Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 1048–1052, February 1984
Biochemistry

Amino acid sequence homology among the major outer membrane proteins of *Escherichia coli*

(porin/phage λ receptor protein/OmpA protein/protein secretion)

HIROSHI NIKAIDO* AND HENRY C. P. WU†

*Department of Microbiology and Immunology, University of California, Berkeley, CA 94720; and †Department of Microbiology, Uniformed Services University of Health Sciences, Bethesda, MD 20814

Communicated by Daniel E. Koshland, Jr., November 4, 1983

ABSTRACT **Analysis of amino acid sequences reported for the major outer membrane proteins of *Escherichia coli*, including the porins (OmpF, OmpC, and PhoE), the phage λ receptor (LamB), and another protein (OmpA), revealed several regions of local homology that is statistically significant. The implications of this observation are discussed in relation to the evolutionary origins of these proteins, as well as to the mechanism of export of these proteins to the outer membrane.**

All Gram-negative bacteria possess an outer membrane (1) that contains a unique lipid component, lipopolysaccharide, as well as a unique set of proteins. The major outer membrane proteins of *Escherichia coli* K-12 are OmpF and OmpC porins, which produce transmembrane diffusion channels, and OmpA protein and murein lipoprotein, which apparently play structural roles (1, 2). In addition, several proteins are induced to become a major component under different growth conditions—e.g., PhoE porin by phosphate starvation and the phage λ receptor, LamB protein, by the presence of maltose.

The possible presence of amino acid sequence homology among these major outer membrane proteins was of interest for several reasons. First, at least two of these classes of proteins, porins and LamB, are similar in possessing a channel-forming function (3) and similar in their physical construction, being β-sheet-rich trimers resistant to denaturation by sodium dodecylsulfate (4–7). Second, porins, LamB, and OmpA are all known to have high affinity for lipopolysaccharide (8, 9) and can interact with the peptidoglycan layer under appropriate conditions (4, 10, 11). Third, because all of these proteins are located in the outer membrane, there may be a common "signal" for their translocation. Various observations suggest that if such a signal were present, it likely would be located within the sequence of the mature proteins. For example, it is known that the amino terminal "leader" sequence of the LamB precursor by itself was unable to direct the translocation to the outer membrane of another polypeptide fused to it (12). Fourth, translocation of proteins into the outer membrane could require a complex transport mechanism. It is unlikely that such a mechanism was "reinvented" for each new outer membrane protein as it evolved; therefore, many outer membrane proteins may have been derived from a common ancestral protein.

Sequence homologies between *E. coli* outer membrane proteins have been examined before. Although strong homologies were found among three porins (OmpF, OmpC, and PhoE) (13, 14), no significant homology was reported between OmpA and OmpF (15) or between LamB and either OmpA or OmpF (16). However, these comparisons looked for global homologies, demanding significant similarities throughout the entire protein sequence. In view of the probably distant relationship between these proteins, sequence similarities may not have been conserved throughout their entire lengths. Therefore, we searched for local homologies involving only short segments of proteins and found statistically significant homologies among all of these outer membrane proteins except the murein lipoprotein.

METHODS

Sources of Sequence Information. Amino acid sequences for *E. coli* K-12 OmpA and *E. coli* B/r OmpF were those determined by Chen and co-workers (15, 17). Nucleotide sequence analysis of the *ompF* gene of *E. coli* K-12 showed that this OmpF protein is identical to the B/r protein except for substitutions at positions 66, 117, and 262 (18). The sequence of OmpA also was confirmed by DNA sequence determination (19). The amino acid sequences for OmpC and PhoE are those deduced from the DNA sequences determined by Mizuno *et al.* (14) and by Overbeeke *et al.* (13). The LamB and TraT sequences are based on the DNA sequences of Clément and Hofnung (16) and Ogata *et al.* (20). The sources of other sequences used are listed in ref. 21.

Searches for Potentially Homologous Regions. All computer programs were written for a Tandy model I personal computer; to speed up the execution, most of the routines were written in Z80 assembly language.‡ Potentially homologous regions between proteins were identified by a two-step process. (*i*) Initially, regions of potential homology were searched for by comparisons of short segments without the creation of gaps. Thus, the program compared all possible short segments (e.g., containing seven amino acid residues) from one protein with all or a limited subset of segments of the same length from the second protein. If a comparison revealed a set number (e.g., four) or more of identical amino acids at corresponding positions, the matched region was analyzed further as described below.

We emphasize that, for a number of reasons, the matches detected here can serve only as the starting points for the search for homology: homology of short sequences is statistically very unreliable (22); gaps may be needed for optimum alignment, but they were not introduced; and, finally, the search was performed purposefully at low degrees of stringency so as not to miss the real regions of homology. (*ii*) We then optimized the alignment of the more extended segments containing 11–30 amino acid residues, centered on the short segments of potential match identified above. This procedure permitted the insertion of gaps, if necessary, with a set amount of penalty ("gap penalty factor" of 2) given to the creation of each gap and also another penalty ("gap bias factor" of 5) given to each missing residue within a gap (23). The extent of homology in the alignment was quantitated by calculating "alignment index" by using the "Mutation Data Matrix" of Barker and Dayhoff (23) for each amino acid pair, and the alignment was optimized (24) to yield the highest

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡The program will be available on request from H. Nikaido.

alignment index. Because a positive gap bias is awarded to regions without gaps in our procedure, alignment indices were corrected by subtracting the number of amino acid residues in the final aligned segment multiplied by 5 (i.e., the gap bias factor). The statistical significance of each aligned segment was then determined as described in the next section.

Statistical Significance of the Proposed Alignment. Usually, the statistical significance of a match is determined by randomizing the sequences and comparing the alignment index of the real sequences with those of randomized sequences (22, 23). However, randomization of the sequences of the short segments we obtained would give statistically incorrect results because these segments were already preselected from the global sequences as those giving the best match. Therefore, we randomized the sequences of the entire proteins and went each time through the procedures described above, as follows. First, regions of potential match were picked out without the creation of gaps. Second, the alignment of longer segments centered on each of these regions of potential match was optimized. Third, alignment indices of all of the potentially homologous segments were compared, and the highest one was recorded as the "best alignment index." This sequence was repeated at least 1000 times in two-way comparisons between outer membrane proteins.

The significance of three-way alignments was assessed in a similar manner—i.e., by randomizing the whole protein sequences, searching for 7-residue segments containing at least 3 residues common to all three proteins at corresponding positions, and calculating the alignment index for 11-residue segments centered on the 7-residue segments after optimizing the alignment. Because of the long time required by the computer for this procedure, three-way comparison of the randomized sequences was repeated only 100 times. Calculation of the alignment index also was done in a simpler manner, by assigning 10 points for any two-way matches of identical amino acid residues and by assigning −5 points for the creation of any gap and −3 points for any missing residue in a gap. Independent three-way comparisons were also carried out with a method similar to that described by Doolittle and co-workers (22, 25)—i.e., by randomizing only the segments to be compared and calculating the alignment index as above after alignment optimization.

RESULTS

Tentative Alignment of OmpF, LamB, and OmpA Sequences. Amino acid sequence comparisons of OmpF with LamB and LamB with OmpA were made by matching all possible 7-residue segments of one protein against similar segments from the other protein. Because we hypothesized that the proteins were derived from a common ancestral protein, we expected that homologous segments would occur at roughly similar positions in the proteins being compared. Therefore, we allowed a relative "shift" of up to 50 residues in either direction, that is, a segment beginning with residue n of the first protein was compared only with segments beginning with residue $n - 50$ through $n + 50$ of the second protein. Fifteen (OmpF vs. LamB) and 11 (LamB vs. OmpA) matches containing 4 or more identical amino acid residues in corresponding positions were found. Some of these matches were probably fortuitous because probability considerations (21) indicate that there should be about 7 such matches in proteins of this size due to chance alone. However, several of these matches appeared to be significant in two ways. First, when the 20-residue stretches centered on these 7-residue segments were optimally aligned and alignment indices were calculated (see *Methods*), some segments had indices much higher than expected from chance matches between segments of randomized sequences (see below). Second, these better-matched segments were in approximate register with each other. These segments included those containing residues 11–30 of OmpF and residues 1–20 of LamB (Fig. 1, region a) and those centered on residues 161–167 of OmpF and 197–203 of LamB as well as those containing 185–195 of LamB and 160–169 of OmpA (Fig. 1, region d).

There were several more pairs of segments picked out by computer as potentially homologous, which were in approximate register with each other and with those described above but, by themselves, did not exhibit the degree of homology that could be proven as statistically significant. These regions were aligned by computer using the Needle-

```
a   OmpF:   11  Val Asp Leu Tyr Gly Lys Ala Val Gly Leu His Tyr Phe Ser Lys Gly Asn Gly Glu --
    LamB:    1  Val Asp Phe His Gly Tyr Ala Arg Ser Gly Ile Gly Trp Thr Gly Ser Gly Gly Glu --

    (PhoE:  27  Ala Ser Lys Asp     Gly Asp Gln Ser Tyr Ile Arg Phe Gly ---)
    (OmpC:  29  Lys Asp Val Asp     Gly Asp Gln Thr Tyr Met Arg Leu Gly ---)
b   OmpF:   32  Tyr Gly     Gly Asn Gly Asp Met Thr Tyr Ala Arg Leu Gly ---
    LamB:   32  Tyr Arg Leu Gly Asn Glu Cys Glu Thr Tyr Ala Glu Leu Lys ---
    OmpA:    1  Ala Pro Lys Asp Asn Thr Trp Tyr Thr Gly Ala Lys Leu Gly ---

c   LamB:  150  Gly Gly Ser Ser Ser Phe Ala Ser Asn Asn Ile Tyr Asp Tyr Thr Asn Glu Thr Ala ----
    OmpA:  118  Gly Val Ser Pro Val Phe Ala Gly Gly Val Glu Tyr Ala Ile Thr Pro Glu Ile Ala ----

    (PhoE:  144  Gly     Leu Asn Leu Thr Leu Gln Tyr Gln Gly Lys     Asn Glu Asn                    Arg Asp Val Lys Lys --)
    (OmpC:  142  Gly     Leu Asn Phe Ala Val Gln Tyr Gln Gly Lys     Asn Gly Asn--(14 amino acids)--Arg Asp Ala Leu Arg --)
d   OmpF:   150  Gly     Leu Asn Phe Ala Val Gln Tyr Leu Gly Lys     Asn Glu                        Arg Asp Thr Ala Arg --
    LamB:   185  Gly Thr Leu Glu Leu Gly Val Asp Tyr     Gly Arg Ala Asn Leu                        Arg Asp Asn Tyr Arg --
    OmpA:   160  Gly Met Leu Ser Leu Gly Val Ser Tyr         Arg Phe     Gly Gln--(hinge region)--

    OmpF:   195  Asp Arg Thr Asn Leu Gln Glu Ala Gln Pro Leu     Gly Asn Gly Lys             Lys Ala
    LamB:   219  His     Thr         Gln     Ser     Val Leu Lys Gly Phe Asn Lys Phe Val Val
    OmpA:   193  His Phe Thr     Leu     Lys Ser Asp Val Leu Phe Asn Phe Asn Lys Ala Thr Leu Lys Pro
e
    OmpF:   212  Glu     Gln Trp Ala Thr Gly Leu Lys Tyr Asp Ala Asn Asn Ile Tyr ----
    LamB:   233          Gln Tyr Ala Thr Asp Ser Met Thr Ser Gln Gly Lys Gly Leu ----
    OmpA:   212  Glu Gly Gln Ala Ala Leu Asp Gln Leu Tyr Ser Gln Leu Ser Asn Leu ----
```

FIG. 1. Homology between porins, LamB, and OmpA. The sequences of the OmpC and PhoE porins were shown only for regions b and d, as no new matches among OmpF, LamB, and OmpA could be found in other regions by considering these sequences. There is no significant homology between porins and the LamB-OmpA pair in region c. Residues common to OmpF, LamB, and OmpA proteins are double-underlined, residues shared by two of the three proteins are underlined, and the residues common to LamB (or OmpA) and PhoE (or OmpC) are underlined with broken lines. The *ompC* gene has been shown to have an insertion, corresponding to 14 extra amino acids between Asn and Arg (14) within region d. In the same area, OmpA protein has a long alternating Ala-Pro sequence (15) that is suspected to act as the hinge between the membrane-embedded and peripheral domains of the protein. In region b, an alternative alignment can be generated by aligning residues 33–35 of LamB with residues 42–44 of OmpF. This alignment leads, however, to excessive numbers of gaps when aligned with OmpA and is not shown.

```
OmpF:   86  Ala Gly Leu Lys     Tyr Ala Asp Val Gly Ser Phe Asp Tyr Gly ----
LamB:  127  Ala Gly Leu Glu     Asn Ile Asp Val Gly     Phe          Gly ----
OmpA:   41  Gly Gly Tyr Gln Val Asn Pro Tyr Val Gly     Phe Glu Met Gly ----
```

FIG. 2. Homology between OmpF, LamB, and OmpA proteins requiring larger displacement of sequences against each other. Underlining of the homologous residues follows the convention described in Fig. 1. For details see text.

man–Wunsch algorithm (Fig. 1, regions b, c, and e). The three-way alignments shown in Fig. 1 were produced in this manner, as modified slightly by consideration of matches involving two other porins, OmpC and PhoE. The three-way comparison further strengthened the apparent significance of the alignment because aligning OmpF with LamB and LamB with OmpA brought out regions of homology between OmpF and OmpA—for example, between residues 35 and 44 of OmpF and residues 5 and 14 of OmpA (Fig. 1, region b) and between residues 150 and 160 of OmpF and residues 160 and 169 of OmpA (Fig. 1, region d). In addition, comparison with PhoE and OmpC sequences showed that, in most areas of homology, these proteins contained identical or functionally similar amino acids to those shared by the three other proteins (Fig. 1, regions b and d). In some segments where the other porins lacked strong homology with OmpF, they showed strong resemblance to other outer membrane proteins—for example, between residues 27 and 30 of PhoE and 1 and 4 of OmpA (Fig. 1, region b).

We also looked for four out of seven matches in searches without any restriction on the size of the shifts. Most of the additional matches found appeared to be fortuitous, as there was little homology in the regions flanking these additional matched segments. Nonetheless, a fairly good alignment could be produced among residues 86–99 of OmpF, 127–137 of LamB, and 41–54 of OmpA (Fig. 2). In order to bring this match in register with those listed in Fig. 1, however, long gaps of about 40 residues had to be introduced into both the OmpF and OmpA sequences. For this reason, this alignment was not investigated further.

Significance of the Matches Found. The statistical significance of the two-way matches between segments of the authentic proteins was examined by comparison to the best alignment index for a pair of segments found after randomization of the entire protein sequences (see *Methods*). The average and standard deviation of the best-alignment indices were calculated after at least 1000 trials. The alignment index of the real sequences was then compared with these figures, and an alignment score was assigned. Alignment score was calculated as (alignment index of the real sequences − average of best alignment indices with randomized sequences)/standard deviation (23). The 20-residue segments centered around residues 185–191 of LamB, and residues 160–166 of OmpA had an alignment score of 2.3 (Table 1); this value indicates that the probability that this match was fortuitous was about 0.01, if we assume that the values of the best alignment index between randomized sequences followed the normal distribution. Similarly, such a probability was about 0.1 for the alignment of 20-residue segments centered around residues 161–167 of OmpF and residues 197–203 of LamB, on the basis of the alignment score of 1.3 (Table 1).

Thus, for the OmpF with LamB match, there was a 1-in-10 chance that our conclusion was wrong. Therefore, we focused our attention on the second most conspicuous homologous region between OmpF and LamB—the region containing residues 11–30 of OmpF and 1–20 of LamB, which gave an alignment index of 33. When the alignment indices were examined after the computer randomization of the protein sequences, there were 91 cases in 1000 trials in which the alignment index of the best-matched segments was equal to or larger than 37 (the alignment index between the 154–173 segment of OmpF and 190–209 segment of LamB), in agreement with the 1-in-10 chance predicted from the alignment score as described above. However, in only 4 of these 91 cases was there also another pair of segments whose alignment index was equal to or exceeded 33. We feel, therefore, that the homology between OmpF and LamB is a significant one.

Significance of the Three-way Alignment. When the significance of the three-way alignment of region b of Fig. 1 was

Table 1. Two-way alignment index of the best-matched 20-residue segments from real and randomized sequences

Protein segments	Alignment index: Real sequences	Alignment index: Randomized sequences Mean	Alignment index: Randomized sequences SD	Alignment score
LamB vs. OmpA	43	22	7.4	2.3
OmpF vs. LamB	37	27	7.6	1.3
OmpF vs. murein lipoprotein	–*	7†	9.1†	
OmpF vs. TEM β-lactamase	15	19	9.2	−1.6
OmpF vs. alkaline phosphatase	13	22	8.2	−1.1
OmpF vs. sulfate-binding protein	23	23	7.0	0
OmpF vs. MalK	9	19	8.5	−1.2
OmpF vs. LacY	29	25	7.6	0.5
OmpF vs. F_0-ATPase, subunit a	6	20	9.3	−1.5
OmpF vs. elongation factor Tu	22	22	6.9	0
OmpF vs. Lac repressor	20	19	7.6	0.1
OmpF vs. tryptophan synthetase, a	14	20	7.6	−0.8

The alignment was done as described in *Methods*—i.e., by first looking for 7-residue segment pairs containing 4 or more identical amino acids at corresponding positions, then optimizing the alignment of 20-residue segments centered on the 7-residue segments. The procedure was repeated after randomizing the sequences of both of the proteins at least 1000 times for comparisons between outer membrane proteins and 400 times for other comparisons.

*There were no four-of-seven matches.

†In 52% of comparisons, no matches were found.

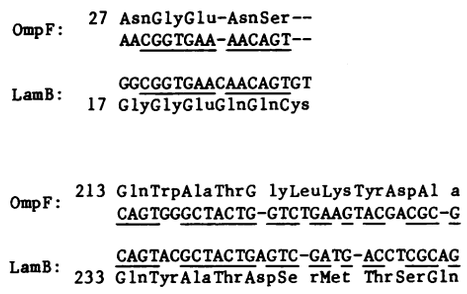

FIG. 3. Examples of DNA sequence homology between *ompF* and *lamB* genes. Because segments with a high degree of amino acid sequence homology obviously will be homologous in terms of DNA sequence, we have chosen here segments that are adjacent to such segments but do not show much amino acid sequence homology, yet are quite homologous in nucleotide sequence. In segment a, Asn-Ser and Gln-Gln-Cys that follow the homologous segments Asn(or Gly)-Gly-Glu (end of region a in Fig. 1) do not show much homology at the amino acid level but are almost identical at the DNA level if the presence of two frameshift mutations is assumed. Similarly, in segment b, the segments after the homologous peptide Gln-Trp(or Tyr)-Ala-Thr (middle of region e in Fig. 1) show high homology at the DNA level, yet practically no homology at the amino acid sequence level. Identical nucleotides are underlined. DNA sequence data were obtained from refs. 16 and 18.

tested as described by global randomization of OmpF, LamB, and OmpA sequences, the alignment index of the real sequences was 140, whereas the average of the indices of the best-aligned segments among randomized sequences was 108, with a standard deviation of 9.8, yielding an alignment score of 3.3. Hence, the probability that the match shown in Fig. 1, region b was merely fortuitous was only about 0.0005.

The significance of the matches in regions b and d of Fig. 1 also was tested by randomizing only the sequences of the tested segments and calculating the alignment indices of randomized segments. The alignment scores of the real sequences for these regions were 5.2 and 6.2, respectively. Although it is impossible in this situation to calculate the probability that the original alignment was due to random chance because we are dealing with short, preselected areas within large proteins, these scores are in the range generally regarded as highly significant by workers who performed three-way alignment studies of limited regions of a variety of other proteins (22, 25).

Comparison of the DNA Sequences. A limited manual comparison of published nucleotide sequences was performed for the homologous regions identified in Fig. 1. In most cases, there appeared to be corresponding homology in the DNA sequences, although no statistical analysis was performed. In some cases, DNA sequences showed striking homology in segments that were adjacent to the homology region identified by amino acid sequence, yet did not show much amino acid sequence homology. Two examples are shown in Fig. 3.

Comparison with the TraT Sequence. TraT is an outer membrane protein coded for by a plasmid gene (20). There were clear similarities betweeen the 200–212 segment of OmpA and the 213–225 segment of TraT and between the 71–79 segment of OmpA and the 75–83 segment of TraT (Fig. 4). The first region of similarity corresponds to the first half of region e in Fig. 1. However, TraT protein did not seem to contain sequences that are strongly homologous to the other major areas of homology among the porins, LamB, and OmpA.

Comparison of OmpF Sequence with Some Other *E. coli* Proteins. Using the same method that was used for the comparison of the major outer membrane proteins, we compared the sequence of OmpF to those of murein lipoprotein and three periplasmic, three cytoplasmic-membrane, and three cytosolic proteins of *E. coli* (Table 1). In one case, there were no seven-residue segments that shared four or more identical residues. Even in those cases where there were some apparently homologous segments, the alignment indices of these regions were uniformly low, about the level seen for matches between randomized sequences, or lower (Table 1). This result further strengthens our conclusion that the homology observed among the major outer membrane proteins is a significant one.

DISCUSSION

The comparison of amino acid sequences described here showed areas of local homology among several major outer membrane proteins of *E. coli* (Fig. 1). It is known, however, from earlier analysis of other protein sequences (26), that sometimes even those alignments perceived visually as very strong turn out to be statistically insignificant. Therefore, a careful statistical treatment was necessary. This problem is even more severe in comparisons of the type we performed—i.e., those involving local rather than global homology. We chose to use the Monte Carlo approach and made local comparisons after the randomization of the sequences of the whole proteins (see *Methods*). The results, especially those of three-way comparisons, clearly showed that the homology was significant.

It is possible to find, by using different search methods, other apparent regions of homology among the proteins. For example, OmpF, LamB, and OmpA share two tripeptide sequences, Gly-Leu-Lys and Asn-Asn-Ile. However, we are skeptical of the significance of these matches for two reasons. First, the regions flanking the tripeptides, as well as in between the tripeptides, showed very little homology. Second, statistical analysis shows that matches of short sequences are not very significant (22). Furthermore, certain short amino acid sequences are likely to occur much more frequently in proteins than anticipated purely from probability considerations because certain combination of residues may perform specific functions in the formation of secondary structures (27). Several areas of local homology between OmpA and OmpF also have been identified by Movva *et al.* (19), presumably by inspection. It is striking, however, that none of these regions showed statistically significant homology according to our method of analysis (results not shown).

The amino acid sequence homologies found (Fig. 1) and the DNA sequence homology (Fig. 3) are best explained by a common evolutionary origin of these outer membrane pro-

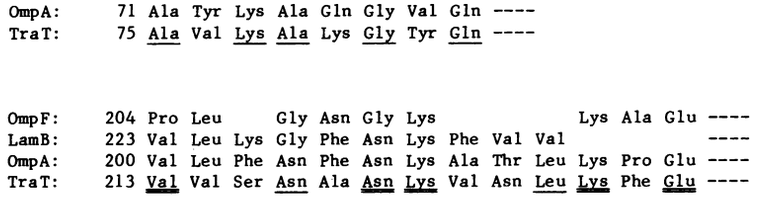

FIG. 4. Homology between TraT and other major outer membrane proteins. The residues common to two sequences are underlined, those common to three or more sequences are double-underlined.

teins. The conservation of homology only in limited local areas would be explained if these areas define domains involved in functions shared by all these proteins. For example, they could represent sites of interaction with a common ligand, such as peptidoglycan or lipopolysaccharide (see the Introduction). Alternatively, one of these regions might act as a recognition signal to determine the final location of the proteins. In fact, Benson and Silhavy (28) have shown, by the introduction of in-frame deletions, that the region containing residues 0–75 of LamB is essential in the export of the protein into the outer membrane. In a more recent work, this estimate was narrowed to the region corresponding to residues 0–49 (S. A. Benson, personal communication), a segment containing one of the homologous regions we identified in this study (Fig. 1, region b). These considerations suggest that a critical test of the functional significance of this region by directed mutagenesis would be worthwhile. Although two outer membrane proteins, lipoprotein and TraT, did not share a homologous sequence in the region b of Fig. 1, we suspect that these cases may be exceptions that actually prove the rule. Unlike the porins, LamB, and OmpA, both lipoprotein and TraT (29) contain large hydrophobic groups with long chain fatty acids, and their mechanism of translocation to the outer membrane may be quite different from that of the other outer membrane proteins.

Note Added in Proof. The nucleotide sequence of the gene coding for another outer membrane protein, TolC, was recently determined by J. Hackett and P. Reeves (30). The region containing residues 5–16 of this protein appeared to be homologous with the region b of Fig. 1, further supporting the potential significance of this region in the proteins of *E. coli* outer membrane.

This work was supported by National Institutes of Health Research Grants AI-09644 (to H.N.) and GM-28811 (to H.C.P.W.). We thank Drs. A. N. Glazer and J. Thorner for valuable criticisms on the manuscript.

1. Nikaido, H. & Nakae, T. (1979) *Adv. Microbiol. Physiol.* **20,** 163–250.
2. Lugtenberg, B. & van Alphen, L. (1983) *Biochim. Biophys. Acta* **737,** 51–115.
3. Nikaido, H., Luckey, M. & Rosenberg, E. Y. (1980) *J. Supramol. Struct.* **13,** 305–313.
4. Rosenbusch, J. P. (1974) *J. Biol. Chem.* **249,** 8019–8029.
5. Nakae, T. (1976) *J. Biol. Chem.* **251,** 2176–2178.
6. Palva, E. T. & Westermann, P. (1979) *FEBS Lett.* **99,** 77–80.
7. Garavito, R. M., Jenkins, J. A., Neuhaus, J. M., Pugsley, A. P. & Rosenbusch, J. P. (1982) *Ann. Microbiol. Inst. Pasteur (Paris)* **133A,** 37–41.
8. Schindler, H. & Rosenbusch, J. P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2302–2306.
9. Schweizer, M., Hindennach, I. & Henning, U. (1978) *Eur. J. Biochem.* **82,** 211–217.
10. Gabay, J. & Yasunaka, K. (1980) *Eur. J. Biochem.* **104,** 13–18.
11. Endermann, R., Krämer, C. & Henning, U. (1978) *FEBS Lett.* **86,** 21–24.
12. Hall, M. N., Emr, S. D. & Silhavy, T. J. (1980) *J. Supramol. Struct.* **13,** 147–163.
13. Overbeeke, N., Bergmans, H., van Mansfeld, F. & Lugtenberg, B. (1983) *J. Mol. Biol.* **163,** 513–532.
14. Mizuno, T., Chou, M.-Y. & Inouye, M. (1983) *J. Biol. Chem.* **258,** 6932–6940.
15. Chen, R., Schmidmayer, W., Krämer, C., Chen-Schmeisser, U. & Henning, U. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 4592–4596.
16. Clément, J.-M. & Hofnung, M. (1981) *Cell* **27,** 507–514.
17. Chen, R., Krämer, C., Schmidmayer, W., Chen-Schmeisser, U. & Henning, U. (1982) *Biochem. J.* **203,** 33–43.
18. Inokuchi, K., Mutoh, N., Matsuyama, S. & Mizushima, S. (1982) *Nucleic Acids Res.* **10,** 6957–6968.
19. Movva, N. R., Nakamura, K. & Inouye, M. (1980) *J. Mol. Biol.* **143,** 317–328.
20. Ogata, R. T., Winters, C. & Levine, R. P. (1982) *J. Bacteriol.* **151,** 819–827.
21. Gilson, E., Higgins, C. F., Hofnung, M., Ames, G. F.-L. & Nikaido, H. (1982) *J. Biol. Chem.* **257,** 9915–9918.
22. Doolittle, R. F. (1981) *Science* **214,** 149–159.
23. Barker, W. C. & Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure 1972*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, pp. 101–110.
24. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48,** 443–453.
25. Jue, R. A., Woodbury, N. W. & Doolittle, R. F. (1980) *J. Mol. Evol.* **15,** 129–148.
26. Haber, J. E. & Koshland, D. E. (1970) *J. Mol. Biol.* **50,** 617–639.
27. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47,** 45–148.
28. Benson, S. A. & Silhavy, T. J. (1983) *Cell* **32,** 1325–1335.
29. Perumal, N. B. & Minkley, E. G., Jr. (1982) *Abstr. Annual Meeting of the American Society for Microbiology* (Washington, DC), 167.
30. Hackett, J. & Reeves, P. (1983) *Nucleic Acids Res.* **11,** 6487–6495.