THE PRIMARY STRUCTURE OF PROTEIN S3 FROM THE SMALL RIBOSOMAL SUBUNIT OF *ESCHERICHIA COLI*

Dieter BRAUER and Roswitha RÖMING
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, D-1000 Berlin-Dahlem, Germany

Received 29 August 1979

1. Introduction

Protein S3 from the small *Escherichia coli* ribosomal subunit is involved in several ribosomal functions such as f-Met–$tRNA_f^{Met}$ binding [1], and it is part of one binding site of dihydrostreptomycin [2]. Using immunoelectron microscopy it has been shown that there are two binding sites on the surface of the 30 S subunit for specific antibodies raised against S3 (reviewed in [3,4]). This finding is consistent with results from low-angle X-ray scattering [5] and from hydrodynamic studies [6,7] which show that protein S3 has an elongated shape.

Protein S3 has been found to be associated with an RNA region comprising about 900 nucleotides at the 5′-end of 16 S RNA [8] and can be crosslinked with ribosomal proteins S2, S4 and S10 [9,10]. This paper describes the complete primary structure of protein S3.

2. Materials and methods

Protein S3 was isolated from *E. coli* strain K as in [11] and was provided by Dr H. G. Wittmann. The primary structure was determined by combined enzymic and chemical digestions.

2.1. *Enzymic digestions*

Trypsin, treated with 1-chloro-4-phenyl-3-tosyl-amidobutan-2-one (Merck, Darmstadt) was used for enzymic cleavage of the intact protein both with and without reversible modification of the protein by protecting the ε-amino groups of the lysine residues with ETPA (exo-*cis*-3,6-endoxo-Δ^4-tetrahydrophthalic anhydride) or after irreversibly modifying the arginine residues with 1,2-cyclohexanedione [12,13]. All tryptic cleavages were performed at pH 8 for 4 h at 37°C.

Digestion with thermolysin (Serva, Heidelberg) was carried out at pH 8 for 90 min at 50°C, and with *Staphylococcus aureus* protease from Miles Biochemicals (Frankfurt/Main) in 0.05% ammonium acetate buffer (pH 4) for 48 h at 37°C [14].

2.2. *Chemical digestion*

Specific cleavage of the protein at its tryptophanyl bonds was achieved by treatment with BNPS–skatole (2-(2-nitrophenylsulfonyl)-3-methyl-3′-bromidolenin) [15] for 40 h at room temperature.

2.3. *Isolation of peptides and amino acid analyses*

Peptides were separated either by gel filtration on Sephadex G-50 and G-75 (superfine; 170 × 1, 140 × 1 cm) in 10% acetic acid [14] or on a microcolumn of Dowex M71 (0.3 × 10 cm) at 55°C using pyridine formate gradients [14,16]. The fractions were analysed either photometrically at 280 nm or by spotting aliquots of every second fraction onto cellulose plates, with subsequent ascending chromatography. If further separation procedures were necessary, the methods of preparative chromatography, electrophoresis or the fingerprint technique on cellulose thin-layer plates [14] were used. These latter proce-

dures, especially the fingerprint technique, were only used when sequencing according to the new sensitive double-coupling method [17] was performed. All amino acid analyses were carried out with a Durrum D-500 analyser. For estimating the cysteine content the protein was oxidized with performic acid before acid hydrolysis. The presence of tryptophan was tested by spraying the fingerprint plates with *p*-aminobenzaldehyde [18] and by the photometric method [19] carried out by Dr P. Woolley.

2.4. *Sequencing methods*

Both automated and manual techniques were used for the determination of amino acid sequences.

(a) Automatic Edman degradation [20] of intact protein was performed in an improved Beckman sequenator [21] for the N-terminal sequence [22].

(b) Manual Edman degradation was combined with dansylation of the free N-terminal amino acid residue after each degradation step [23]. Aspartic acid, glutamic acid and their corresponding amides were determined by converting the released 2-anilino-5-thiazolinone derivatives to their corresponding phenylthiohydantoin derivatives. These were identified by ascending chromatography [22]. Ambiguities in the identification of dansylated amino acids were resolved by liberation of the free amino acids from their 2-anilino-5-thiazolinones or phenylthiohydantoin derivatives by hydrolysis and amino acid analysis.

(c) Manual Edman-type degradation was performed using the double-coupling method with 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate [17]. Nearly all peptides were sequenced by both manual methods.

3. Results and discussion

The primary structure of protein S3 could be established by various digestions producing the required overlapping peptides. The N-terminal region of the protein was determined by automatic Edman-degradation in an improved Beckman sequenator [22]. As shown in fig.1 the most important cleavages in the sequence analyses of protein S3 were the tryptic digestion (supplemented by ETPA-modification of lysine residues), digestion with *Staphylococcus aureus* protease and chemical cleavage with BNPS–skatole. The tryptic peptides were separated by either ion-exchange chromatography or gel-filtration followed in both cases by the fingerprint technique. The peptides obtained by SP-cleavage and from BNPS–skatole treatment were separated by gel filtration on Sephadex G-50 (SP) and Sephadex G-75 (skatole). Nearly all peptides isolated were sequenced with both manual methods. The long peptides (BNPS, CHD and some SP peptides) were sequenced manually as far as possible, and analyses were completed by treating with other enzymes and separation by the fingerprint technique. Application of carboxypeptidase B resulted in a nearly equimolar amount of Lys and Arg in C-terminal analyses. Combination of all the results (which will be detailed elsewhere) led to the complete primary structure determination of S3 as shown in fig.1.

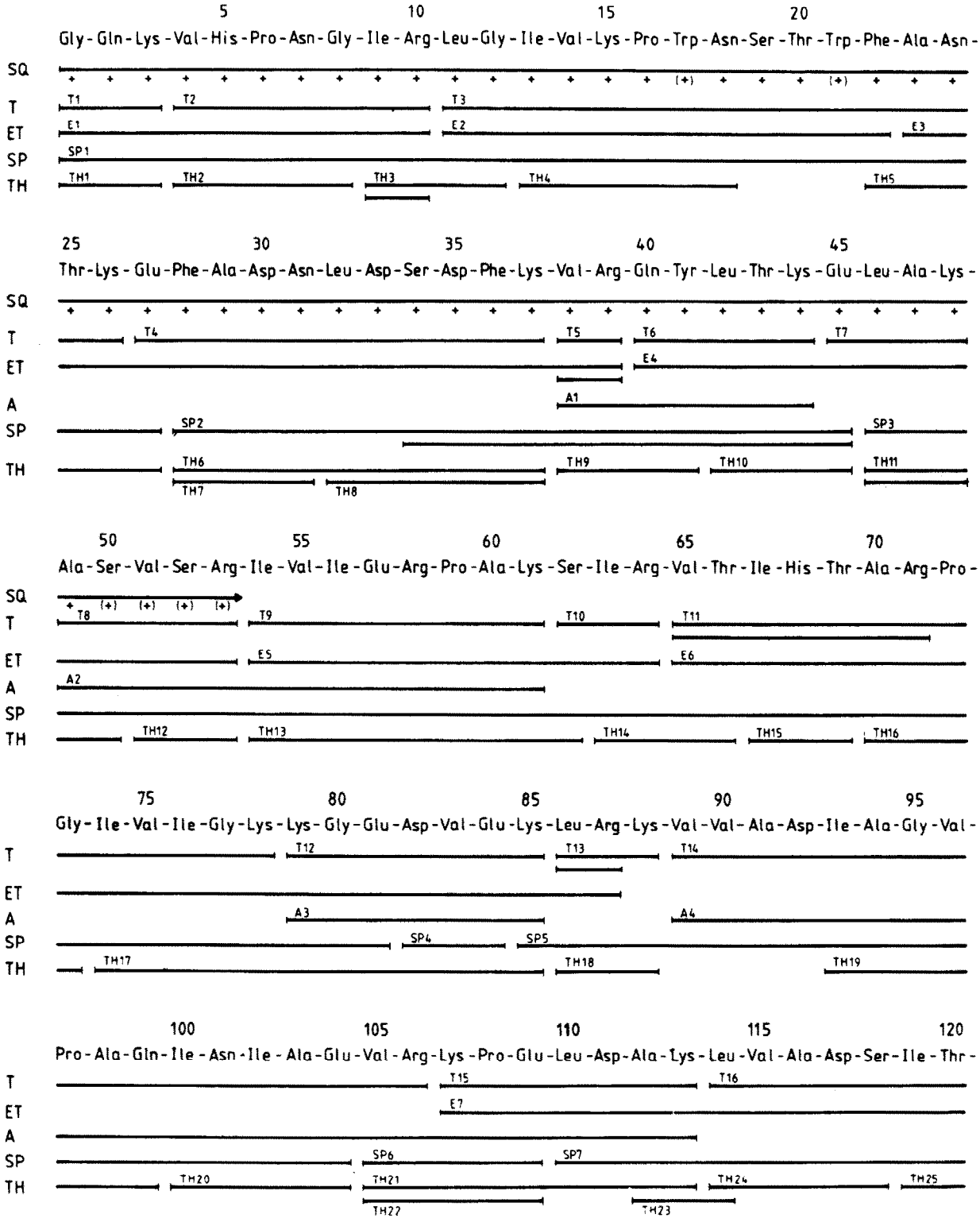
The protein consists of 232 amino acid residues with the following amino acid composition:

Asp_9, Asn_7, Thr_{11}, Ser_9, Glu_{16}, Gln_9, Pro_{10}, Gly_{19}, Ala_{24}, Val_{22}, Met_4, Ile_{18}, Leu_{13}, Tyr_4, Phe_5, His_4, Lys_{23}, Arg_{21} and Trp_4.

This composition leads to mol. wt 25 852, which is in excellent agreement with that determined by hydrodynamic methods, namely 25 800 (± 500) [7].

Our sequence analysis gave no hint of the presence of modified amino acids, e.g., no unusual spots were observed by different manual degradation methods and no unusual peaks were observed during the amino acid analysis. The presented primary structure of S3 corresponds to the longer form of this protein. Differences to a second form of S3 with lower molecular weight, which can be isolated by chromatographic procedures (H. G. Wittmann, personal communication) are under investigation and will be detailed elsewhere.

In fig.2 the result of secondary structure prediction of protein S3 by four different methods is presented. This procedure was carried out by Dr B. Wittmann-Liebold and is detailed in [24]. In the line 'PRE' of fig.2, regions for helix, extended structure and β-turns are indicated from at least three out of four methods in agreement. Based on this prediction, protein S3 was predicted to contain at least 28% helix, 12% extended structure and 15% β-turns.


5 10 15 20
Gly-Gln-Lys-Val-His-Pro-Asn-Gly-Ile-Arg-Leu-Gly-Ile-Val-Lys-Pro-Trp-Asn-Ser-Thr-Trp-Phe-Ala-Asn-
SQ
T
ET
SP
TH
T1 T2 T3
E1 E2 E3
SP1
TH1 TH2 TH3 TH4 TH5
25 30 35 40 45
Thr-Lys-Glu-Phe-Ala-Asp-Asn-Leu-Asp-Ser-Asp-Phe-Lys-Val-Arg-Gln-Tyr-Leu-Thr-Lys-Glu-Leu-Ala-Lys-
SQ
T
ET
A
SP
TH
T4 T5 T6 T7
E4
A1
SP2 SP3
TH6 TH7 TH8 TH9 TH10 TH11
50 55 60 65 70
Ala-Ser-Val-Ser-Arg-Ile-Val-Ile-Glu-Arg-Pro-Ala-Lys-Ser-Ile-Arg-Val-Thr-Ile-His-Thr-Ala-Arg-Pro-
SQ
T
ET
A
SP
TH
T8 T9 T10 T11
E5 E6
A2
TH12 TH13 TH14 TH15 TH16
75 80 85 90 95
Gly-Ile-Val-Ile-Gly-Lys-Lys-Gly-Glu-Asp-Val-Glu-Lys-Leu-Arg-Lys-Val-Val-Ala-Asp-Ile-Ala-Gly-Val-
T
ET
A
SP
TH
T12 T13 T14
A3 A4
SP4 SP5
TH17 TH18 TH19
100 105 110 115 120
Pro-Ala-Gln-Ile-Asn-Ile-Ala-Glu-Val-Arg-Lys-Pro-Glu-Leu-Asp-Ala-Lys-Leu-Val-Ala-Asp-Ser-Ile-Thr-
T
ET
A
SP
TH
T15 T16
E7
SP6 SP7
TH20 TH21 TH22 TH23 TH24 TH25

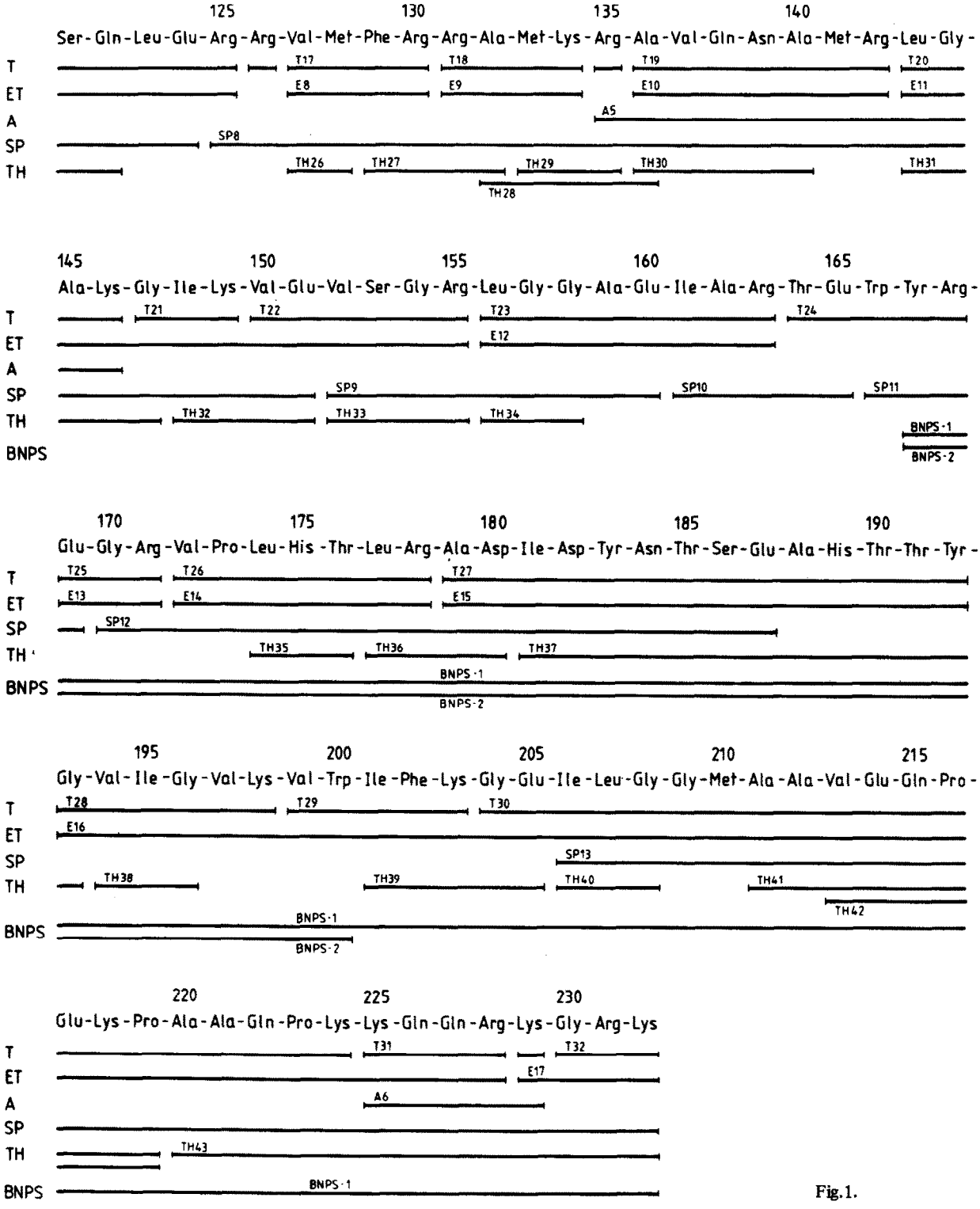

125 130 135 140
Ser-Gln-Leu-Glu-Arg-Arg-Val-Met-Phe-Arg-Arg-Ala-Met-Lys-Arg-Ala-Val-Gln-Asn-Ala-Met-Arg-Leu-Gly-
T
ET
A
SP
TH
T17 T18 T19 T20
E8 E9 E10 E11
A5
SP8
TH26 TH27 TH29 TH30 TH31
TH28
145 150 155 160 165
Ala-Lys-Gly-Ile-Lys-Val-Glu-Val-Ser-Gly-Arg-Leu-Gly-Gly-Ala-Glu-Ile-Ala-Arg-Thr-Glu-Trp-Tyr-Arg-
T
ET
A
SP
TH
BNPS
T21 T22 T23 T24
E12
SP9 SP10 SP11
TH32 TH33 TH34
BNPS-1
BNPS-2
170 175 180 185 190
Glu-Gly-Arg-Val-Pro-Leu-His-Thr-Leu-Arg-Ala-Asp-Ile-Asp-Tyr-Asn-Thr-Ser-Glu-Ala-His-Thr-Thr-Tyr-
T
ET
SP
TH
BNPS
T25 T26 T27
E13 E14 E15
SP12
TH35 TH36 TH37
BNPS-1
BNPS-2
195 200 205 210 215
Gly-Val-Ile-Gly-Val-Lys-Val-Trp-Ile-Phe-Lys-Gly-Glu-Ile-Leu-Gly-Gly-Met-Ala-Ala-Val-Glu-Gln-Pro-
T
ET
SP
TH
BNPS
T28 T29 T30
E16
SP13
TH38 TH39 TH40 TH41 TH42
BNPS-1
BNPS-2
220 225 230
Glu-Lys-Pro-Ala-Ala-Gln-Pro-Lys-Lys-Gln-Gln-Arg-Lys-Gly-Arg-Lys
T
ET
A
SP
TH
BNPS
T31 T32
E17
A6
TH43
BNPS-1

Fig.1.

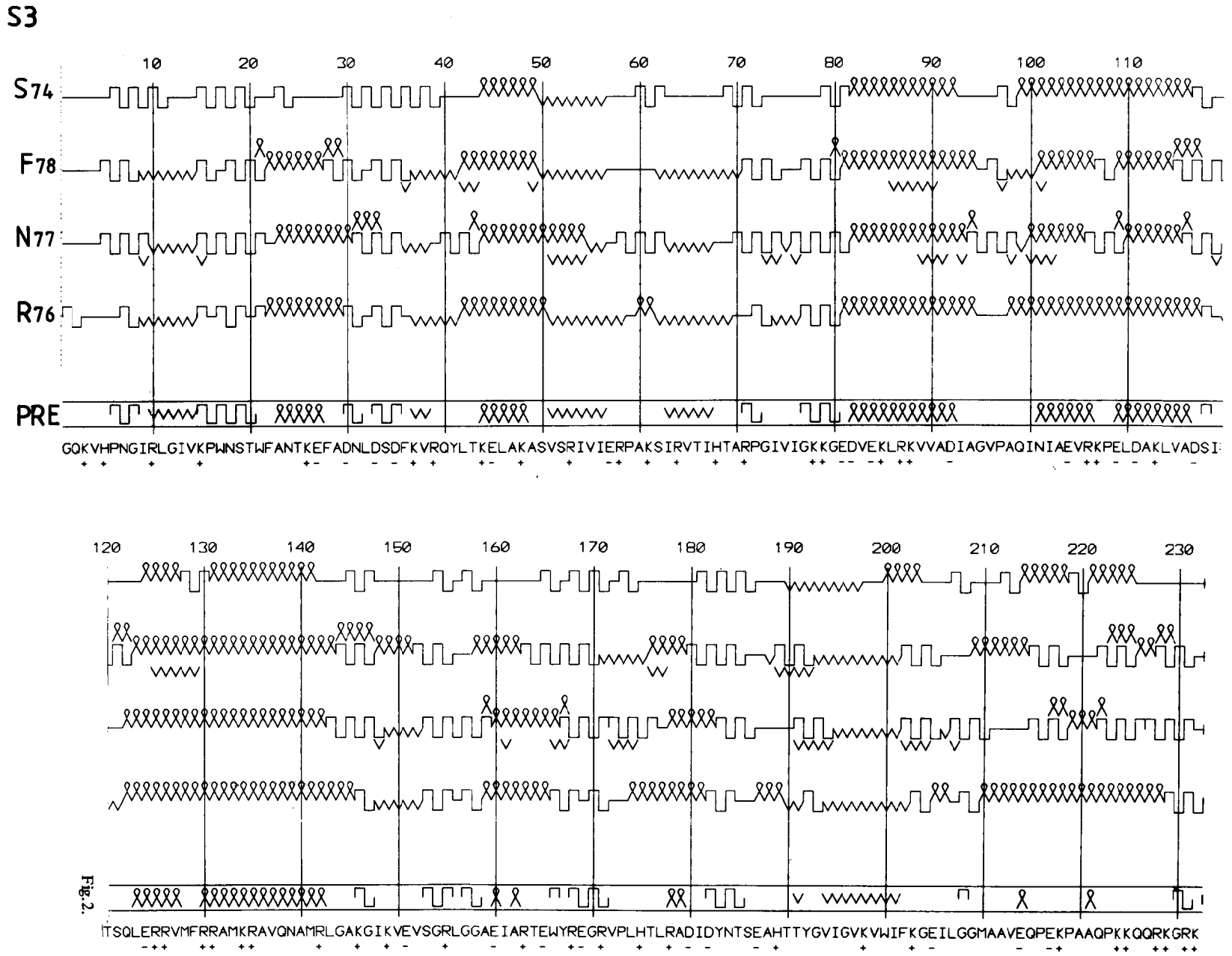

S3
10
20
30
40
50
60
70
80
90
100
110
S74
F78
N77
R76
PRE
GQKVHPNGIRLGIVKPWNSTWFANTKEFADNLDSDFKVRQYLTKELAKASVSRIVIERPAKSIRVTIHTARPGIVIGKKGEDVEKLRKVVADIAGVPAQINIAEVRKPELDAKLVADSI
120
130
140
150
160
170
180
190
200
210
220
230
TSQLERRVMFRRAMKRAVQNAMRLGAKGIKVEVSGRLGGAEIARTEWYREGRVPLHTLRADIDYNTSEAHTTYGVIGVKVWIFKGEILGGMAAVEQPEKPAAQPKKQQRKGRK

Fig.2.

Acknowledgements

We thank Dr H. G. Wittmann for support and discussion, and we are thankful to Irene Öchsner, Gisela Ehrke and Dr Robert Chen for supplying us with the primary structures of several tryptic peptides.

References

[1] Van Duin, J., Van Knippenberg, P. H., Dieben, M. and Kurland, C. G. (1972) Mol. Gen. Genet. 116, 181–191.
[2] Schreiner, G. and Nierhaus, K. H. (1973) J. Mol. Biol. 81, 71–82.
[3] Stöffler, G. and Wittmann, H. G. (1977) in: Molecular Mechanisms of Protein Biosynthesis (Weissbach, A. and Pestka, S. eds) pp. 117–202, Academic Press, London, New York.
[4] Brimacombe, R., Stöffler, G. and Wittmann, H. G. (1978) Ann. Rev. Biochem. 47, 217–249.
[5] Gulik, A., Freund, A. M. and Vachette, P. (1978) J. Mol. Biol. 119, 391–397.
[6] Rhode, M. F., O'Brien, S., Cooper, S. and Aune, K. C. (1975) Biochemistry 14, 1079–1087.
[7] Franz, A., Georgalis, Y. and Giri, L. (1979) Biochim. Biophys. Acta 578, 365–371.
[8] Möller, K., Rinke, J., Ross, A., Buddle, G. and Brimacombe, R. (1977) Eur. J. Biochem. 76, 175–187.
[9] Lutter, L. C., Ortandel, F. and Fasold, H. (1974) FEBS Lett. 48, 228–292.
[10] Sommer, A. and Traut, R. R. (1975) J. Mol. Biol. 97, 471–481.
[11] Hindenach, I., Stöffler, G. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 7–11.
[12] Riley, M. and Perham, R. N. (1970) Biochem. J. 118, 733–739.
[13] Toi, K., Bynum, E., Norris, E. and Itano, H. (1967) J. Biol. Chem. 242, 1036–1043.
[14] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.
[15] Fontana, A. (1972) Methods Enzymol. 25, 419–423.
[16] Chen, R. and Wittmann-Liebold, B. (1975) FEBS Lett. 52, 139–140.
[17] Chang, J. Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[18] Needleman, S. B. (1975) in: Protein Sequence Determination; Molecular Biology, Biochemistry and Biophysics, vol. 8, p. 201, Springer Verlag, Berlin, Heidelberg, New York.
[19] Edelhoch, A. (1967) Biochemistry 6, 1948–1954.
[20] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[21] Wittmann-Liebold, B., Graffunder, H. and Kohls, H. (1976) Analyt. Biochem. 75, 621–633.
[22] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[23] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[24] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.

Fig.1. The primary structure of protein S3 from *Escherichia coli. Abbreviations:* SQ, sequence determination with a modified Beckman sequenator (+ means unambiguously identified); T, trypsin digestion; ET, trypsin digestion after blocking the amino groups with ETPA; A, trypsin digestion after blocking the arginines by 1,2-cyclohexandione; SP, digestion with *Staphylococcus aureus* protease; TH, digestion with thermolysin; BNPS, chemical cleavage with BNPS–skatole.

Fig.2. Secondary structure prediction of protein S3. The symbols represent residues in helical ( ), β-turn or loops ( ), extended () and coil (——) conformation states, respectively. In the line 'PRE' the secondary structure is summarized for three out of four predictions in agreement.