Proc. Natl. Acad. Sci. USA
Vol. 76, No. 10, pp. 5014-5017, October 1979
Biochemistry

Primary structure of major outer membrane protein I of *Escherichia coli* B/r

(transmembrane protein/microscale sequence determination/porin)

ROBERT CHEN, CORNELIA KRÄMER, WALDTRAUD SCHMIDMAYR, AND ULF HENNING

Max-Planck-Institut für Biologie, 7400 Tübingen, West Germany

Communicated by Feodor Lynen, July 16, 1979

ABSTRACT **The amino acid sequence of the pore-forming outer membrane protein I (porin) from *Escherichia coli* B/r has been determined. The polypeptide contains 340 amino acid residues resulting in a molecular weight of 37,205. The transmembrane polypeptide has no stretches of nonpolar residues, uninterrupted by charged side chains, longer than 11 amino acid residues. Regarding polarity, the chain can be subdivided into three regions: a distinctly hydrophilic region between residues 1 and 82 (51.2% polarity), a fairly nonpolar region between residues 83 and 194 (33.9% polarity), and a more hydrophilic region up to the COOH terminus (48% polarity). These results are interpreted as evidence against a simple transmembrane structure in which the membrane is spanned by a single contiguous sequence of hydrophobic amino acids, as has been proposed, for example, for glycophorin.**

The cell envelope of Gram-negative bacteria possesses, in addition to the plasma membrane, an outer membrane which has perhaps more correctly also been called a porous skeletal organ (see ref. 1 for recent review) (2). This porosity is provided by proteins, the porins (e.g., refs. 3–5), which form hydrophilic channels allowing the diffusion of various low molecular weight solutes. In *Escherichia coli* B/r, protein I (6) [closely related to Rosenbusch's matrix protein from *E. coli* BE, (7)] is the porin responsible for the existence of these channels, which have a diameter of about 0.9 nm (8, 9).

Aside from the fact that information is sparse regarding structure–function relationships of integral membrane proteins in general, the porins pose a number of interesting questions in addition to those connected with their physiological functions. In *E. coli* K-12 a whole family of such proteins of similar size and properties exists, and it seems that, under usual laboratory conditions, several of the corresponding structural genes are silent or nearly so (10–13). These genes are not clustered on the *E. coli* chromosome, and if they have arisen by duplications they might allow some insight into the evolution of this chromosome (11, 13–17). Furthermore, several of these proteins can serve as at least parts of phage receptors (18–21) and they are required for an apparent uptake of protein into the cell: mutants lacking certain such polypeptides are highly tolerant to several colicins (e.g., refs. 22 and 23). It is not known what constitutes a phage receptor area on such proteins and what is their function in colicin sensitivity.

Finding answers to all these and related questions would be helped by a knowledge of the amino acid sequence of such a protein. It is also likely that the gene for the protein under study will soon become available by DNA cloning. The determination of the DNA sequence, of much interest because of the unknown control region(s), should of course also be much aided by knowledge of the primary structure of the protein. We have determined the sequence of protein I from *E. coli* B/r and here present its primary structure. [Strains of *E. coli* B/r can differ in regard to the type of protein I produced; the protein used for this study is polypeptide Ia (24).]

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.

RESULTS AND DISCUSSION

The amino acid sequence proposed for protein I (the porin from *E. coli* B/r) is shown in Fig. 1. The protein contains 340 amino acid residues and from this composition (Table 1) a molecular weight of 37,205 is calculated. The composition derived from the sequence and that calculated from amino acid analyses of acid hydrolyzates of the whole protein agree well (Table 1), and the molecular weight calculated from the sequence is in excellent agreement with earlier estimates (7). Protein I is the largest integral membrane polypeptide that has so far been sequenced.

The primary structure of protein I was derived by analyses of peptides obtained by cleavage with cyanogen bromide (CNBr), trypsin, thermolysin, and the *Staphylococcus aureus* protease specific for glutamic acid residues (26). The strategy was first to determine the alignment of the four CNBr fragments (27) [CNBr1 (38 residues), CNBr2 (76 residues), CNBr3 (193 residues), and CNBr4 (33 residues)] and then to establish their amino acid sequences, starting with small, soluble peptides. Additional information was gained from overlapping tryptic peptides obtained from protein I with citraconylated (28) lysine residues. Sequencing was performed exclusively by manual methods, normally with the micro-dansyl-Edman technique (29, 30). For the assignment of glutamine or asparagine, the 4-*N*,*N*-dimethylaminoazobenzene 4′-isothiocyanate method introduced by Chang and Creaser (31) was refined to microscale and found to be much superior to the identification of the corresponding 3-phenyl-2-thiohydantoins. In some cases, glutamic acid and its amide were distinguished by their electrophoretic mobilities (32) on cellulose thin-layer plates. The details concerning peptide isolation and sequence determination will be published elsewhere (for most of the methods used, see refs. 27 and 30).

The major difficulties in the determination of the primary structure of protein I were the often rather low yields of tryptic peptides and the strong tendency of larger peptides to aggregate. For example, the CNBr fragments could not be separated by chromatography on Sephadex G-100 in 8 M urea although they are soluble under this condition, and this separation only became possible upon citraconylation of these fragments (27). Even then, and for unknown reasons, resolution of the COOH- and NH_2-terminal fragments CNBr1 and CNBr4, respectively, could not be reproduced in each chromatography run. A methodological generalization was not possible: several larger tryptic peptides showed increased aggregation upon citraconylation. The micromethods used (see above), which generally required only nanomolar quantities of peptides, were therefore of great help in the establishment of the sequence.

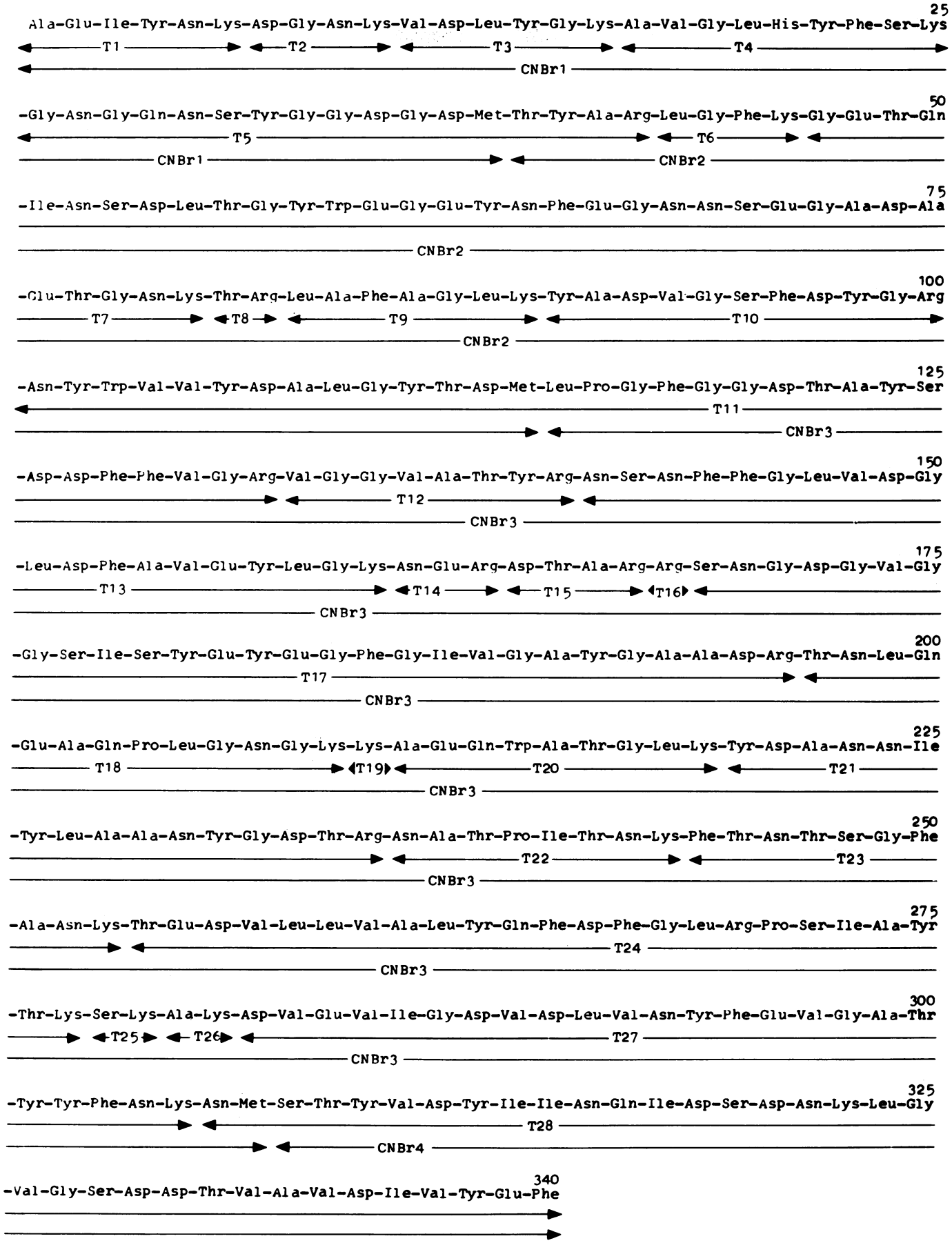

FIG. 1. Amino acid sequence of protein I. CNBr, cyanogen bromide fragments; T, tryptic peptides.

Table 1. Amino acid composition of protein I

Amino acid	Sequence	No. of residues found* 1	2	3
Asp	31	57	54	49
Asn	28			
Thr	21	21	21	20
Ser	16	15	17	16
Glu	17	27	27	26
Gln	7			
Pro	4	4	5	7
Gly	46	45	45	43
Ala	29	30	30	30
Cys	0	ND	1	1
Val	24	22	22	22
Met	3	2	4	4
Ile	12	12	11	13
Leu	21	21	21	23
Tyr	29	29	25	20
Phe	19	20	18	17
His	1	1–2	2	2–3
Lys	18	18	18	17
Arg	11	11	12	11
Trp	3	ND	3	5

* Analyses: 1, Hydrolysate of protein I used for sequence analysis; 2, matrix protein from *E. coli* BE (7); 3, protein I from *E. coli* B/r (from ref. 25). ND, not determined.

The protein from *E. coli* B/r and BE has been reported to contain one cysteine residue (7, 25) which was not found in the sequence and which has been reported to be absent from the corresponding protein from *E. coli* K-12 (33). Although treatment of our protein I with [^{14}C]iodoacetamide caused radioactive labeling of the protein, much less than 1 mol of reagent was incorporated and a clearly labeled radioactive tryptic peptide was not found. We believe, therefore, that the protein investigated does not contain cysteine.

Diedrich and Schnaitman (34) have recently reported that some of the lysine residues of this protein from *E. coli* K-12 are posttranslationally modified and exist as allysine (α-aminoadipic acid δ-semialdehyde) residues. Identification of all lysine residues in the sequence was unambiguous, and we have not found any evidence for the occurrence of allysine. This apparent contradiction may be attributable either to the absence of modified lysine residues from the *E. coli* B/r protein or to the more likely possibility (34) that only a relatively small fraction of a lysine at a given position is modified, thus escaping detection.

A few features of the polypeptide chain stand out. In contrast to phage M13 coat protein and glycophorin (transmembrane proteins with known primary structure; see below), protein I does not contain a contiguous sequence of ≈20 hydrophobic residues that might span the membrane by interacting with its hydrocarbon phase. The longest nonpolar sequence that is uninterrupted by charged residues is that from residues 184 through 194. Indeed, it seems unlikely that a protein that forms a membrane pore, probably as a trimer (35–38), would simply span the membrane with a linear lipophilic segment. On the other hand, the chain may be subdivided into three regions with differing degrees of polarity (39): a distinctly hydrophilic region from residues 1 to 82 (51.2% polarity), a fairly nonpolar region from residues 83 to 194 (33.9% polarity), and a more hydrophilic region up to the COOH-terminus (48% polarity). It remains to be seen if a rather simplistic view has anything to recommend it—namely, that the first (NH_2-terminal) region is located outside the outer membrane and that at least portions of the more lipophilic part interact with lipid components of that membrane. Another conspicuous property of the sequence consists in the distribution of lysine and arginine residues. The former are clustered at the COOH and NH_2 termini whereas the latter are arranged predominantly in the middle of the chain. Also, a rather uneven distribution of glutamate and aspartate residues can be seen. These properties, for which functional or topological interpretation is not yet possible, are shown in Fig. 2.

Calculation of the α-helix and β-sheet conformational features of the protein (40) yielded 19.7% for the former and 37.7% for the latter type of secondary structure with much of the β-sheet regions in the COOH-terminal half of the polypeptide. However, circular dichroism and infrared spectroscopy indicate that a larger fraction of the polypeptide exists in β-conformation and that almost 20% α-helix certainly does not appear to be at all realistic (ref. 7; J. P. Rosenbusch, personal communication). It seems, therefore, that the predictive rules derived mainly from studies on globular hydrophilic proteins are not necessarily applicable to other types of proteins.

So far, the primary structures of integral membrane proteins have been established only in a few cases—lipoprotein from *E. coli* (41); erythrocyte glycophorin (42); phage M13 coat protein (43–45); hepatic cytochrome b_5 (46); subunit 9 of yeast mitochondrial ATPase (47); and bacteriorhodopsin (48). This limited information, of course, is insufficient to permit general conclusion. It certainly appears already, however, that rather different means of association of a protein with its membrane have been devised. It has been mentioned above that the porin lacks the conspicuous linear "membrane segment" present in glycophorin and the phage coat protein. Although content and distribution of polar, nonpolar, and hydrophilic residues of the ATPase subunit and bacteriorhodopsin show some overall similarity, these features are entirely different in, for example, glycophorin. In line with this difference is the fact that the membrane topology of bacteriorhodpsin (49) is certainly not at all similar to that of glycophorin (50, 51). In view of different

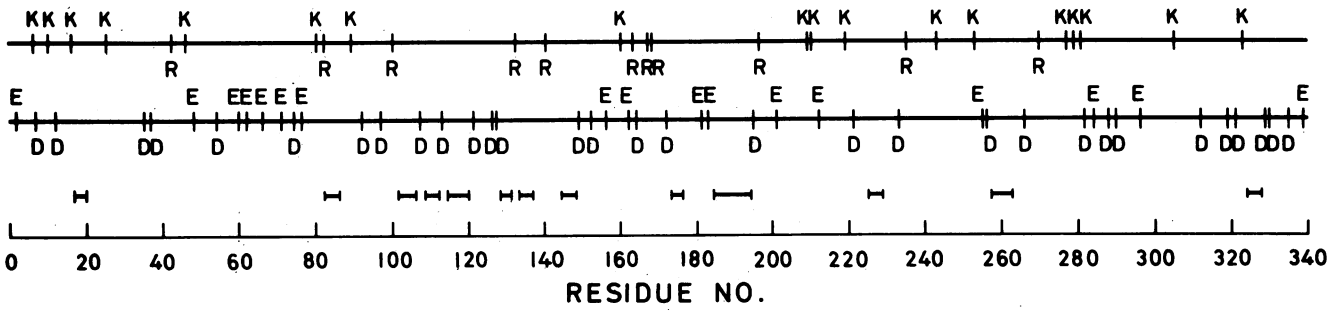

FIG. 2. Distribution of charged residues and hydrophobic sequences in protein I. Top line: lysine (K) and arginine (R) residues. Middle line: glutamate (E) and aspartate (D) residues. Bottom line: hydrophobic sequences equal to or exceeding four residues (the nonpolar residues alanine and glycine are included).

functions, these differences are not really unexpected but should probably be stressed because the glycophorin sequence may have been somewhat overemphasized in regard to the biosynthetic incorporation of such a protein into its membrane (52). It is also generally assumed that at least one domain of an integral membrane protein interacts with the hydrocarbon phase of the membrane. This need not be so; it is just as conceivable that even a transmembrane polypeptide is anchored to the membrane, without direct lipid contact, by other membrane proteins. The sequence information required to establish such an (hypothetical) arrangement could then be quite different from that presumably operating in, for example, the case of the M13 coat protein.

We express our deep gratitude to the sponsor of this paper, F. Lynen, who died on 6 Aug. 1979. We thank Dr. U. Chen-Schmeisser for performing the Chou-Fasman calculation, Dr. I. Crowlesmith for his critical review of the manuscript, Mr. M. Klaus for excellent technical assistance, and the Fonds der Chemischen Industrie for financial support.

1. Lodish, H. F. & Rothman, J. E. (1979) *Sci. Am.* **240** (1), 38–53.
2. DiRienzo, J. M., Nakamura, K. & Inouye, M. (1978) *Ann. Rev. Biochem.* **47**, 481–532.
3. Nakae, T. (1976) *J. Biol. Chem.* **251**, 2176–2178.
4. Nakae, T. (1976) *Biochem. Biophys. Res. Commun.* **71**, 877–884.
5. Bavoil, P., Nikaido, H. & von Meyenburg, K. (1977) *Mol. Gen. Genet.* **158**, 23–33.
6. Garten, W., Hindennach, I. & Henning, U. (1975) *Eur. J. Biochem.* **59**, 215–221.
7. Rosenbusch, J. P. (1974) *J. Biol. Chem.* **249**, 8019–8029.
8. Benz, R., Janko, K., Boos, W. & Läuger, P. (1978) *Biochim. Biophys. Acta* **511**, 305–319.
9. Schindler, H. & Rosenbusch, J. P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3751–3755.
10. Foulds, J. & Chai, T.-J. (1978) *J. Bacteriol.* **133**, 1478–1483.
11. Pugsley, A. P. & Schnaitman, C. A. (1978) *J. Bacteriol.* **135**, 1118–1129.
12. Van Alphen, W., van Selm, N. & Lugtenberg, B. (1978) *Mol. Gen. Genet.* **159**, 75–83.
13. Henning, U., Schmidmayr, W. & Hindennach, I. (1977) *Mol. Gen. Genet.* **154**, 293–298.
14. Ichihara, S. & Mizushima, S. (1978) *J. Biochem. (Tokyo)* **83**, 1095–1100.
15. Gamon, K., Chen, R. & Henning, U. (1978) *Mol. Gen. Genet.* **166**, 187–192.
16. Zipkas, D. & Riley, M. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1354–1358.
17. Foulds, J. & Chai, T.-J. (1978) *J. Bacteriol.* **136**, 501–506.
18. Datta, D. B., Arden, B. & Henning, U. (1977) *J. Bacteriol.* **131**, 821–829.
19. Verhoef, C., de Graaff, P. J. & Lugtenberg, E. J. J. (1977) *Mol. Gen. Genet.* **150**, 103–105.
20. Chai, T.-J. & Foulds, J. (1978) *J. Bacteriol.* **135**, 164–170.
21. Hantke, K. (1978) *Mol. Gen. Genet.* **164**, 131–135.
22. Chai, T. & Foulds, J. (1974) *J. Mol. Biol.* **85**, 465–474.
23. Chai, T.-J. & Foulds, J. (1977) *J. Bacteriol.* **130**, 781–786.
24. Schmitges, C. J. & Henning, U. (1976) *Eur. J. Biochem.* **63**, 47–52.
25. Garten, W. & Henning, U. (1974) *Eur. J. Biochem.* **47**, 343–352.
26. Drapeau, G. R., Boily, Y. & Houmard, I. (1972) *J. Biol. Chem.* **247**, 6720–6726.
27. Chen, R., Hindennach, I. & Henning, U. (1978) *Hoppe-Seyler's Z. Physiol. Chem.* **359**, 1807–1810.
28. Dixon, H. B. F. & Perham, R. N. (1968) *Biochem. J.* **109**, 312–314.
29. Gray, W. R. & Hartley, B. S. (1963) *Biochem. J.* **89**, 379–380.
30. Chen, R. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* **357**, 873–886.
31. Chang, J. Y. & Creaser, E. H. (1976) *Biochem. J.* **157**, 77–85.
32. Offord, R. E. (1966) *Nature (London)* **311**, 591–593.
33. Diedrich, D., Summers, A. O. & Schnaitman, C. A. (1977) *J. Bacteriol.* **131**, 598–607.
34. Diedrich, D. L. & Schnaitman, C. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3708–3712.
35. Palva, E. T. & Randall, L. L. (1978) *J. Bacteriol.* **133**, 279–286.
36. Steven, A. C., Ten Heggeler, B., Müller, R., Kistler, J. & Rosenbusch, J. P. (1977) *J. Cell. Biol.* **72**, 292–301.
37. Nakae, T., Ishii, J. & Tokunaga, M. (1979) *J. Biol. Chem.* **254**, 1457–1461.
38. Yu, F., Ichihara, S. & Mizushima, S. (1979) *FEBS Lett.* **100**, 71–74.
39. Capaldi, R. A. & Vanderkooi, G. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 930–932.
40. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–222.
41. Braun, V. & Bosch, V. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 970–974.
42. Tomita, M., Furthmayr, H. & Marchesi, V. T. (1978) *Biochemistry* **17**, 4756–4770.
43. Asbeck, F., Beyreuther, K., Köhler, H., von Wettstein, G. & Braunitzer, G. (1969) *Hoppe Seyler's Z. Physiol. Chem.* **350**, 1047–1066.
44. Nakashima, Y. & Konigsberg, W. (1974) *J. Mol. Biol.* **88**, 598–600.
45. Wickner, W. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1159–1163.
46. Ozols, J. & Gerard, C. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3725–3729.
47. Wachter, E., Sebald, W. & Tsagoloff, A. (1977) in *Mitochondria 1977*, eds. Bandlow, W., Schweyen, R. J., Wolf, K. & Kaudewitz, F. (de Gruyter, Berlin), pp. 441–449.
48. Ovchinnikov, Yu. A., Abdulaev, N. G., Feigina, M. Yu., Kiselev, A. V. & Lobanov, N. A. (1979) *FEBS Lett.* **100**, 219–224.
49. Henderson, R. & Unwin, P. N. T. (1975) *Nature (London)* **257**, 28–32.
50. Bretscher, M. S. (1975) *J. Mol. Biol.* **98**, 831–833.
51. Cotmore, S. F., Furthmayr, H. & Marchesi, V. T. (1977) *J. Mol. Biol.* **113**, 539–553.
52. Rothman, J. E. & Lenard, J. (1977) *Science* **195**, 743–753.