Proc. Natl. Acad. Sci. USA
Vol. 77, No. 8, pp. 4592–4596, August 1980
Biochemistry

Primary structure of major outer membrane protein II* (*ompA* protein) of *Escherichia coli* K-12

(transmembrane protein/microscale sequence determination/phage receptor)

ROBERT CHEN, WALTRAUD SCHMIDMAYR, CORNELIA KRÄMER, URSULA CHEN-SCHMEISSER, AND ULF HENNING

Max-Planck-Institut für Biologie, Corrensstrasse 38, D-7400 Tübingen, West Germany

Communicated by David E. Green, May 8, 1980

ABSTRACT **The amino acid sequence of major outer membrane protein II* (*ompA* protein) from *Escherichia coli* K-12 has been determined. The transmembrane polypeptide consists of 325 residues, resulting in a molecular weight of 35,159. The transmembrane part of the protein is located between residues 1 and 177. In this part of the protein a predominantly lipophilic 27-residue segment exists that perhaps spans the membrane in a mostly α-helical conformation, or a 19-residue stretch of this segment might traverse the membrane linearly. Inside the outer membrane a sequence -Ala-Pro-Ala-Pro-Ala-Pro-Ala-Pro- exists that, analogous to the -Cys-Pro-Pro-Cys-Pro- sequence in the hinge region of immunoglobulin, could assume the conformation of a polyproline helix. Computer analysis did not reveal a clear overall pattern of internal homology in the protein; besides the -Ala-Pro- repeat, only one local area (two adjacent dodecapeptide segments) shows some repetitiveness. The same analysis did not produce evidence for internal homology in the previously determined sequence of outer membrane protein I (porin) nor was any marked resemblance detected between transmembrane proteins I and II*.**

Polypeptide II* (1) is one of the most abundant proteins of the outer membrane of the *Escherichia coli* cell envelope (for other such proteins and other nomenclature, see ref. 2). This transmembrane protein (3) is multifunctional: besides its nonphysiological function as a phage receptor (4, 5) and its requirement for the action of a colicin (6), it can serve as a mediator in F-dependent conjugation (5, 7, 8), and, in combination with the outer membrane lipoprotein (9), it is required for the structural integrity of this membrane and the generation of normal cell shape (10). Its role in the uptake of amino acids is controversial (11, 12). It may also have other functions.

We have previously presented the primary structure of another major outer membrane protein, the pore-forming transmembrane protein I (porin; e.g., see refs. 13 and 14) of *E. coli* B/r (15). We have extended this work to protein II* in order to answer a number of questions specifically about this polypeptide and about integral membrane proteins in general. Very little information exists concerning the details of association of such proteins with their membranes, and very little is known about relevant structure–function relationships. There are also a number of questions specific to protein II*, but of more general interest. Some examples follow. What constitutes a phage receptor area? The protein interacts with the outer membrane's lipopolysaccharide (16). Exactly which part of the protein recognizes this membrane component and how does the protein bind to it? The protein is synthesized in precursor form with an extended NH_2-terminal sequence (17, 18). Which rules of polypeptide folding are followed during membrane incorporation?

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 structural gene for protein II*, *ompA*, has recently been cloned (19). Knowledge of the amino acid sequence will be helpful for determination of the corresponding DNA sequence (of considerable interest because of the unknown mechanism of regulation of synthesis of this protein).

MATERIALS AND METHODS

Materials. Citraconic anhydride and ethyleneimine were purchased from Pierce, fluorescamine from Serva (Heidelberg, W. Germany), succinic anhydride from Merck, and heptafluorobutyric acid from Fluka (Buchs, Switzerland). Trypsin treated with L-(tosylamido-2-phenyl)ethyl chloromethyl ketone was obtained from Merck, thermolysin from Serva, and *Staphylococcus aureus* protease SV8 from Miles. Protein II* was isolated from *E. coli* K-12 strain P400 (8) according to a modification (4) of a procedure described earlier (20).

Preparation of Cyanogen Bromide Fragments. Protein II* (140 nmol) was treated with CNBr (≈10 μM) in 1 ml of 75% (vol/vol) formic acid for 24 hr at room temperature. Under the same conditions, the protein was cleaved at Met and Trp residues (21) with CNBr in 88% (vol/vol) formic acid/heptafluorobutyric acid, 1:1 (vol/vol). Fragments were separated on Sephadex G-75 superfine (2.5 × 150 cm column) in 50% (vol/vol) acetic acid and at a flow rate of 4.5 ml/hr (see Fig. 1).

Partial Tryptic Digestion of CNBr Fragments. Purified CNBr fragments (50 nmol/ml) were acylated with citraconic anhydride (22) and separated from excess reagent by gel filtration on Sephadex G-25 superfine (1 × 30 cm column) in 25 mM NH_4HCO_3. The same procedure was used for succinylation. Samples were digested with trypsin for 2 hr at 37°C at an enzyme-to-substrate ratio of 1:200. Peptides were deacylated [50% (vol/vol) acetic acid, 4 hr at room temperature] and chromatographed on Sephadex G-50 superfine (2.5 × 150 cm column) in 50% acetic acid.

Isolation of Peptides by Fingerprinting (Two-Dimensional Tryptic Chromatography). Peptides obtained by the action of various proteases were separated on cellulose thin-layer plates (Macherey & Nagel, Cel 300 MN; Düren, W. Germany) as described (23) except that electrophoresis in the first dimension was at pH 5.4 [2% (vol/vol) pyridine in 5% (vol/vol) acetone adjusted with acetic acid]. Positions of peptides were determined with Fluram spray (24). Peptides were eluted with 50% acetic acid as described (23).

Sequence Determination. Amino acid sequences were determined by the micro-dansyl–Edman technique (23). For assignments of glutamine or asparagine, the 4-*N*,*N*-dimethylazobenzene 4′-isothiocyanate method was used (25).

RESULTS

The general strategy for elucidation of the primary structure of protein II* was first to establish the sequence of the individual CNBr fragments and then to determine their order. Only selected examples for the alignment of peptides are given here; more detail is shown in Fig. 2.

Purification of CNBr Fragments. Chromatography of the CNBr fragments of protein II* resulted, before the appearance of the included volume, in the profile shown in Fig. 1. The corresponding NH_2-terminal residues were determined as: Leu (CNBr5), Val (CNBr4), Ala and Gly (CNBr1 and CNBr6), and Pro (CNBr3). Obviously, and as was confirmed later, fragments CNBr1 and CNBr6 eluted together. The smallest fragment, CNBr2, appeared with the included volume. It was separated from the marker for this volume (dinitrophenyllysine) by chromatography on Sephadex G-25 fine; its NH_2 terminus was Gly.

Sequence of CNBr5. With the exception of tryptic peptides T13 and T16, all corresponding tryptic peptides could be isolated from fingerprints on thin-layer plates. For establishment of the sequence encompassed by T13, citraconylated CNBr5 was digested with trypsin. Chromatography on Sephadex G-50 fine resolved four peaks. The largest peptide was CNBr5LB1, consisting of tryptic peptides T13–T17. Sequence determination (25 cycles) yielded the sequence of T13 and two residues of T14. To obtain the T16 sequence, a tryptic digest of citraconylated CNBr5 was succinylated, decitraconylated, and chromatographed on Sephadex G-50 superfine. The α-succinylated peptide T13–T17 was treated with trypsin and chromatographed on Sephadex G-25 superfine. Peptides T16 and T13 (the latter blocked) eluted in the first peak, and the sequence of T16 could be determined.

Sequence of CNBr4. The citraconylated peptide was digested with trypsin; the resulting peptides were chromatographed on Sephadex G-50 superfine. Sequence determination (25 cycles) of the first eluting 35-residue peptide CNBr4LB1 revealed the order T9-T10-T11. T12 eluted in the second peak. T11 could be isolated from a tryptic digest of deacylated CNBr4LB1.

Sequence of CNBr1 and CNBr6. When the mixture of these two fragments was obtained from succinylated protein II*, only Gly was found as the NH_2 terminus. Ala is the NH_2 terminus of the protein; this result indicated that one of the fragments was the NH_2-terminal fragment. The 41-residue peptide T3 was purified from tryptic peptides obtained from the mixture of the two fragments by chromatography on Sephadex G-50 superfine. Its sequence was then determined as indicated in Fig. 2, and all relevant peptides could be isolated by fingerprinting. The NH_2-terminal sequence of CNBr6 was obtained by sequence analysis of the mixture of CNBr1 and CNBr6 from succinylated protein. All tryptic peptides could be isolated from fingerprints. Peptide T29 lacks Arg, Lys, and homoserine. Thus we concluded that it was at the COOH terminus of protein II*.

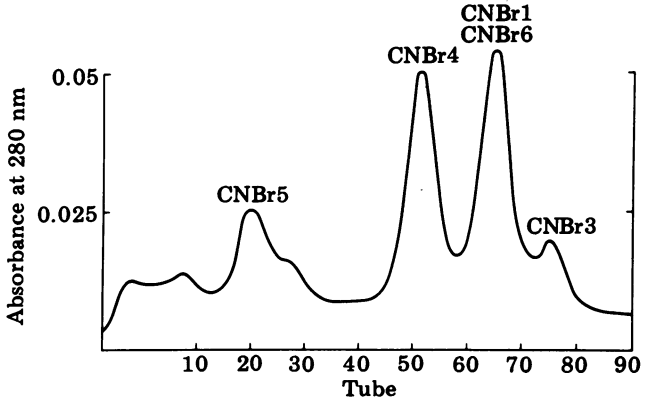

FIG. 1. Sephadex chromatography of CNBr fragments of protein II*. The small fragment CNBr2 eluted with the included volume (not shown, tubes 170–190). The shoulder of the CNBr5 peak contained a dimer of CNBr6 (see text).

Sequence of CNBr3. All tryptic peptides were isolated by fingerprinting. Sequence analysis of CNBr3 for 27 cycles established the order T4-T5-T6-T7. Peptide T8N (the NH_2-terminal moiety of T8) contained homoserine and was therefore placed at the COOH terminus of CNBr3.

Sequence of CNBr2. This fragment consisted of seven residues, and their sequence could be determined directly.

Alignment of CNBr Fragments. When protein II* was treated with trypsin, tyrosine in position 48 was cleaved to some degree, releasing two subfragments of tryptic peptide T3. The COOH-terminal part of T3 (T3C) provided the overlap of CNBr1-CNBr2. The alignment of CNBr2-CNBr3 was determined by peptide T4 and by one thermolytic peptide (PH1), obtained by digestion of protein II*. Fragment CNBr4 was placed after CNBr3 by the data obtained from peptide T8. Determination of the sequence of peptide T12 for 26 cycles established the overlap of CNBr4-CNBr5. The sequence of peptide T23 confirmed that CNBr6 came after CNBr5.

The amino acid sequence proposed for protein II* is shown in Fig. 2. The protein consists of 325 residues, and from this composition a molecular weight of 35,159 was calculated. The amino acid composition derived from the sequence agrees with earlier estimates (Table 1).

DISCUSSION

The anomalous mobility of protein II* upon $NaDodSO_4$/polyacrylamide gel electrophoresis can be explained as follows. The protein can be modified by heat (20, 27, 28); when it is boiled in the detergent, its apparent molecular weight is 33,000, whereas without the heat treatment a value of 28,000 is found. It has been suggested that the lower value may be closer to the true molecular weight (1); however, evidence to the contrary has been published (29). The higher apparent molecular weight is only 6% short of the value calculated in this paper. Thus, the modification by heat most likely reflects incomplete unfolding of the protein in the absence of heating in the detergent.

It had been reported that protein II* from *E. coli* K-12 might contain hexosamine whereas no amino sugar was found in this protein from *E. coli* B/r (1). The data reported here do not show an amino acid residue substituted correspondingly. The protein therefore does not appear to be glycosylated.

The small peak eluting as a shoulder of that of CNBr5 apparently contained a dimer of CNBr6; tryptic digestion of part of the corresponding material essentially generated peptides belonging to CNBr6. This fragment contains the two cysteine residues of the protein and, thus, the dimer most likely represents the disulfide. We have found no evidence that interchain disulfides exist *in vivo*. $NaDodSO_4$/polyacrylamide gel electrophoresis of cell envelopes reduced with 2-mercaptoethanol or not reduced did not show differences in staining intensities of protein II* (i.e., the band moving with an apparent molecular weight of 33,000), and there was no evidence of a dimer of the protein.

In contrast to some other transmembrane proteins, outer membrane protein I (porin) does not contain a contiguous sequence of ≈20 hydrophobic residues that might span the membrane (15). This situation may be different in protein II*. In cell envelopes, protein II* can be partially degraded by proteases; e.g., trypsin and Pronase generate fragments re-

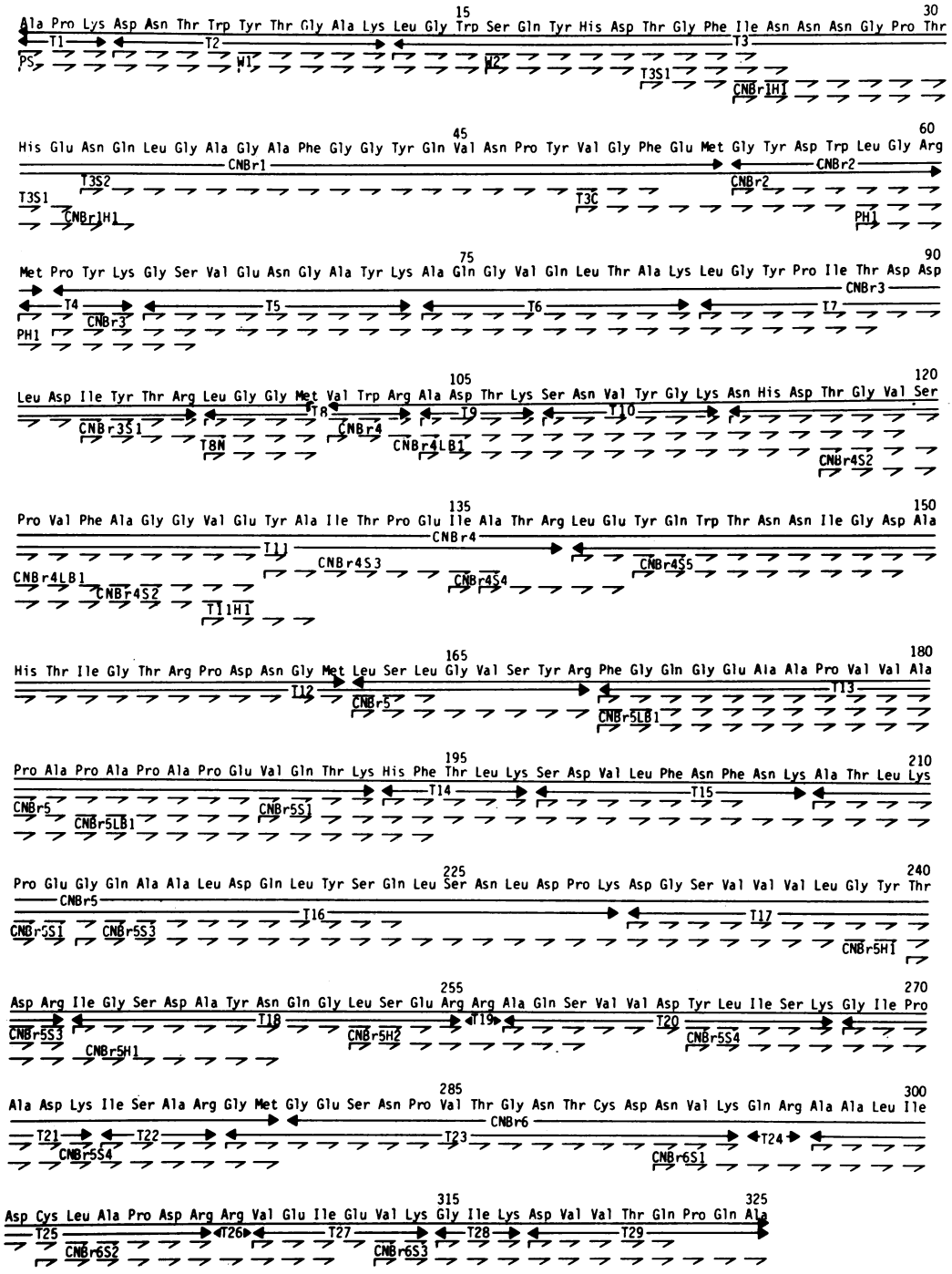

FIG. 2. Amino acid sequence of protein II*. CNBr, cyanogen bromide fragments; T, tryptic peptides; H, thermolysine peptides; S, *S. aureus* protease peptides; LB, lysine-blocked (citraconylated) tryptic peptides; W, peptides resulting from CNBr cleavage at Trp and Met residues; PS, sequence determined from the complete protein; PH1, thermolysin peptide obtained by digestion of the complete protein; T3C, tryptic peptide originating from trypsin cleavage of the complete protein (see text); T8N, tryptic peptide obtained from CNBr3. Arrows (—) indicate the cycles of micro-dansyl–Edman degradations. The entire sequence has recently been confirmed by sequence determination of the corresponding cloned DNA (19, 26). The DNA sequence did not agree, at four positions, with the amino acid sequence originally determined. These discrepancies were resolved at three positions in favor of the DNA sequence and at one position in favor of the protein sequence.

maining associated with the outer membrane and exhibiting apparent molecular weights of 24,000 and 19,000, respectively (16). These fragments represent the NH_2-terminal portion of the polypeptide, they remain active as phage receptors, and the parts of the protein removed by the proteases are exposed inside the outer membrane (16). We now have determined the site of cleavage by Pronase. The Pronase fragment of protein II* was isolated (30), citraconylated, and digested with trypsin, and the peptides were separated on Sephadex G-50 superfine in 50% acetic acid. The large tryptic peptide encompassing residues 170–242, known from analyses of lysine-blocked CNBr5, was missing. Instead, a peptide with the sequence Phe-Gly-Gln-Gly-Glu-Ala-Ala-Pro was recovered—i.e., Pronase cleaves at residue 177. Therefore, the transmembrane part(s) of the protein must reside between residues 1 and 177. (The molecular weight of the Pronase fragment calculated from the sequence,

Table 1. Amino acid composition of protein II*

Amino acid	No. of residues found by sequence analysis	No. of residues found by analyses of acid hydrolysates[a] A	B
Asp	22	38	41
Asn	19		
Thr	21	20	20
Ser	16	15	14
Glu	14	29	29
Gln	15		
Pro	19	18	19
Gly	37	35	35
Ala	29	28	31
Cys	2	2	1
Val	25	23	25
Met	5	6	4
Ile	14	14	14
Leu	22	23	22
Tyr	17	16	17
Phe	8	8	10
His	5	6	6
Lys	17	18	18
Arg	13	13	14
Trp	5	6	5

[a] A, data published earlier (1) and recalculated for a molecular weight of ≈35,000; B, data from Van Alphen *et al.* (5) calculated for 325 residues.

19,282, agrees with the value mentioned above and determined electrophoretically. Because there is not such close agreement for the molecular weight of the whole protein, it is possible that even boiling in $NaDodSO_4$ does not completely unfold the polypeptide chain. In accord with this notion is the fact that neither the tryptic nor the Pronase fragment is able to be modified by heat.)

A linear lipophilic segment 19 residues long may exist at positions 33–51 (although it contains two glutamine and two asparagine residues). Another possible lipophilic segment may be considered in view of evidence from studies of bacteriorhodopsin. On the basis of its three-dimensional structure (31), a model for the accommodation of the amino acid sequence (32, 33) of the protein in the purple membrane of halophilic bacteria has been proposed (32). Each of the seven α-helical segments of bacteriorhodopsin spanning the membrane consists of ≈28 residues, and almost all of these segments contain charged residues. The rather lipophilic 27-residue segment in protein II* from position 33 to position 59 (containing two aspartic acid residues) possesses only one proline residue and perhaps may be accommodated, mostly in an α-helical conformation, by the outer membrane. Then the following sequence, -Arg-Met-Pro-Tyr-Lys-, could support this hypothesis. The sequences that occur after the proposed transmembrane segments of glycophorin and phage M13 coat protein (transmembrane proteins with known primary structure) are -Arg-Arg-Leu-Ile-Lys-Lys- and -Lys-Leu-Phe-Lys-Lys-, respectively (34, 35).

The most conspicuous part of the sequence of protein II* is the stretch from residue 171 to residue 187 containing the 4-fold -Ala-Pro- repeat. This sequence, exposed inside the outer membrane, may assume the conformation of a polyproline helix. It is interesting that the inter-heavy-chain disulfide-linked -Cys-Pro-Pro-Cys-Pro- sequences (36) in the hinge region of an immunoglobulin (IgG1 Kol) exist as two parallel polyproline helices (refs. 37 and 38; R. Huber, personal communication). To draw further analogies to the immunoglobulins is highly speculative, but the comparable susceptibility to enzymatic proteolysis (generation of Fab fragments) may allow us to consider that this region of protein II* also has segmental flexibility.

Finally, because of the striking repetitiveness of the outer membrane lipoprotein sequence (9), the sequence of protein II* and that previously published for protein I (15) have been tested for internal homologies and similarities between the two polypeptides by computer analysis (program RELATE performed by W. C. Barker, National Medical Research Foundation, Washington, DC) with the mutation data matrix (39, 40, 41). For protein I, no evidence for internal repetitiveness was found. There was also no clear overall pattern of internal homology in protein II*. Besides the -Ala-Pro- repeat, one local area showed some repetitiveness: the adjacent dodecapeptide segments 127–138 and 139–150. For two other segments (residues 33–69 and 119–155) the score for possible internal duplication was almost significant. Any marked resemblance between proteins I and II* was not detected.

We thank Dr. Robert Huber for drawing our attention to the immunoglobulin polyproline helix and Dr. Colin Manoil for critically reviewing the manuscript. We are grateful to Dr. W. C. Barker for performing the computer analyses. This work was supported by the Fonds der Chemischen Industrie.

1. Garten, W., Hindennach, I. & Henning, U. (1975) *Eur. J. Biochem.* **59**, 215–221.
2. DiRienzo, J. M., Nakamura, K. & Inouye, M. (1978) *Annu. Rev. Biochem.* **47**, 481–532.
3. Endermann, R., Krämer, C. & Henning, U. (1978) *FEBS Lett.* **86**, 21–24.
4. Datta, D. B., Arden, B. & Henning, U. (1977) *J. Bacteriol.* **131**, 821–829.
5. Van Alphen, L., Havekes, L. & Lugtenberg, B. (1977) *FEBS Lett.* **75**, 285–290.
6. Chai, T. & Foulds, J. (1974) *J. Mol. Biol.* **85**, 465–474.
7. Schweizer, M. & Henning, U. (1977) *J. Bacteriol.* **129**, 1651–1652.
8. Skurray, R. A., Hancock, R. E. W. & Reeves, P. (1974) *J. Bacteriol.* **119**, 726–735.
9. Braun, V. (1975) *Biochim. Biophys. Acta* **415**, 335–337.
10. Sonntag, I., Schwarz, H., Hirota, Y. & Henning, U. (1978) *J. Bacteriol.* **136**, 280–285.
11. Manning, P. A., Pugsley, A. P. & Reeves, P. (1977) *J. Mol. Biol.* **116**, 285–300.
12. Lugtenberg, B., van Boxtel, R., Verhoef, C. & van Alphen, W. (1978) *FEBS Lett.* **96**, 99–105.
13. Nakae, T. (1976) *Biochem. Biophys. Res. Commun.* **71**, 877–884.
14. Nikaido, H. (1979) *Angew. Chem.* **91**, 394–407.
15. Chen, R., Krämer, C., Schmidmayr, W. & Henning, U. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5014–5017.
16. Schweizer, M., Hindennach, I., Garten, W. & Henning, U. (1978) *Eur. J. Biochem.* **82**, 211–217.
17. Sekizawa, J., Inouye, S., Halegoua, S. & Inouye, M. (1977) *Biochem. Biophys. Res. Commun.* **77**, 1126–1133.
18. Movva, N. R., Nakamura, K., & Inouye, M. (1980) *J. Biol. Chem.* **255**, 27–29.
19. Henning, U., Royer, H.-D., Teather, R. M., Hindennach, I. & Hollenberg, C. P. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4360–4364.
20. Hindennach, I. & Henning, U. (1975) *Eur. J. Biochem.* **59**, 207–213.
21. Ozols, J. & Gerard, C. (1977) *J. Biol. Chem.* **252**, 5986–5989.
22. Dixon, H. B. F. & Perham, R. N. (1968) *Biochem. J.* **109**, 312–314.
23. Chen, R. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* **357**, 873–886.
24. Udenfriend, S. Stein, S., Bohlen, P., Dairman, W., Leimgruber, W. & Weigele, M. (1972) *Science* **178**, 871–872.

25. Chang, J. Y. & Creaser, E. H. (1976) *Biochem. J.* **157**, 77–85.
26. Beck, E. & Bremer, E. (1980) *Nucleic Acids Res.*, in press.
27. Schnaitman, C. A. (1974) *J. Bacteriol.* **118**, 442–453.
28. Reithmeier, R. A. F. & Bragg, P. D. (1974) *FEBS Lett.* **41**, 195–198.
29. Heller, K. B. (1978) *J. Bacteriol.* **134**, 1181–1183.
30. Garten, W. & Henning, U. (1974) *Eur. J. Biochem.* **47**, 343–352.
31. Henderson, R. & Unwin, P. N. T. (1975) *Nature (London)* **257**, 28–32.
32. Ovchinnikov, Y. A., Abdulaev, N. G., Feigina, M. Y., Kiselev, A. V. & Lobanov, N. A. (1979) *FEBS Lett.* **100**, 219–224.
33. Khorana, H. G., Gerber, G. E., Herlihy, W. C., Gray, C. P., Anderegg, R. J., Nihei, K. & Biemann, K. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5046–5050.
34. Tomita, M., Furthmayr, H. & Marchesi, V. T. (1978) *Biochemistry* **17**, 4756–4770.
35. Wickner, W. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1159–1163.
36. Edelman, G. M., Cunningham, B. A., Gall, W. E., Gottlieb, P. D., Rutishauser, U. & Waxdal, M. J. (1969) *Proc. Natl. Acad. Sci. USA* **63**, 78–85.
37. Colman, P. M., Deisenhofer, J., Huber, R. & Palm, W. (1976) *J. Mol. Biol.* **100**, 257–282.
38. Huber, R., Deisenhofer, J., Colman, P. M., Matsushima, M. & Palm, W. (1976) *Nature (London)* **264**, 415–420.
39. Dayhoff, M. O. (1976) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 5–7.
40. Dayhoff, M. O. (1976) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 311–312.
41. Barker, W. C., Ketcham, L. K. & Dayhoff, M. O. (1978) *J. Mol. Evol.* **10**, 265–281.