PRIMARY STRUCTURE OF PROTEIN S2 FROM THE *ESCHERICHIA COLI* RIBOSOME

Brigitte WITTMANN-LIEBOLD and Arnim BOSSERHOFF
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, D-1000 Berlin 33 (Dahlem), Germany

Received 13 April 1981

1. Introduction

Protein S2 is a slightly elongated protein [1] which is much more sensitive to proteolytic attack than other ribosomal proteins under the same conditions [2]. It can bind to 16 S RNA [3] and can be crosslinked to proteins S1, S3, S5 and S8 (review [4]). Furthermore, crosslinking between the initiation factor IF-2 and a group of proteins including S2 has been reported [5]. Protein S2 is located near the cleft between the head and body of the 30 S particle, as revealed by immuno-electron microscopy [6,7].

Treatment of 30 S subunits with LiCl yields core particles which are inactive in the ribosome–EF-G and EF-T GTPase reaction. Addition of proteins S2, S5 and S9 can restore the enzymatic activities to the original level [8,9].

Several mutants with alterations in protein S2 have been isolated [10–12], and the gene for this protein has been mapped at 4 min on the *E. coli* chromosome map [10].

Here, we report the complete amino acid sequence of protein S2 and compare its primary structure with that of other ribosomal proteins. Furthermore, the secondary structure of this protein is deduced using 4 different predictive programmes.

2. Materials and methods

Protein S2 was isolated from 30 S subunits of *E. coli*, strain K12, in the presence of urea as in [13], and was kindly provided by Dr H. G. Wittmann. The purity of the protein was checked by two-dimensional gel electrophoresis [14].

Peptides were isolated as follows:

(i) Tryptic peptides (made in 0.1 M *N*-methylmorpholine acetate buffer, pH 8.1, at 37°C for 4–6 h with TPCK-trypsin from Worthington, at 1:50) were isolated by one of the following methods: fingerprinting on cellulose thin-layer sheets; Dowex 50 (2 × 90 mm) column chromatography; Sephadex G-50 (in 10% acetic acid) or G-100 (in 0.07% ammonia) gel filtration as in [15].

(ii) Peptides obtained from digestion with thermolysin (made in 0.1 M *N*-methyl morpholine acetate buffer, pH 8.1, at 50°C for 4 h at 1:100) were isolated by thin-layer fingerprinting and by column chromatography on a Dowex 50 microcolumn.

(iii) *Staphylococcus aureus* protease peptides (SP-peptides) from digestions at pH 8.1 (see above) and at pH 4.4 (in 0.1 M ammonium acetate buffer) at 37°C for 24 h at 1:30 enzyme:substrate were separated by gel filtration on Sephadex G-50 in 50% formic acid. The sediment from the digest at pH 4.4 was treated further with trypsin (5 h at 37°C) and chromatographed on Dowex 50. The amounts of protein used for the different digestions were 1 mg for the separation on 10 fingerprints, 2 mg for each gel filtration and 3–4 mg for Dowex 50 column chromatography.

(iv) For specific cleavage at arginine residues, the lysines were blocked with ETPA (exo-*cis*-3,6-endoxo-Δ^4-tetrahydrophthalic acid) (as detailed in [15]) with 12 mg reagent/2 mg protein in 0.3 M borate buffer at pH 8 for 3 h at 4°C. After dialysis in dilute ammonia (pH 8) the tryptic digestion was performed, followed by deblocking of the lysines in 10% acetic acid overnight at room temperature. The resulting peptides were isolated on Sephadex G-50 in 10% acetic acid.

(v) Cyanogen bromide fragments were made from 2.5 mg protein in 300 µl 70% formic acid (with the addition of 0.2% mercaptoethanol) with

2 mg reagent for 24 h at room temperature (in the dark), and were separated on Sephadex G-50 (in 10% acetic acid). Larger peptides were cleaved further by *Staphylococcus aureus* protease, pepsin or thermolysin and then separated by fingerprinting.

(vi) Partial cleavage in dilute formic acid (pH 2.0) was made at 110°C for 16 h, and the peptides were separated on Sephadex G-50 and G-75 in 10% acetic acid.

The cysteine-containing tryptic peptide T15 was isolated on thin-layer fingerprints after oxidation of the protein, and the tryptophan-containing peptides were isolated by reverse-phase high-pressure liquid chromatography with a gradient of dilute ammonium formate buffer (pH 4.3) and methanol at 40°C on RP-18, 5 μm support (from Merck, Germany). Peptides were detected by monitoring at 294 nm [16].

The Edman degradations were performed manually by the combined dansyl-Edman technique, or by the DABITC/PITC (4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate) double coupling method as reviewed in [15]. The intact protein was degraded in an improved Beckman sequencer as detailed in [17,18].

Amino acid analyses were performed on a Durrum D-500 analyser, using 1–2 nmol peptides for the hydrolysis (in 50 μl double-distilled 5.7 N HCl, for 20 h at 110°C; sensitivity scale of 0.5 A). Protein S2 was hydrolysed in 100 μg amounts (in 50 μl HCl at 110°C for 20, 48 and 72 h), using aliquots of 20% of the sample at 2.0 A sensitivity for each analysis.

3. Results and discussion

3.1. *Sequence determination*

The sequence of protein S2 was determined by combining results obtained from peptides which were derived from enzymatic digestions with trypsin, thermolysin and *Staphylococcus aureus* protease as well as from cyanogen bromide treatment. If necessary, these peptides were further digested as in fig.1. All peptides shown in fig.1 were analysed and sequenced manually by the combined dansyl-Edman procedure, or more recently by the DABITC/PITC double coupling method. Details of the methods employed for separating and sequencing the various peptides are reviewed in [15,17].

The alignment of the peptides in the N-terminal region was achieved by automated liquid-phase degradation of the protein [19] in an improved Beckman sequencer [17,18] as shown in fig.1. Beyond position 33, the interpretation of the results from the sequencer became ambiguous due to partial cleavages of labile peptide bonds in this protein, causing the formation of new end groups during the degradation. Labile peptide bonds are present at the tryptophan residues (pos. 22, 95 and 103) and at Asx–Pro bonds (e.g., in pos. 23–24 and 191–192). The presence of an Asp–Pro bond also caused difficulties in the sequence determination of the ribosomal protein S7 [20].

Isolation and sequencing of the Asx–Pro containing peptides T4 and T5 (pos. 21–34) as well as of the tryptophan-containing peptide (pos. 95–104) were rather difficult. The latter peptides were isolated by HPLC-separation on a reverse-phase support and were monitored at 294 nm ([16]; submitted). Another problem was the isolation of hydrophobic or rather insoluble peptides, e.g., T15 (pos. 81–94); T20 (pos. 115–127); T26 (pos. 152–173); T27 (pos. 174–207) and T30 (pos. 225–240). They could be isolated by combination of the methods in section 2.

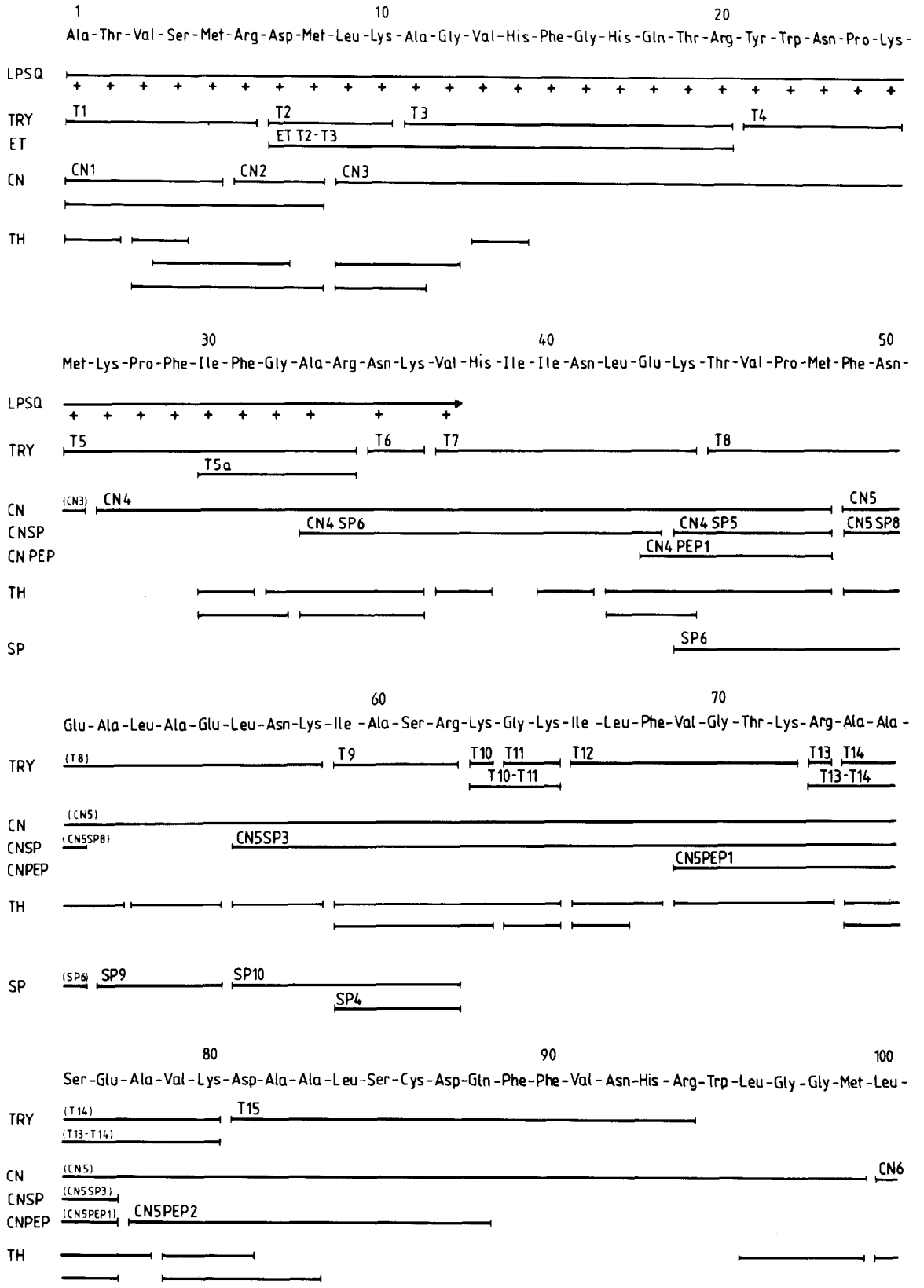
The tryptic peptides were aligned within the entire sequence by the isolation and analysis of the cyanogen bromide peptides CN4 (pos. 27–48), CN5 (pos. 49–99), CN6 (pos. 100–135), CN7 (pos. 136–153), CN8 (pos. 154–240), and of the thermolysin and SP-peptides which are shown in fig.1.

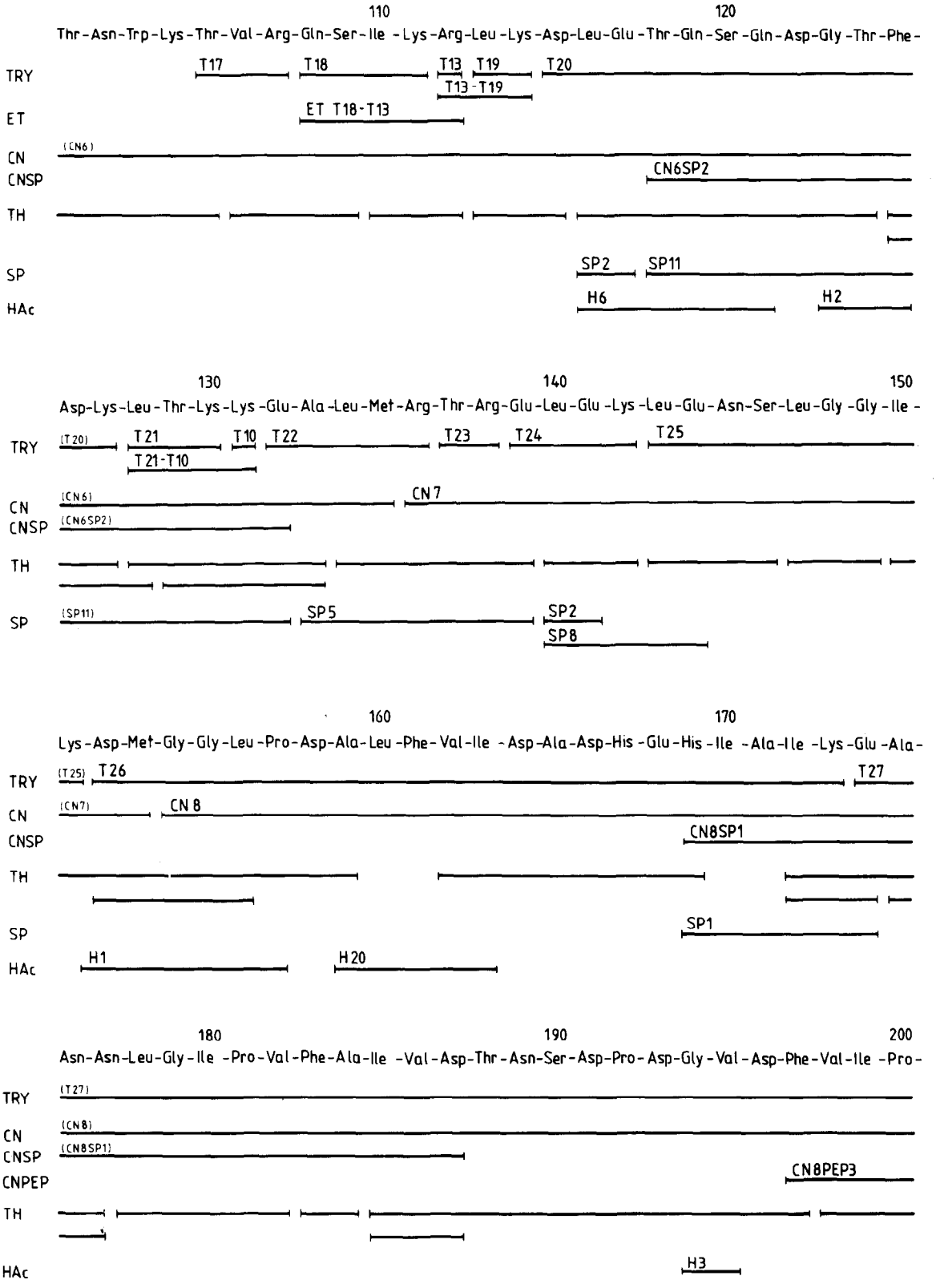
The primary structure of protein S2 (fig.1) is in full agreement with the amino acid sequence deduced from the nucleotide sequence of the gene coding for this protein (An, Mamelak, Bendiak and Friesen, personal communication).

3.2. *Characterization of the sequence*

Protein S2 has a molecular mass of 26 613 and its amino acid composition is: Asp_{17}, Asn_{12}, Thr_{13}, Ser_{11}, Glu_{16}, Gln_{7}, Pro_{7}, Gly_{17}, Ala_{25}, Val_{17}, Met_{7}, Ile_{14}, Leu_{21}, Tyr_{2}, Phe_{12}, His_{6}, Lys_{19}, Arg_{13}, Cys_{1} and Trp_{3}. This is in excellent agreement with the results from the amino acid analysis of this protein.

About 2/3rds of the 33 acidic residues of the protein are located in the C-terminal half of the protein chain, and many of these are clustered, e.g., in peptide T27 (pos. 174–207) which contains 10 Asx residues, and in peptide T30 (pos. 225–240) with 6 Glx residues. The basic amino acids lysine and arginine are more evenly distributed along the protein chain than the acidic residues, although one long


1
10
20
Ala-Thr-Val-Ser-Met-Arg-Asp-Met-Leu-Lys-Ala-Gly-Val-His-Phe-Gly-His-Gln-Thr-Arg-Tyr-Trp-Asn-Pro-Lys-
LPSQ
TRY
T1
T2
T3
T4
ET
ET T2-T3
CN
CN1
CN2
CN3
TH
30
40
50
Met-Lys-Pro-Phe-Ile-Phe-Gly-Ala-Arg-Asn-Lys-Val-His-Ile-Ile-Asn-Leu-Glu-Lys-Thr-Val-Pro-Met-Phe-Asn-
LPSQ
TRY
T5
T6
T7
T8
T5a
CN
(CN3)
CN4
CN5
CNSP
CN4 SP6
CN4 SP5
CN5 SP8
CN PEP
CN4 PEP1
TH
SP
SP6
60
70
Glu-Ala-Leu-Ala-Glu-Leu-Asn-Lys-Ile-Ala-Ser-Arg-Lys-Gly-Lys-Ile-Leu-Phe-Val-Gly-Thr-Lys-Arg-Ala-Ala-
TRY
(T8)
T9
T10
T11
T12
T13
T14
T10-T11
T13-T14
CN
(CN5)
CNSP
(CN5SP8)
CN5SP3
CNPEP
CN5PEP1
TH
SP
(SP6)
SP9
SP10
SP4
80
90
100
Ser-Glu-Ala-Val-Lys-Asp-Ala-Ala-Leu-Ser-Cys-Asp-Gln-Phe-Phe-Val-Asn-His-Arg-Trp-Leu-Gly-Gly-Met-Leu-
TRY
(T14)
T15
(T13-T14)
CN
(CN5)
CN6
CNSP
(CN5SP3)
CNPEP
(CN5PEP1)
CN5PEP2
TH


110
120
Thr-Asn-Trp-Lys-Thr-Val-Arg-Gln-Ser-Ile-Lys-Arg-Leu-Lys-Asp-Leu-Glu-Thr-Gln-Ser-Gln-Asp-Gly-Thr-Phe-
TRY
T17
T18
T13
T19
T20
T13-T19
ET
ET T18-T13
CN
(CN6)
CNSP
CN6SP2
TH
SP
SP2
SP11
HAc
H6
H2
130
140
150
Asp-Lys-Leu-Thr-Lys-Lys-Glu-Ala-Leu-Met-Arg-Thr-Arg-Glu-Leu-Glu-Lys-Leu-Glu-Asn-Ser-Leu-Gly-Gly-Ile-
TRY
(T20)
T21
T10
T22
T23
T24
T25
T21-T10
CN
(CN6)
CN7
CNSP
(CN6SP2)
TH
SP
(SP11)
SP5
SP2
SP8
160
170
Lys-Asp-Met-Gly-Gly-Leu-Pro-Asp-Ala-Leu-Phe-Val-Ile-Asp-Ala-Asp-His-Glu-His-Ile-Ala-Ile-Lys-Glu-Ala-
TRY
(T25)
T26
T27
CN
(CN7)
CN8
CNSP
CN8SP1
TH
SP
SP1
HAc
H1
H20
180
190
200
Asn-Asn-Leu-Gly-Ile-Pro-Val-Phe-Ala-Ile-Val-Asp-Thr-Asn-Ser-Asp-Pro-Asp-Gly-Val-Asp-Phe-Val-Ile-Pro-
TRY
(T27)
CN
(CN8)
CNSP
(CN8SP1)
CNPEP
CN8PEP3
TH
HAc
H3

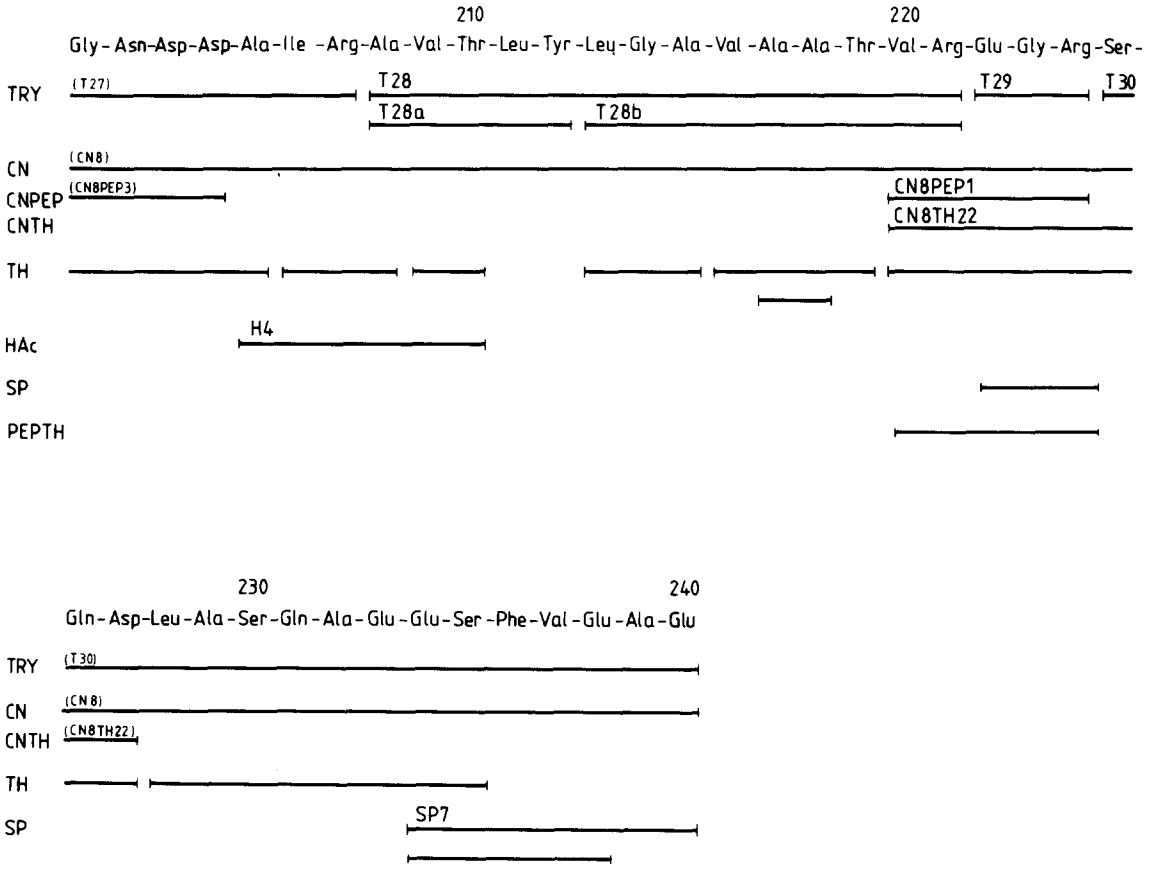


Fig.1. Primary structure of protein S2: LPSQ, Edman degradation of intact protein S2 by means of an improved Beckman sequencer; +, means identification by thin-layer chromatography; TRY, digestion with trypsin; ET, with trypsin after reaction with ETPA; SP, with *Staphylococcus aureus* protease; PEP, with pepsin; HAc, cleavage with dilute acid; CN, cleavage with cyanogen bromide. Several of the cyanogen bromide peptides were further cleaved with *Staphylococcus aureus* protease, pepsin or thermolysin as indicated.

region (pos. 174–206) is completely free of basic amino acids. Four out of the 6 histidines are present in 2 regions (at pos. 14–17 and 167–169). The only cysteine is located at pos. 86, and the majority of the methionines and of the aromatic amino acids are clustered in the N-terminal half of the protein chain.

3.3. *Secondary structure predictions*

Four different prediction methods were employed for calculating the secondary structure of protein S2 as presented in fig.2 and in [21]. In the case of the predictions according to Chou and Fasman, the most recent parameters [22,23] were used. According to these predictions, protein S2 contains 31% helix, 21% turns or loops and only a small amount of β-sheet structure. Pronounced helix areas are predicted for pos. 48–60, 73–84 and 129–144.

3.4. *Comparison with sequences of other ribosomal proteins*

The sequence of protein S2 was compared with the 51 *E. coli* ribosomal proteins whose sequences have been determined (review [2]), and with the primary structures of 7 ribosomal proteins from other organisms (yeast, *Bacillus subtilis, Bacillus stearothermophilus, Artemia salina* and rat liver; review [24]). No long stretches of identical or strongly homologous sequence regions were found. Details will be presented elsewhere in the near future when the primary structures of all *E. coli* ribosomal proteins have been determined.

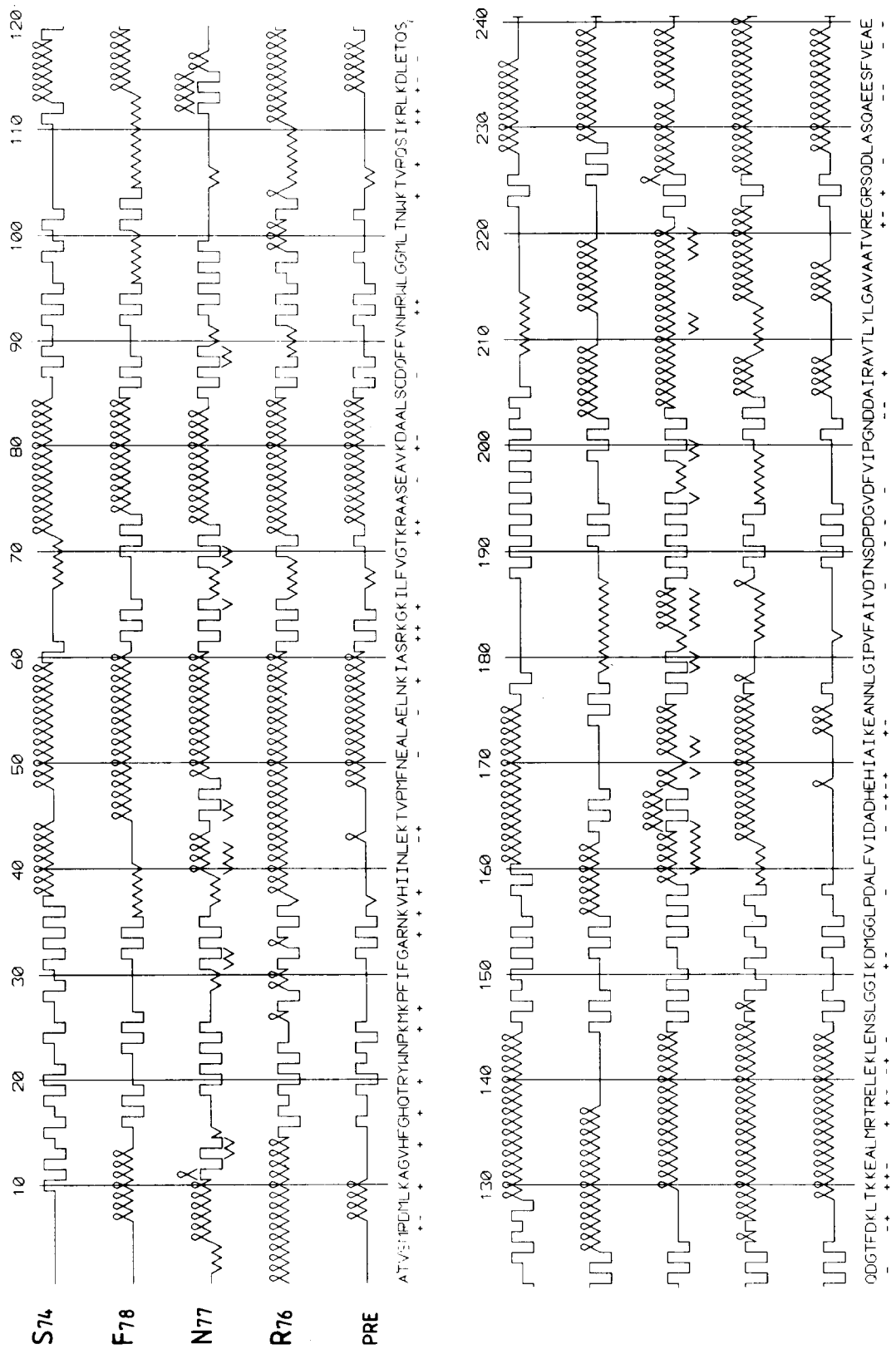


Fig.2. Prediction of secondary structure of protein S2 according to 4 different methods (detailed in text and [21]).

Acknowledgement

We are pleased to thank Dr J. D. Friesen for a very useful exchange of information.

References

[1] Georgalis, Y., Littlechild, J. A. and Giri, L. (1981) Biochemistry 20, 1061–1064.
[2] Wittmann, H. G., Littlechild, J. A. and Wittmann-Liebold, B. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 51–88, University Park Press, Baltimore MD.
[3] Littlechild, J. A., Dijk, J. and Garrett, R. A. (1977) FEBS Lett. 74, 292–300.
[4] Traut, R. R., Lambert, J. M., Boileau, G. and Kenny, J. W. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 89–110, University Park Press, Baltimore MD.
[5] Bollen, A., Heimark, R. L., Cozzone, A., Traut, R. R., Hershey, J. W. and Kahan, L. (1975) J. Biol. Chem. 250, 4310–4314.
[6] Tischendorf, G. W., Zeichhardt, H. and Stöffler, G. (1975) Proc. Natl. Acad. Sci. USA 72, 4820–4824.
[7] Lake, J. A. (1978) in: Advanced Techniques in Biological Electron Microscopy II (Koehler, J. ed) pp. 173–211, Springer-Verlag, Berlin, Heidelberg, New York.
[8] Marsh, R. C. and Parmeggiani, A. (1973) Proc. Natl. Acad. Sci. USA 70, 151–155.
[9] Sander, G., Marsh, R. C. and Parmeggiani, A. (1973) FEBS Lett. 33, 132–134.
[10] Nashimoto, H. and Uchida, H. (1975) J. Mol. Biol. 96, 443–453.
[11] Isono, S., Isono, K. and Hirota, Y. (1978) Mol. Gen. Genet. 165, 15–20.
[12] Dabbs, E. R. (1978) Mol. Gen. Genet. 165, 73–78.
[13] Hindennach, I., Stöffler, G. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 7–11.
[14] Kaltschmidt, E. and Wittmann, H. G. (1970) Anal. Biochem. 36, 401–412.
[15] Wittmann-Liebold, B., Brauer, D., Dognin, J. (1977) in: Solid Phase Methods in Protein Sequence Analysis (Previero, A. and Coletti-Previero, M. A. eds) pp. 219–232, Elsevier/North-Holland, Amsterdam, New York.
[16] Wittmann-Liebold, B. (1981) in: High Pressure Liquid Chromatography in Protein and Peptide Chemistry (Lottspeich, F. et al. eds) Walter de Gruyter, Berlin, in press.
[17] Wittmann-Liebold, B. (1980) in: Polypeptide Hormones (Beers, R. F. jr and Bassett, E. G. eds) pp. 87–120, Raven, New York.
[18] Wittmann-Liebold, B. (1981) in: Proceedings of the International Conference on the Chemical Synthesis and Sequencing of Peptides and Proteins (Schechter, A. N. and Liu, T. Y. eds) in press.
[19] Wittmann, H. G. and Wittmann-Liebold, B. (1974) in: Ribosomes (Nomura, N. et al. eds) pp. 115–140, Cold Spring Harbor Laboratory, NY.
[20] Reinbolt, J., Tritsch, D. and Wittmann-Liebold, B. (1979) Biochimie 61, 501–522.
[21] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.
[22] Chou, P. Y. and Fasman, G. D. (1978) Annu. Rev. Biochem. 47, 251–276.
[23] Chou, P. Y. and Fasman, G. D. (1977) J. Mol. Biol. 115, 135–175.
[24] Wittmann-Liebold, B. (1980) in: RNA Polymerase, tRNA and Ribosomes (Osawa, S. et al. eds) pp. 639–654, University of Tokyo Press, Tokyo.