissolved in 50 µl of double-distilled water, 25 µl of pyridine was added. Then 50 µl of phenyl isothiocyanate solution (phenyl isothiocyanate/pyridine, 1:15, v/v) was pipetted into the sample, initiating the coupling reaction. The cleavage was in 50 µl of trifluoroacetic acid. The extraction step with n-butyl acetate was repeated once with 75 µl of the solvent. Dansylation of a peptide sample was in 5 µl of 0.2 M-*N*-methylmorpholine/acetate buffer, pH 8.1, instead of 0.2 M-$NaHCO_3$.

Assignment of amides

Each peptide that was found to contain aspartic acid/asparagine or glutamic acid/glutamine at a given position was degraded up to the residue before that position. In the next cycle 4-*NN*-dimethylaminoazobenzene 4′-isothiocyanate (Chang & Creaser, 1976) instead of phenyl isothiocyanate was coupled to the amino acid, since DABTH-Glu/DABTH-Gln as well as DABTH-Asp/DABTH-Asn

Fig. 1. *Amino acid sequence of protein I (OmpF protein) of E. coli B/r*

Key: CN, CNBr-cleavage fragments; T, tryptic peptides; S, proteinase SV8 peptides; H, thermolytic peptides; LB or L, lysine-blocked (tryptic) peptides. Arrows (⇁) indicate the cycles of micro dansyl–Edman degradations. T24* and T27*–T28: trypsin to a certain degree cleaved at the -Tyr-Gln- and -Tyr-Tyr- sites respectively, and the corresponding peptides could be sequenced as indicated. The lysine-blocked peptides CN3LB1 (residues 115–132), CN3LB2 (residues 133–140), CN3LB3 (residues 141–168), CN3LB4 (residues 169–196) and CN3LB5 (residues 197–235) have not been sequenced directly; they are therefore not included in this Figure. The tryptophan residues are in parentheses because they have not been identified directly.


15 30
Ala-Glu-Ile-Tyr-Asn-Lys-Asp-Gly-Asn-Lys-Val-Asp-Leu-Tyr-Gly-Lys-Ala-Val-Gly-Leu-His-Tyr-Phe-Ser-Lys-Gly-Asn-Gly-Glu-Asn
CN1
T1 T2 T3 T4
CN1S1 CN1S2
45 60
Ser-Tyr-Gly-Gly-Asn-Gly-Asp-Met-Thr-Tyr-Ala-Arg-Leu-Gly-Phe-Lys-Gly-Glu-Thr-Gln-Ile-Asn-Ser-Asp-Leu-Thr-Gly-Tyr-Gly-Gln
T5 T6 T7
T7S1 T7S2
75 90
(Trp)Glu-Tyr-Asn-Phe-Glu-Gly-Asn-Asn-Ser-Glu-Gly-Ala-Asp-Ala-Gln-Thr-Gly-Asn-Lys-Thr-Arg-Leu-Ala-Phe-Ala-Gly-Leu-Lys-Tyr
CN2
T8 T9
T7S3 CN2H1
CN2S1
105 120
Ala-Asp-Val-Gly-Ser-Phe-Asp-Tyr-Gly-Arg-Asn-Tyr-Gly-Val-Val-Tyr-Asp-Ala-Leu-Gly-Tyr-Thr-Asp-Met-Leu-Pro-Gly-Phe-Gly-Gly
T10 T11
T11N T11C
CN2S2 CN2S3
PS1
135 150
Asp-Thr-Ala-Tyr-Ser-Asp-Asp-Phe-Phe-Val-Gly-Arg-Val-Gly-Gly-Val-Ala-Thr-Tyr-Arg-Asn-Ser-Asn-Phe-Phe-Gly-Leu-Val-Asp-Gly
T12 T13
CN3S1
CN3L3H1 CN3L3H2 CN3L3H3
165 180
Leu-Asn-Phe-Ala-Val-Gln-Tyr-Leu-Gly-Lys-Asn-Glu-Arg-Asp-Thr-Ala-Arg-Arg-Ser-Asn-Gly-Asp-Gly-Val-Gly-Gly-Ser-Ile-Ser-Tyr
T14 T15 T16 T17
CN3S2
CN3L3H4 CN3L3H5 CN3L3H6 CN3L3H7 CN3L3H8 T17S1
CN3L4H1

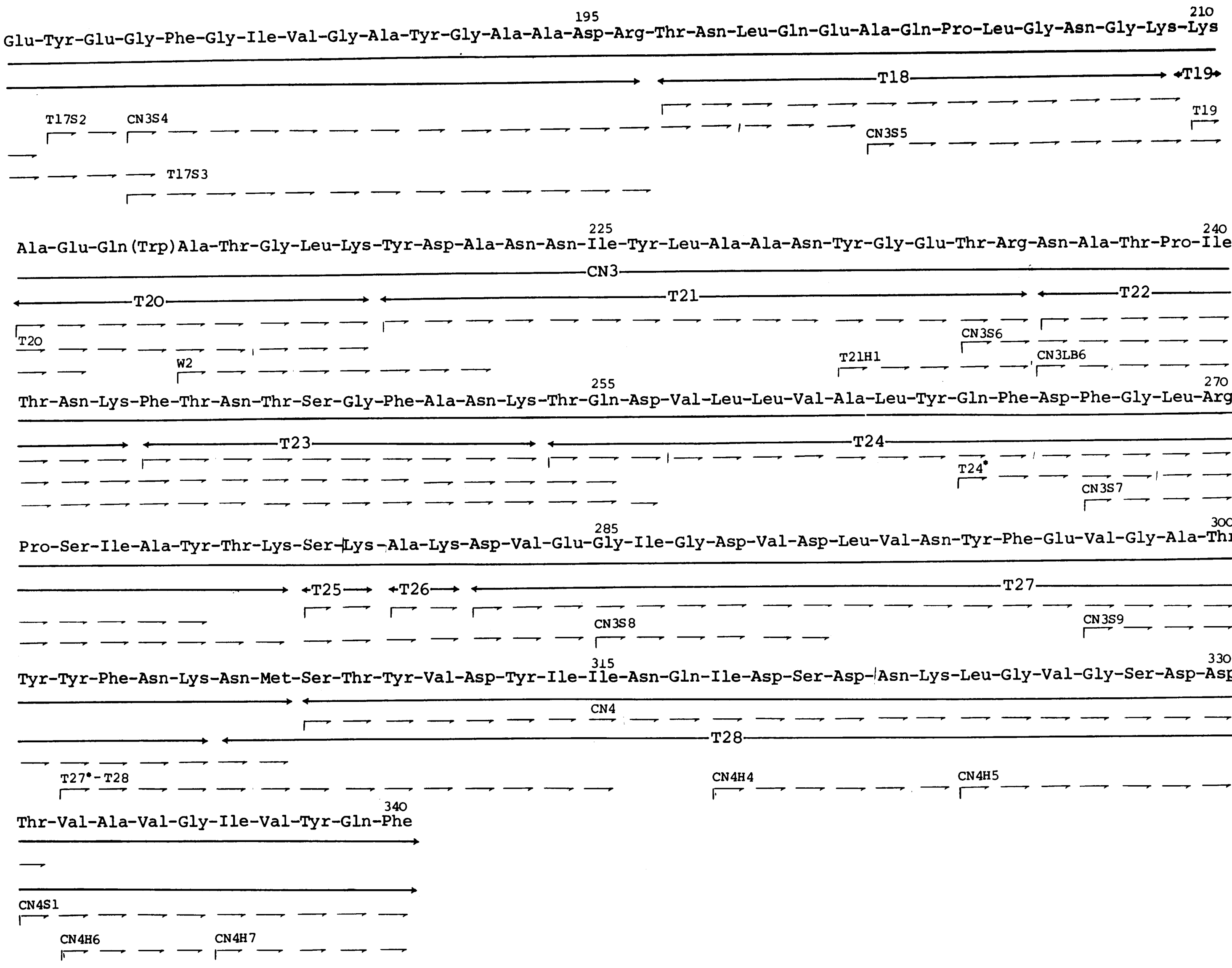

195
210
Glu-Tyr-Glu-Gly-Phe-Gly-Ile-Val-Gly-Ala-Tyr-Gly-Ala-Ala-Asp-Arg-Thr-Asn-Leu-Gln-Glu-Ala-Gln-Pro-Leu-Gly-Asn-Gly-Lys-Lys
T18
T19
T17S2
CN3S4
T19
CN3S5
T17S3
225
240
Ala-Glu-Gln(Trp)Ala-Thr-Gly-Leu-Lys-Tyr-Asp-Ala-Asn-Asn-Ile-Tyr-Leu-Ala-Ala-Asn-Tyr-Gly-Glu-Thr-Arg-Asn-Ala-Thr-Pro-Ile
CN3
T20
T21
T22
T20
CN3S6
W2
T21H1
CN3LB6
255
270
Thr-Asn-Lys-Phe-Thr-Asn-Thr-Ser-Gly-Phe-Ala-Asn-Lys-Thr-Gln-Asp-Val-Leu-Leu-Val-Ala-Leu-Tyr-Gln-Phe-Asp-Phe-Gly-Leu-Arg
T23
T24
T24*
CN3S7
285
300
Pro-Ser-Ile-Ala-Tyr-Thr-Lys-Ser-Lys-Ala-Lys-Asp-Val-Glu-Gly-Ile-Gly-Asp-Val-Asp-Leu-Val-Asn-Tyr-Phe-Glu-Val-Gly-Ala-Thr
T25
T26
T27
CN3S8
CN3S9
315
330
Tyr-Tyr-Phe-Asn-Lys-Asn-Met-Ser-Thr-Tyr-Val-Asp-Tyr-Ile-Ile-Asn-Gln-Ile-Asp-Ser-Asp-Asn-Lys-Leu-Gly-Val-Gly-Ser-Asp-Asp
CN4
T28
T27*-T28
CN4H4
CN4H5
340
Thr-Val-Ala-Val-Gly-Ile-Val-Tyr-Gln-Phe
CN4S1
CN4H6
CN4H7

Table 1. *Amino acid compositions of protein I*

	Composition (mol of residue/mol)	
	Amino acid analysis	Sequence analysis
Aspartic acid	57	27
Asparagine		30
Threonine	21	21
Serine	15	16
Glutamic acid	27	14
Glutamine		11
Proline	4	4
Glycine	47	49
Alanine	30	29
Cysteine*	—	—
Valine	22	23
Methionine	2	3
Isoleucine	12	12
Leucine	21	21
Tyrosine	29	29
Phenylalanine	20	19
Histidine	1	1
Lysine	18	18
Arginine	11	11
Tryptophan†	2	2
Total		340

* Cysteine was detected as cysteic acid.

† Tryptophan was determined after hydrolysis in 4 M-methanesulphonic acid.

can be distinguished by two-dimensional chromatography on polyamide sheets. This method was adjusted to the sensitivity of the dansyl–Edman technique by two innovations. First, the peptides were treated with a smaller amount of DABITC solution (20 μl of 0.35 mg/ml in pyridine), and, secondly, a drying step was introduced after the coupling reaction. The sample was then dissolved in 5 μl of pyridine, and 5 μl of redistilled water was added before the extraction with heptane/ethyl acetate (2:1, v/v).

Results and discussion

The primary structure (and the segments used for sequence determination) of the porin from *E. coli* B/r is shown in Fig. 1. In contrast with earlier reports (Rosenbusch, 1974; Garten & Henning, 1974), no cysteine was detected (Table 1). Our results agreed with the data of Diedrich & Schnaitman (1978). In addition, careful re-investigations indicated only two tryptophan residues, whereas three had been reported in our preliminary report (Chen *et al.*, 1979). Amino acid analysis of a protein sample hydrolysed in methanesulphonic acid supported this finding, whereas other authors had found three to five tryptophan residues (Rosenbusch, 1974; Garten & Henning, 1974). The amino acid compositions of the peptides as derived from the sequence agreed well with the amino acid analyses of the individual peptides. All peptides shown in Fig. 1 were quantitatively analysed and fit the sequence, except for tryptophan, which was not determined quantitatively within the peptides. The isolation of the protein's four CNBr-cleavage fragments (CN1–CN4) as well as their order have been described (Chen *et al.*, 1978). The amino acid sequences of the individual fragments were determined as follows.

Fragment CN1

Peptide 'mapping' of a tryptic digest of this fragment resolved five T-peptides (Fig. 2), all of which could be sequenced up to their *C*-terminal residues. The *N*-terminal sequence of fragment CN1 provided the order T1–T2–T3. Digestion of fragment CN1 with proteinase SV8 allowed the separation, by 'fingerprinting', of four peptides. The sequence of one of these, peptide CN1S1, established the order T3–T4–T5N.

Fragment CN2

Seven tryptic peptides from this fragment were separated by 'fingerprinting' (Fig. 3). The largest peptide (T7, 34 residues) remained at the origin. It could be obtained in pure form by chromatography of the tryptic digest on a Sephadex G-50 (superfine grade) column (2.5 cm × 120 cm) in 50% acetic acid. The first peak eluted contained peptide T7, which was sequenced for 19 cycles. Peptide T7 was further digested with proteinase SV8. Three peptides (T7S1, T7S2 and T7S3) were isolated by 'fingerprinting', and could be sequenced up to their *C*-terminal residues. Their order was established by the *N*-terminal sequence of peptide T7. Sequencing of fragment CN2 for 18 cycles provided the order T5–T6–T7.

Fragment CN2 was digested with proteinase SV8, and the S-peptides were fractionated on a Sephadex G-50 (superfine grade) column (2.5 cm × 120 cm) in 50% acetic acid, which resolved five peaks. Peptides CN2S1 and CN2S2 were purified by 'fingerprinting' of the material in peaks 1 and 4 respectively. The sequence of peptide CN2S1 provided the order T7–T8–T9. The results for peptide CN2S2 showed the position of peptide T11 to be *C*-terminal to peptide T10. Peptide T11 contained homoserine, and thus is the *C*-terminal tryptic peptide of fragment CN2. Finally, the overlap of peptides T9–T10 was obtained from the thermolytic peptide CN2H1, which also was recovered by 'fingerprinting'.

Fragment CN3

Tryptic peptides. This large, 193-residue, fragment gives rise to 17 tryptic peptides. Most of them could be isolated from 'fingerprints' (Fig. 4). Peptides T13, T17, T24 and T27 were not resolved

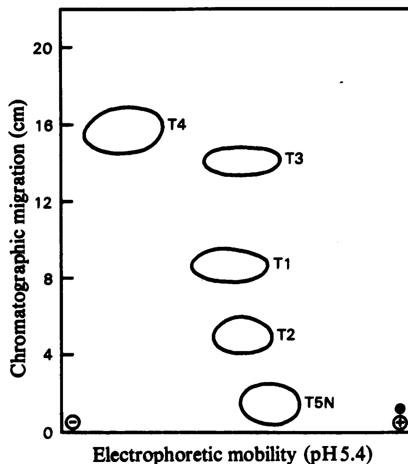


Fig. 2. *Separation of tryptic peptides from CNBr-cleavage fragment CN1*

For full experimental details for this and other Figures see the text. T5N indicates *N*-terminal moiety of peptide T5 as shown in Fig. 1. The dot indicates the point of origin on the thin-layer sheet in all pertinent Figures.

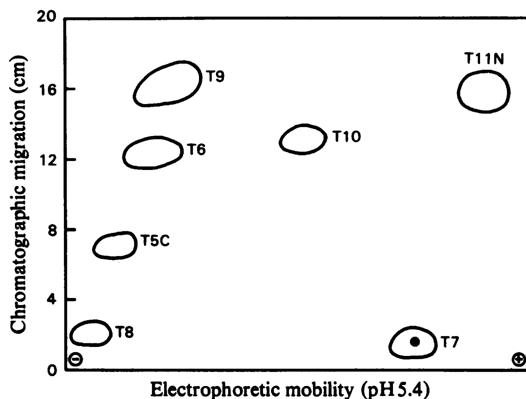


Fig. 3. *Separation of tryptic peptides from CNBr-cleavage fragment CN2*

T5C and T11N indicate the *C*-terminal and *N*-terminal moieties of peptides T5 and T11 respectively (see Fig. 1).

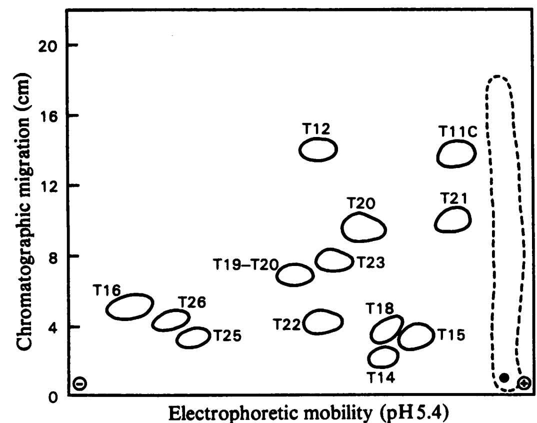


Fig. 4. *Separation of tryptic peptides from CNBr-cleavage fragment CN3*

T11C indicates the *C*-terminal moiety of peptide T11 (see Fig. 1). The Lys–Lys bond between peptides T18 and T19 was often only poorly cleaved by trypsin and therefore peptide T19–T20 arose.

by this procedure, but could be recovered from tryptic peptides obtained from lysine-blocked fragment CN3. A tryptic digest from citraconylated fragment CN3 was chromatographed on a Sephadex G-50 (superfine grade) column (2.5 cm × 120 cm) in 50% acetic acid, and seven peaks appeared (Fig. 5). The material in peak A was deacylated, digested with trypsin and freeze-dried. Peptide T27 did not dissolve in water when this digest had been freeze-dried. Therefore the soluble material was chromatographed on a Sephadex G-25 (superfine grade) column (1 cm × 120 cm) in 36% acetic acid. The first peak contained peptide T24, which was sequenced for 17 cycles (the arginine residue at position 17 of this peptide is not, as a rule, attacked by trypsin because of the following proline residue), and the seven *C*-terminal residues of peptide T24 were derived from the peptide CN3S7. The isolation of the latter and the other proteinase SV8 peptides derived from fragment CN3 is described below (see under 'Order of lysine-blocked peptides').

The water-insoluble peptide T27 was readily soluble in 30 mM-NH_4HCO_3 and was chromatographed in this solvent on a Sephadex G-25 (superfine grade) column (1 cm × 120 cm). Final isolation was achieved by 'fingerprinting'. The peptide remained at the origin, so that it was easy to isolate, and it was sequenced for 19 cycles. The five *C*-terminal residues were derived from the sequence of peptide CN3S9 spanning residues 16–24 of peptide T27. Because of contamination the precise amino acid composition of peptide T27 was not determined by amino acid analysis. To circumvent our difficulties, peptide T27 was redigested with thermolysin. From the hydrolysate seven peptides were isolated (peptide 'map' and amino acid analyses not shown) and sequenced (Fig. 6). The results eliminated any doubts about the composition of peptide T27. In contrast with our previous report (Chen *et al.*, 1979), the fourth residue of peptide T27 proved to be glycine (and not valine). This result was confirmed by the existence of *N*-terminal glycine in peptide CN3S8.

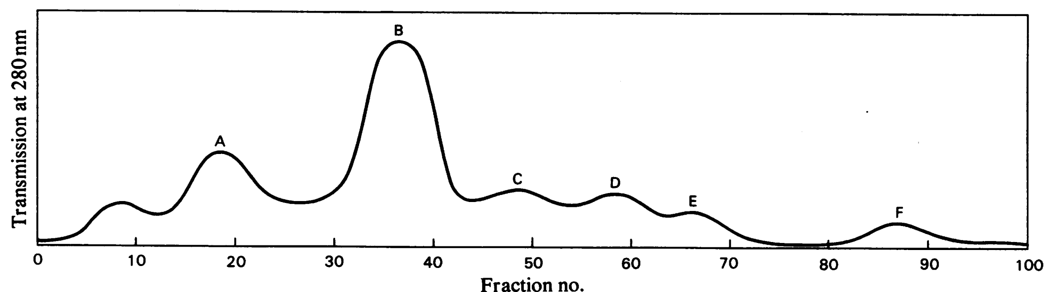


Fig. 5. *Chromatography of tryptic peptides from citraconylated fragment CN3*
Peptides from about 70nmol of citraconylated fragmented CN3 were applied to a column (see the text) which was developed with a flow rate of 6ml/h, and 3ml fractions were collected.

Asp-Val-Glu-Val-Ile-Gly-Asp-Val-Asp-Leu-Val-Asn-Tyr-Phe-Glu-Val-Gly-Ala-Thr-Tyr-Tyr-Phe-Asn-Lys

H1 ←——— H2 ———→ ←——— H3 ———→ ←— H4 —→ ←— H5 —→ ←— H6 ——→ ←—— H7 ——→

Fig. 6. *Amino acid sequence of tryptic peptide T27*
Key: H, thermolytic peptides obtained by subdigestion of peptide T27 with thermolysin.

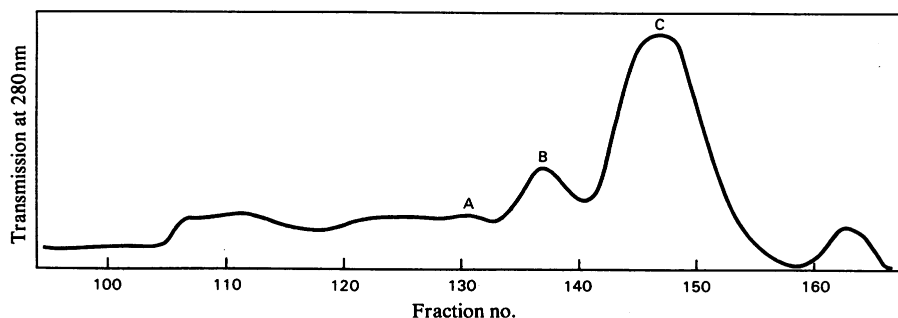


Fig. 7. *Chromatography of proteinase SV8 peptides obtained from fragment CN3*
Peptides from about 70nmol of fragment CN3 were chromatographed as indicated in Fig. 5 legend (fraction volume 2.5ml).

Tryptic digestion of the deacylated material corresponding to peak C (Fig. 5) and subsequent 'fingerprinting' separated peptides T13, T14, T15 and T16. Peptide T13 remained at the origin, but it was pure, and it could be sequenced up to the *C*-terminal lysine residue. Peak D (Fig. 5) contained only peptide T17. From a digest of peptide T17 with proteinase SV8 three peptides (T17S1, T17S2 and T17S3) were isolated by 'fingerprinting'. They could be sequenced to completion, and their order was established by the thermolytic peptide CN3L4H1 (obtained from lysine-blocked peptide CN3LB4 by 'fingerprinting').

Order of tryptic peptides. The lysine-blocked tryptic peptide CN3LB6 was eluted in peak A (Fig. 5). It was sequenced up to residue 21, providing the order T22–T23–T24. The sequence of peptide CN3S7 established the order T24–T25–T26–T27.

The position of the dipeptide Asn-Met C-terminal to peptide T27 was found by the sequence of peptide CN3S9. Peak B (Fig. 5) contained peptide CN3LB5, which was deacylated and cleaved with trypsin, releasing peptides T18 and T20 and lysine (T19). The sequence of peptide CN3S4 established the order T17–T18, and that of peptide CN3S5 the order T18–T19–T20. CNBr treatment of protein I in formic acid/heptafluorobutyric acid (cleavage at methionine and tryptophan residues; see the Experimental section) and separation of the fragments on a Sephadex G-100 column (150cm x 2.5cm) in 50% acetic acid led to the isolation of a peptide with the N-terminal sequence Ala-Thr-Gly-Leu-Lys-Tyr-Asp-Ala.... This sequence showed the order T20–T21.

Peak C (Fig. 5) contained peptide CN3LB3. Peptide 'mapping' of a tryptic digest of deacylated peptide CN3LB3 resolved three peptides (T13, T14, T15) and arginine (T16). Thermolytic peptides, finally, obtained from peptide CN3LB3 (CN3LB3H1–CN3LB3H8) and isolated by 'fingerprinting' clarified the order T13–T14–T15–T16.

Order of lysine-blocked peptides. This order was mainly established by analyses of peptides obtained from fragment CN3 by digestion with proteinase SV8. Partial separation of these S-peptides was achieved by chromatography on a Sephadex G-50 (superfine grade) column (2.5cm x 120cm) in 36% acetic acid (Fig. 7). The material in each peak was further purified by 'fingerprinting' (Figs. 8–10). Sequencing of peptide CN3S1 for 21 cycles showed the order CN3LB1–CN3LB2–CN3LB3. This sequence was followed by peptide CN3LB4 as determined by the sequence of peptide CN3S2. The position of peptide CN3LB5 next to peptide CN3LB4 was established by the sequence of peptide CN3S4. Finally, the N-terminal sequence of peptide CN3S6 provided the overlap CN3LB5–CN3LB6.

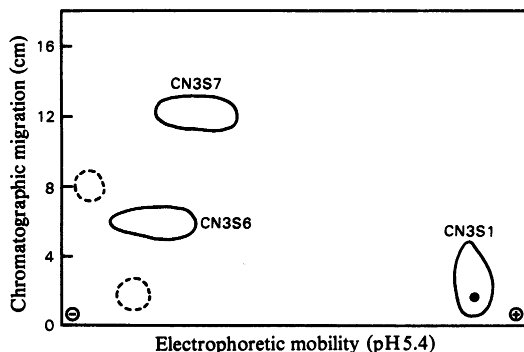


Fig. 8. *Separation of proteinase SV8 peptides obtained from the material in peak A of Fig. 7*

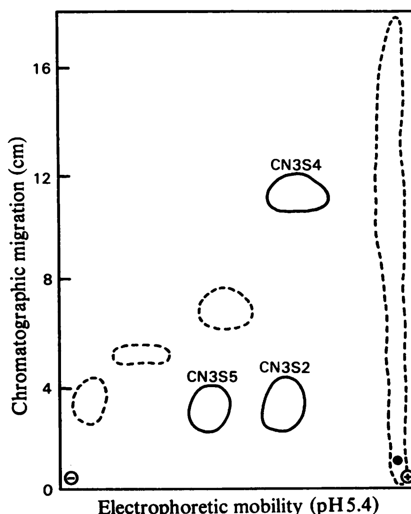


Fig. 9. *Separation of proteinase SV8 peptides obtained from the material in peak B of Fig. 7*

Fragment CN4

This fragment was considerably more problematical to analyse than any of the others. It was difficult to isolate in sufficient quantities because it separated poorly from fragment CN1 (Chen *et al.*, 1978). It has a strong tendency to aggregate and to become insoluble. The single lysine residue at position 16 (residue 323 of the complete protein) proved to be resistant towards trypsin. Fragment CN4 could be sequenced for 24 cycles. Peptide CN4S1, isolated by 'fingerprinting' a proteinase SV8 digest of fragment CN4, could be sequenced to completion. Because of the problems mentioned, we have in addition sequenced all thermolytic peptides derived from fragment CN4 that separate well on 'fingerprints' (Fig. 11). A mistake was discovered at position 325 of the protein, which is glycine instead of the aspartic acid reported previously (Chen *et al.*, 1979).

Tryptophan residues

Tryptophan cannot be identified by using the micro dansyl–Edman technique. For the location of tryptophan residues we had previously used Ehrlich's reagent. The tryptic peptide 'map' was treated with Ehrlich's reagent, which revealed two positive spots. One mobile spot was easily identified to contain peptide T20, whereas the other tryptophan-positive material remained at the origin. Under our experimental conditions several peptides remained in this spot: T7 (34 residues), T11 (32 residues), T17 (28 residues) and T24 (24 residues). In order to determine which of the tryptic peptides contained tryptophan, purified peptides derived from

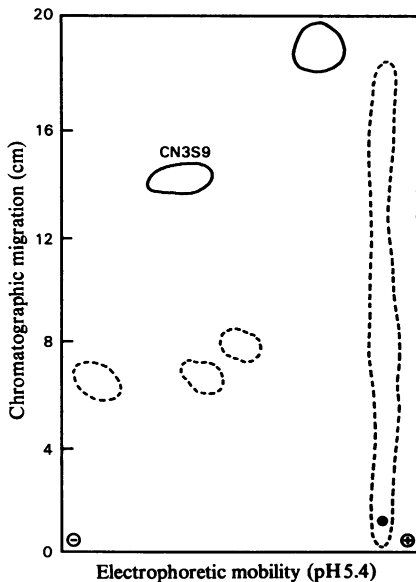


Fig. 10. *Separation of proteinase SV8 peptides obtained from the material in peak C of Fig. 7*
The peptide corresponding to the unidentified spot at the top of the 'fingerprint' could not be sequenced. It may represent a peptide from residues 213–233 (see Fig. 1), possessing an *N*-terminal pyroglutamic acid residue.

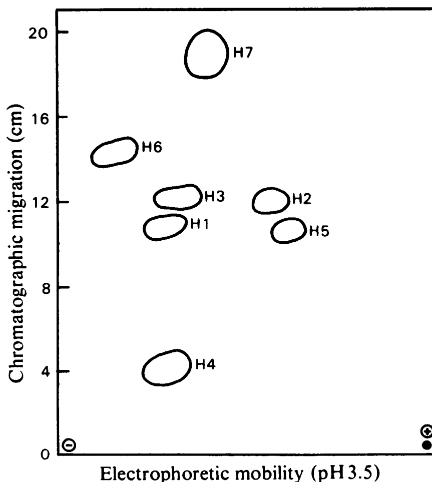


Fig. 11. *Separation of thermolytic peptides from CNBr-cleavage fragment CN4*

a proteinase SV8 digestion that contained the respective tryptic peptides were cleaved again by trypsin. They were 'fingerprinted' and checked for tryptophan. In this way the second tryptophan was located in peptide T7. Since two tryptophan residues seemed to be too low for protein I as compared with five residues reported by Rosenbusch (1974), we spent much time searching for more tryptophan residues. Eventually we believed that we had found a third residue in peptide T11. In our assay the protein was first digested with proteinase SV8. The peptides were then digested with trypsin and thermolysin. Finally they were separated on the peptide 'map' and stained for tryptophan. However, owing to incomplete digestions of the protein by proteinase SV8 and thermolysin, a strong background occurred on the 'map', apparently leading to the incorrect identification of a third tryptophan in peptide T11. Tryptophan-positive peptides were further digested with thermolysin, and positive H-peptides were sequenced. The occurrence of a gap was taken as evidence for the presence of a tryptophan residue. In this way three such residues had erroneously been identified at positions 59, 103 and 214.

We have now used a new cleavage method. CNBr in heptafluorobutyric acid cleaves polypeptides at methionine and tryptophan residues (Ozols & Gerard, 1977). Protein I was cleaved this way, and the fragments were separated as described previously for the CNBr-cleavage fragments (Chen *et al.*, 1978). The elution profile revealed two additional peaks as compared with the earlier experiment. Each peak was sequenced for several cycles, and by comparison of the data with the known sequences of the CNBr-cleavage fragments (methionine cleavage sites) the newly generated fragments (tryptophan cleavage sites) were identified. It had to be concluded that only two tryptophan residues exist: Trp-61 and Trp-214. The correct sequence from residues 59–61 thus is -Gly-Gln-Trp- and not -Trp-Glu-Gly- (Chen *et al.*, 1979). Obviously, the Trp-61 had erroneously been allocated to position 59, and vice versa for Gly-59.

Amide assignments

The number of aspartic acid and glutamic acid residues and their respective amides make up about one-quarter of the 340 residues of protein I, which is the most acidic major outer-membrane protein (Henning *et al.*, 1977, 1978). Since the micro dansyl–Edman method does not distinguish between aspartic acid/asparagine or glutamic acid/glutamine, this procedure in itself is most unsuited for such a protein. By using the DABITC method (Chang & Creaser, 1976) for the assignment of amides in combination with the dansyl–Edman technique, the problem has been overcome. Careful re-examination of our previous results has revealed that nine out of 82 positions have to be corrected as follows: Glu-29, Asn-35, Gln-60, Gln-76, Asn-152, Gln-156, Glu-233, Gln-255 and Gln-339 (see Table

Table 2. *Assignment of amides in the sequence of protein I*

		Method				Method	
Residue	Position	SV8*	DABITC†	Residue	Position	SV8	DABITC†
Glu	2	+	+	Asp	172		+
Asn	5		+	Glu	181	+	+
Asp	7		+	Gln	183		+
Asn	9		+	Asp	195		+
Asp	12	+	+	Asn	198		+
Asn	27		+	Gln	200		+
Glu	29	+	+	Glu	201	+	+
Asn	30		+	Gln	203		+
Asn	35		+	Asn	207		+
Asp	37		+	Glu	212	+	+
Glu	48	+	+	Gln	213		+
Gln	50		+	Asp	221		+
Asn	52		+	Asn	223		+
Asp	54	+	+	Asn	224		+
Gln	60		+	Asn	230		+
Glu	62	+	+	Glu	233	+	+
Asn	64		+	Asn	236		+
Glu	66	+	+	Asn	242		+
Asn	68		+	Asn	246		+
Asn	69		+	Asn	252		+
Glu	71	+	+	Gln	255		+
Asp	74	+	+	Asp	256	+	+
Gln	76		+	Gln	264		+
Asn	79		+	Asp	266	+	+
Asp	92		+	Asp	282		+
Asp	97	+	+	Glu	284	+	+
Asn	101		+	Asp	288		+
Asp	107	+	+	Asp	290		+
Asp	113		+	Asn	293		+
Asp	121		+	Glu	296	+	+
Asp	126	+	+	Asn	304		+
Asp	127	+	+	Asn	306		+
Asn	141		+	Asp	312		+
Asn	143		+	Asn	316		+
Asp	149		+	Gln	317		+
Asn	152		+	Asp	319		+
Gln	156		+	Asp	321		+
Asn	161		+	Asn	322		+
Glu	162	+	+	Asp	329		+
Asp	164		+	Asp	330		+
Asn	170		+	Gln	339		+

* Specificity of staphylococcal proteinase SV8.
† Analysis of 4-*NN*-dimethylaminoazobenzene 4′-thiohydantoin derivative.

2). Seven Asx–Gly bonds exist in protein I. Four of them are Asn–Gly, namely 27–28, 35–36, 170–171 and 207–208. No rearrangements of these bonds have been observed. We had pointed out that in protein I a rather uneven distribution of glutamic acid and aspartic acid residues can be seen (Chen *et al.*, 1979). This fact remains, and the corrections alter the published scheme only slightly.

Concluding remarks

The pore-forming protein (or close relatives thereof) that is the subject of the present work exhibits a number of rather unusual properties. It does not tightly bind dodecyl sulphate as long as it is not denatured by boiling in the detergent (Rosenbusch, 1974), yet, in spite of its acidic nature, it has

some solubility in benzene or toluene (Rosenbusch & Mueller, 1977). It exists in the outer membrane in trimeric form (Steven *et al.*, 1977; Palva & Randall, 1978; Nakae *et al.*, 1979; Yu *et al.*, 1979) and possesses a high degree of β-conformation (Rosenbusch, 1974; Nakamura & Mizushima, 1976) not predictable from the primary structure (Chen *et al.*, 1979). The protein from *E. coli* BE has been crystallized (Garavito & Rosenbusch, 1980), and elucidation of three-dimensional structures can be awaited with much interest.

Financial support by the Fonds der Chemischen Industrie is gratefully acknowledged.

References

Benz, R., Janko, K., Boos, W. & Läuger, P. (1978) *Biochim. Biophys. Acta* **511**, 305–319

Benz, R., Janko, K. & Läuger, P. (1979) *Biochim. Biophys. Acta* **551**, 238–247

Chang, J. Y. & Creaser, E. H. (1976) *Biochem. J.* **157**, 77–85

Chen, R. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* **357**, 873–886

Chen, R., Hindennach, I. & Henning, U. (1978) *Hoppe-Seyler's Z. Physiol. Chem.* **359**, 1807–1810

Chen, R., Krämer, C., Schmidmayr, W. & Henning, U. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5014–5017

Chen, R., Schmidmayr, W., Krämer, C., Chen-Schmeisser, U. & Henning, U. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4592–4596

Diedrich, D. L. & Schnaitman, C. A. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3708–3712

Gamon, K., Chen, R. & Henning, U. (1978) *Mol. Gen. Genet.* **166**, 187–192

Garavito, R. M. & Rosenbusch, J. P. (1980) *J. Cell Biol.* **86**, 327–329

Garten, W. & Henning, U. (1974) *Eur. J. Biochem.* **47**, 343–352

Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 379–380

Hantke, H. (1976) *FEBS Lett.* **70**, 109–112

Henning, U., Schmidmayr, W. & Hindennach, I. (1977) *Mol. Gen. Genet.* **154**, 293–298

Henning, U., Sonntag, I. & Hindennach, I. (1978) *Eur. J. Biochem.* **92**, 491–498

Heuzenroeder, M. W. & Reeves, P. (1980) *J. Bacteriol.* **141**, 431–435

Hindennach, I. & Henning, U. (1975) *Eur. J. Biochem.* **59**, 207–213

Ichihara, S. & Mizushima, S. (1978) *J. Biochem. (Tokyo)* **83**, 1095–1100

Luckey, M. & Nikaido, H. (1980) *Biochem. Biophys. Res. Commun.* **93**, 166–171

Nakae, T. (1976) *Biochem. Biophys. Res. Commun.* **71**, 877–884

Nakae, T. (1979) *Biochem. Biophys. Res. Commun.* **88**, 774–781

Nakae, T., Ishii, J. & Tokunaga, M. (1979) *J. Biol. Chem.* **254**, 1457–1461

Nakamura, K. & Mizushima, S. (1976) *J. Biochem. (Tokyo)* **80**, 1411–1422

Nikaido, H. (1979*a*) *Angew. Chem.* **91**, 394–407

Nikaido, H. (1979*b*) in *Bacterial Outer Membranes* (Inouye, M., ed.), pp. 361–407, John Wiley and Sons, New York

Ozols, J. & Gerard, C. (1977) *J. Biol. Chem.* **252**, 5986–5989

Palva, E. T. & Randall, L. L. (1978) *J. Bacteriol.* **133**, 279–286

Pugsley, A. P. & Schnaitman, C. A. (1978) *J. Bacteriol.* **135**, 1118–1129

Pugsley, A. P., Lee, D. R. & Schnaitman, C. A. (1980) *Mol. Gen. Genet.* **177**, 681–690

Rosenbusch, J. P. (1974) *J. Biol. Chem.* **249**, 8019–8029

Rosenbusch, J. P. & Mueller, R. (1977) in *Solubilization of Lipoprotein Complexes* (Peeters, H. & Massué, J. P., eds.), pp. 59–68, European Press, Ghent

Schindler, H. & Rosenbusch, J. P. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3751–3755

Spackman, D. H., Stein, W. H. & Moore, S. (1958) *Anal. Chem.* **30**, 1190–1206

Steven, A. C., Ten Heggeler, B., Müller, R., Kistler, J. & Rosenbusch, J. P. (1977) *J. Cell Biol.* **72**, 292–301

Tomassen, J. & Lugtenberg, B. (1980) *J. Bacteriol.* **143**, 151–157

Yu, F., Ichihara, S. & Mizushima, S. (1979) *FEBS Lett.* **100**, 71–74

Zalman, L. S., Nikaido, H. & Kagawa, Y. (1980) *J. Biol. Chem.* **255**, 1771–1774.