PRIMARY STRUCTURE OF PROTEIN S11 FROM *ESCHERICHIA COLI* RIBOSOMES

R. KAMP and B. WITTMANN-LIEBOLD
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, D-1000 Berlin 33 (Dahlem), Germany

Received 15 September 1980

1. Introduction

Protein S11 has by immune electron microscopy been localized on the head and the large lobe [1] or on the platform [2] of the 30 S subunit. It has been crosslinked to proteins S8, S13 and S21 by bifunctional reagents [3]. The distance between the gravity centers of mass of proteins S11 and S12 has been determined to be 114 Å and that between S11 and S15 to be 95 Å [4]. However, the location of protein S11 within the 30 S subunit as deduced from neutron scattering studies is at present difficult to reconcile with the immune electron microscopical results.

Reconstruction tests have identified protein S11 as a fidelity protein since its omission leads to a decreased fidelity of translation [5]. Interestingly, protein S11 was one of the proteins found to be affinity-labeled using a mRNA derivative [6].

In this paper the complete primary structure of protein S11 which consists of 128 amino acids residues and has M_r 13 728 is described. We also show the secondary structure of protein S11 using 4 different prediction programmes. Furthermore, we present data obtained from computer searching for homologous structures between protein S11 on one hand and ribosomal proteins from *Escherichia coli* and other organisms on the other hand.

2. Materials and methods

Protein S11 has been isolated from 30 S subunits of *E. coli* K as in [7] and was provided in salt-free lyophilized form by Dr H. G. Wittmann. For sequencing studies, fragments of protein S11 were isolated applying the following cleavage methods:

(i) Trypsin digestion at pH 8.1 (in 0.1 M *N*-methylmorpholine acetate buffer, at 37°C for 4–6 h; enzyme/substrate ratio of 1:50 with TPCK-trypsin from Worthington).

(ii) Chymotrypsin digestion at pH 8.1 (same buffer as above, at 37°C for 1 h; enzyme/substrate ratio of 1:100 with TLCK-α-chymotrypsin from Merck).

(iii) Thermolysin digestion at pH 8.1 of the protein and some tryptic peptides (same buffer, at 37°C and 52°C for 2 h; enzyme/substrate ratio of 1:100 with thermolysin from Serva).

(iv) Digestion with *Staphylococcus aureus* protease at pH 4.0 (0.05 M ammonium acetate buffer, at 37°C for 16 h; enzyme/substrate ratio of 1:50 with protease V-8 from Miles).

(v) Partial tryptic digestion was obtained after the protein had been acylated at its lysine residues with citraconic anhydride at pH 7.5 [8]. The protein was then desalted on a Sephadex G-25 column in 0.2% NH_4HCO_3 before submitting it to tryptic digestion.

(vi) Cleavage with cyanogen bromide (CN) was performed in 70% formic acid for 24 h at room temperature (reagent/protein ratio 1:1).

(vii) Cleavage with 2-(2-nitrophenyl-sulphenyl)-3-methyl-3′-bromoindolenine skatole (BNPS) of the protein was performed in 50% acetic acid [9] in the presence of an excess of tyrosine.

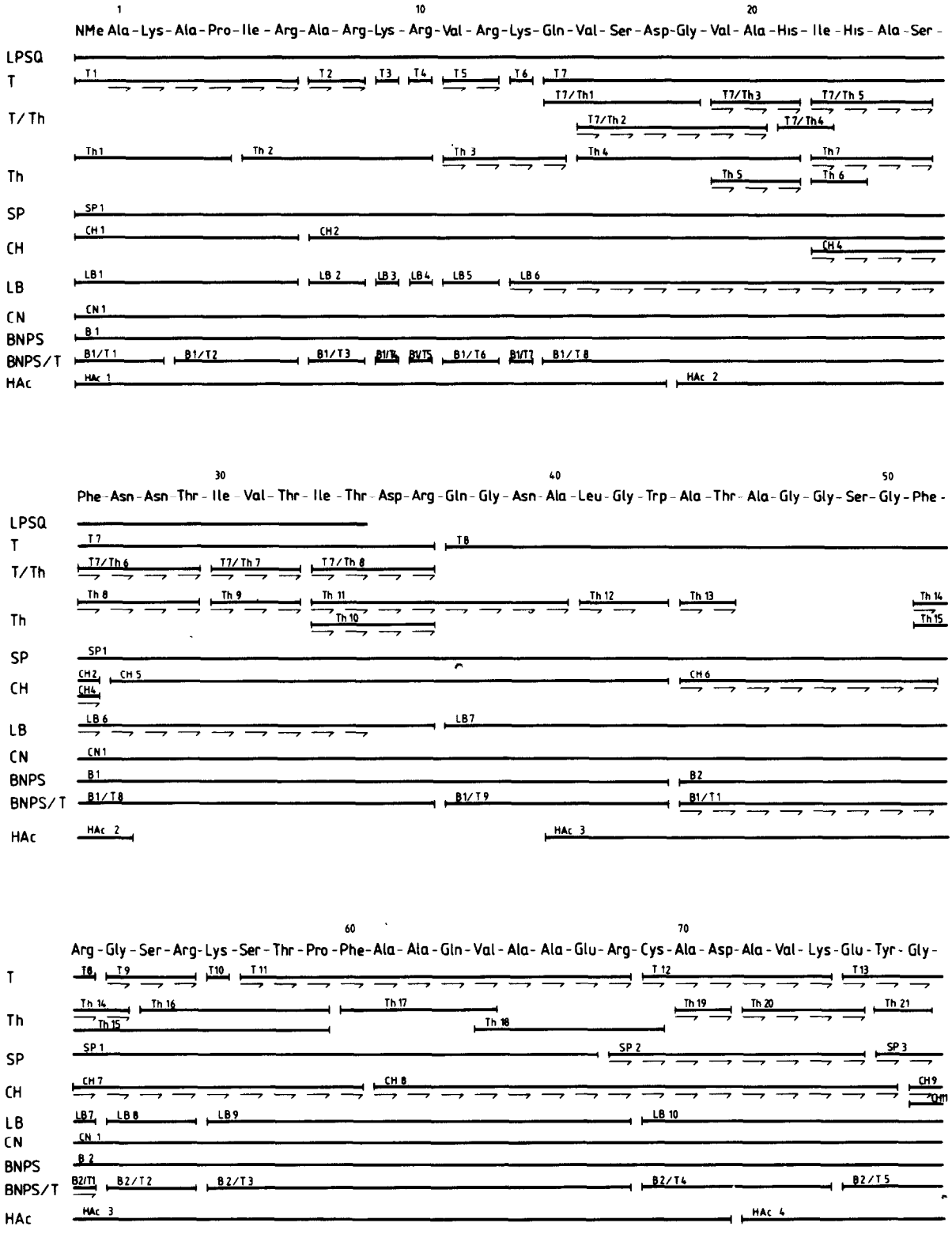
(viii) Partial acid digestion of the protein and of tryptic peptide T18 (after oxidation) was made at pH 2.8 (in 2% acetic acid for 20 h at 110°C).

The peptides were isolated employing mainly the following methods:

(a) Fingerprint technique on cellulose thin-layer sheets (reviewed in [10]);

(b) Column chromatography on Dowex 50 (microcolumns 0.2–0.3 × 9 cm); detailed in [11,12];

(c) Gel filtration on Sephadex G-50 superfine (columns 1 × 195 cm) in 10% acetic acid.


1
10
20
NMe Ala-Lys-Ala-Pro-Ile-Arg-Ala-Arg-Lys-Arg-Val-Arg-Lys-Gln-Val-Ser-Asp-Gly-Val-Ala-His-Ile-His-Ala-Ser-
LPSQ
T
T/Th
Th
SP
CH
LB
CN
BNPS
BNPS/T
HAc
30
40
50
Phe-Asn-Asn-Thr-Ile-Val-Thr-Ile-Thr-Asp-Arg-Gln-Gly-Asn-Ala-Leu-Gly-Trp-Ala-Thr-Ala-Gly-Gly-Ser-Gly-Phe-
60
70
Arg-Gly-Ser-Arg-Lys-Ser-Thr-Pro-Phe-Ala-Ala-Gln-Val-Ala-Ala-Glu-Arg-Cys-Ala-Asp-Ala-Val-Lys-Glu-Tyr-Gly-

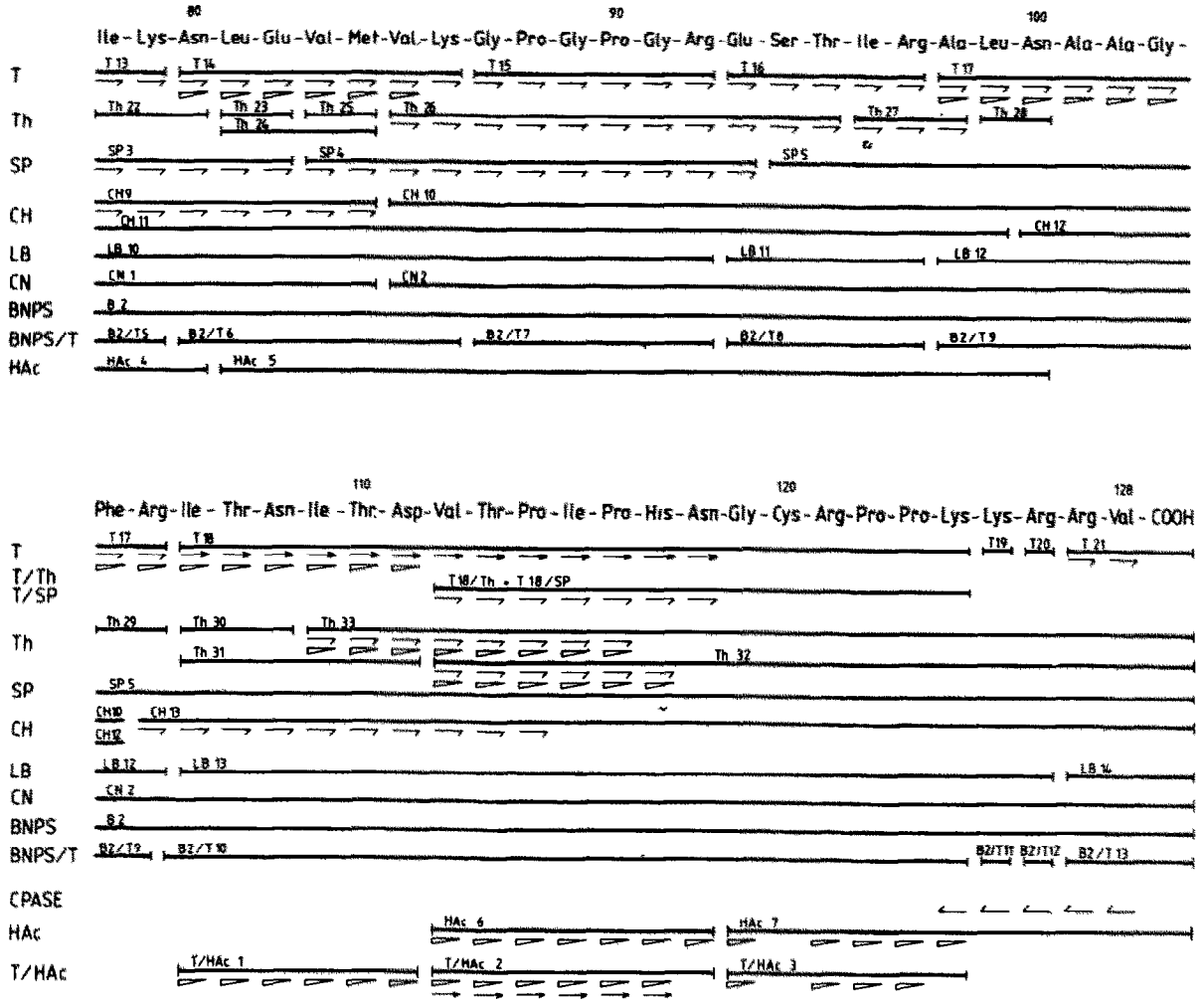


Fig.1. Primary structure of protein S11 from *Escherichia coli* ribosomes: LPSQ, liquid-phase Edman degradation in a modified Beckman sequencer with automatic conversion device; T, trypsin digestion; T/Th, thermolysine digestion of tryptic peptides; TH, thermolysine digestion; SP, digestion with *Staphylococcus aureus* protease; CH, chymotrypsin digestion; LB, tryptic digestion of citraconylated protein; CN, cleavage of the protein with cyanogen bromide; BNPS, cleavage with 2-(2-nitrophenyl-sulphenyl)-3-methyl-3'-bromoindolenine skatole; BNPS/T, tryptic digestion of BNPS-peptides; HAc partial acid cleavage; ⟶ combined micro dansyl-Edman method; ⟶● solid-phase sequencing by attachment of the C-terminal carboxyl group of the peptides to polystyrene, degradation performed with PITC, identification of PTH-amino acid derivatives by thin-layer technique; ⟹, DABITC/PITC double coupling method.

The peptides were sequenced by the combined micro dansyl-Edman technique [13] or by the DABITC/PITC (4-*N*,*N*'-dimethylaminoazobenzene-4'-isothiocyanate/phenylisothiocyanate) double coupling method [14].

Amino acid analysis were performed in a Durrum D-500 analyser. Details of the experimental procedures employed are reviewed in [10].

3. Results and discussion

3.1. *N-terminal region (pos. 1–32)*

The N-terminal sequence of protein S11 up to pos. 32 was determined [15] by liquid-phase Edman degradation performed in a modified Beckman sequencer as detailed in [16]. In the present studies on peptides of this region the previously published

sequence is confirmed. However, discrepancies derived from the identification of the released first amino acid which comigrated (in thin-layer chromatography and different solvent mixtures) with different standard PTH-amino acids, e.g., with either PTH-Phe or PTH-Pro. There was no detectable free amino acid in the analyser after acid hydrolysis of the PTH-derivative (M. Yaguchi, personal communication).

By means of mass spectroscopy and dansylation the unusual N-terminal amino acid was identified as *N*α-monomethylalanine in protein S11 [17]. Other problems derived from the fact that protein S11 gave a heterogenous N-terminal sequence in the Edman degradation (performed either automatically or manually). Part of the second residue (a lysine) occurred in the first degradation cycle, the third amino acid in the second, etc. This 'pre-lapping' could be caused by: either (a) a peptide bond between the methylated alanine and the *N*ε-amino group of the lysine (a branched peptide); or (b) the presence of two chains in protein S11 with one residue difference. The reaction of the PITC-analogue DABITC (4-*N*,*N*'-dimethylaminoazobenzene-isothiocyanate) with the protein S11 showed however, that an abnormal Edman degradation can take place with peptides having an N-terminal *N*α-monomethylated amino acid [18]: the cleavage occurs at alkaline conditions and leads to the release of the thiazolinone amino acid of the first residue at the coupling stage. Therefore, at the same time the following residue can partly be coupled with the reagent and is released simultanously [18].

3.2. *The middle region (pos. 33–105)*

The protein chain was cleaved by BNPS-skatole at the single tryptophan residue of protein S11, which is located in the middle area of the chain. This cleavage released the large fragment B1 (pos. 1–43). It was further digested by trypsin to reach smaller peptides, which could be purified easily. Trypsin, thermolysin and chymotrypsin released a number of smaller peptides which were sequenced by the dansyl-Edman technique (see fig.1). Tryptic peptide T8 (pos. 37–52) could not be obtained in sufficient amounts for direct sequence analysis. However, treatment of the protein with dilute acid resulted in a peptide, bridging the region around the tryptophan residue: HAc3 (pos. 40–71). Fragments obtained after digesting the protein with *S. aureus* protease were employed for sequencing beyond pos. 68–93. Glutamine and asparagine were identified by the manually performed double coupling method with the reagent DABITC [14] which results in red coloured DABTH-derivatives, an advantage for the direct identification of these residues. Cyanogen bromide cleaved the protein at its single methionine residue at pos. 84.

3.3. *The C-terminal region (pos. 106–128)*

Not only the N-terminal region but also the C-terminal part of protein S11 was difficult to sequence. The reasons were the accumulation of proline and basic residues in the C-terminal area. Thermolysin and chymotrypsin peptides served for confirming the N-terminal part of peptide T18 (pos. 106–124) (see fig.1). However, no enzymatic cleavages were obtained beyond pos. 111. This was caused by the region Pro–Ile–Pro (pos. 114–116). Edman degradation performed manually (dansyl-Edman or PITC/DABITC-technique) or automatically (pos. 106–118 by solid-phase technique) did not enable us to degrade peptide T18 beyond this sequence without severe overlapping. Thermolysin and staphylococcal peptides derived by further digestion of T18 were sequenced up to pos 118, employing 3–4 cleavages at the Pro–Ile–Pro sequence. The sequence could then be established by sequencing peptides obtained with dilute acid treatment of the protein. This resulted in peptides HAc6 (pos. 112–118) and HAc7 (pos. 119–128). The cleavage of the peptide chain was directly after the difficult sequence area, allowing a degradation without overlapping at pos. 119–123. Peptide LB13 (pos. 106–126) served as proof for the alignment of the basic residues at the C-terminus of protein S11. This result was confirmed by additional enzymatic digestions of the protein with carboxypeptidase A and B.

3.4. *Characterization of the sequence*

Several independent sets of peptides deriving from cleavage by trypsin, trypsin after citraconylation, chymotrypsin, cyanogen bromide and BNPS-skatole were isolated in order to elucidate the entire sequence of protein S11, as shown in fig.1. The amino acid composition is: Asp_4, Asn_7, Thr_9, Ser_6, Glu_4, Gln_3, Pro_8, Gly_{13}, Ala_{17}, Val_{10}, Met_1, Ile_9, Leu_3, Tyr_1, Phe_4, His_3, Lys_9, Arg_{14}, Trp_1 and Cys_2. This agrees very well with the results obtained from amino acid analysis of the entire protein. Based on the sequence the M_r is 13 728. The amino acid sequence of protein S11 (fig.1) is in complete agreement with the nucleotide

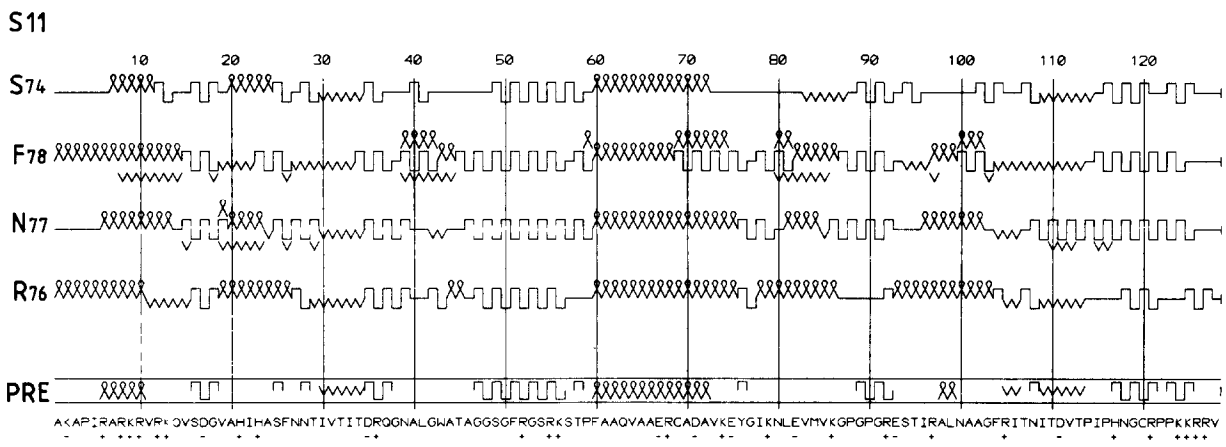


Fig.2. Secondary structure of protein S11 as predicted according to 4 different methods (S, F, N, R). The line 'PRE' summarizes the secondary structure obtained when at least 3 out of the 4 predictions are in agreement. The amino acid sequence of protein S11 is shown in the bottom line in the one-letter code (details in [19]): ⌇⌇⌇⌇⌇⌇⌇, helical; ∿∿∿∿∿, β-sheet; ⊓⊓⊓⊓, turn or loop; ——, random coil conformational stages.

sequence within the gene for this protein (L. Post and M. Nomura, personal communication).

The following characteristic structural features of ribosomal protein S11 can be deduced from this primary sequence:

(i) Two very basic areas occur at the N- and C-terminus of the chain. This was also observed for several other *E. coli* ribosomal proteins.

(ii) The C-terminal basic region has clusters of prolines and contains one histidine and one cysteine residue.

(iii) The two other histidines are located near each

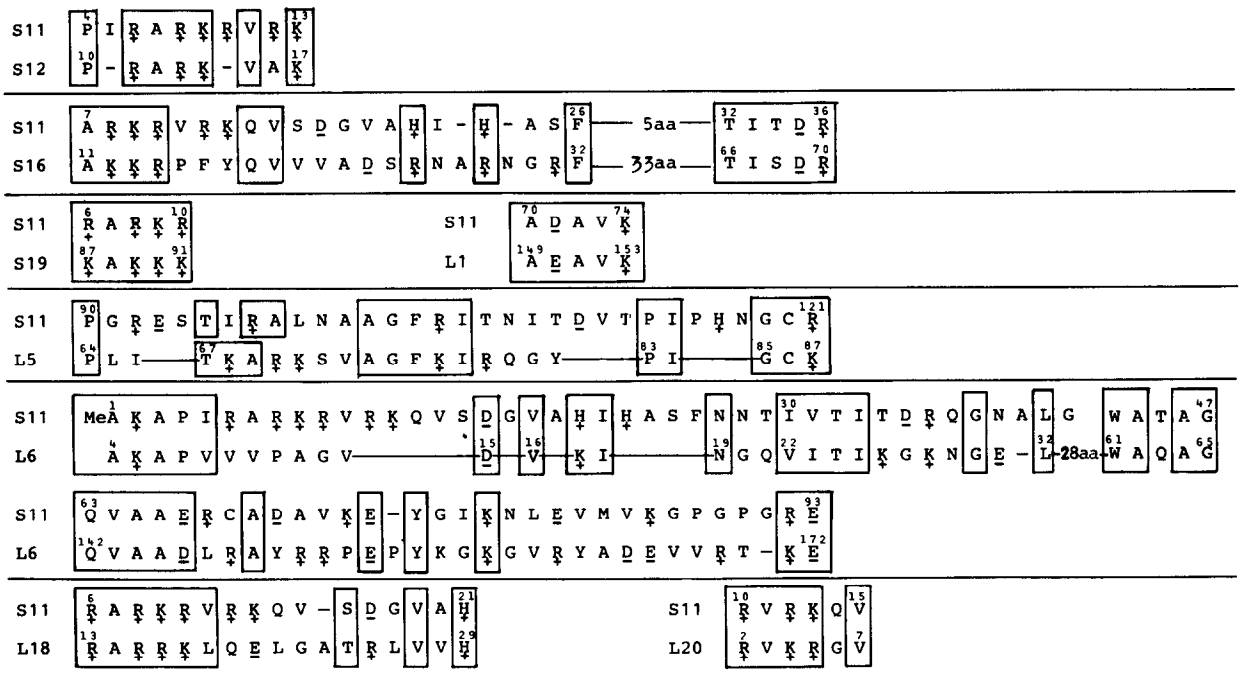


Fig.3. Identical or homologous regions in protein S11 and other *Escherichia coli* ribosomal proteins.

Table 1
Secondary structure prediction of protein S11

	Helix (%)	β-Sheet (%)	Turn (%)
Scheraga 74	18	13	33
Fasman 78	15–38	19–37	34–48
Nagano 77	27–31	8–11	38–45
Robson 76	44	14	27

other at the beginning of the hydrophobic middle section of the chain (pos. 19–33).

(iv) All aromatic residues occur in the inner part of the molecule (at pos. 26, 43, 51, 60, 76, 104) which also contains the single methionine residue and the second cysteine.

3.5. *Secondary structure predictions of protein S11*

Based on the primary structure, predictions of the secondary structure were made employing 4 different algorithms, as in [19] with the difference that more recent parameters were used in the calculations according to Chou and Fasman. The prediction results for protein S11 are presented in fig.2. Notable are the two pronounced turn areas (pos. 46–58 and 116–125), the two areas of β-sheet structure (pos. 29–34 and 105–114) and one dominating helix area (pos. 60–74) predicted by all methods. Table 1 presents the calculated percentage of helical, extended and turn structure for this protein.

3.6. *Comparison with known ribosomal sequences*

The sequence of protein S11 was compared by means of computer programmes with 51 known *E. coli* ribosomal proteins searching for identical and homologous sequence areas. In this search serine and threonine, valine and isoleucine, aspartic acid and glutamic acid and lysine and arginine, respectively, were considered as conservative replacements. The obtained similarities are as given in fig.3.

Acknowledgements

The work reported in this paper was supported by a grant from the Deutsche Forschungsgemeinschaft. Part of the sequence has been established by Dr Robert Chen.

References

[1] Stöffler, G., Bald, R., Kastner, B., Lührmann, R., Stöffler-Meilicke, M. and Tischendorf, G. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 171–205, University Park Press, Baltimore, MD.
[2] Lake, J. A. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 207–236, University Park Press, Baltimore, MD.
[3] 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.
[4] Moore, P. B. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 111–133, University Park Press, Baltimore, MD.
[5] Nomura, M., Mizushima, S., Ozaki, M., Traub, P. and Lowry, C. W. (1969) Cold Spring Harbor Symp. Quant. Biol. 34, 49–61.
[6] Pongs, O., Stöffler, G. and Lanka, E. (1975) J. Mol. Biol. 72, 1505–1509.
[7] Hindennach, I., Stöffler, G. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 7–11.
[8] Dixon, H. B. F. and Perham, R. N. (1968) Biochem. J. 109, 312–314.
[9] Fontana, A. (1972) Methods Enzymol. 25, 419–423.
[10] Wittmann-Liebold, B. and Lehmann, A. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, C. ed.) pp. 49–72, Elsevier/North-Holland, Amsterdam, New York.
[11] Chen, R. (1977) Hoppe Seyler's Z. Physiol. Chem. 358, 1415–1430.
[12] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe Seyler's Z. Physiol. Chem. 357, 1751–1770.
[13] Chen, R. (1976) Hoppe Seyler's Z. Physiol. Chem. 357, 873–886.
[14] Chang, J. Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[15] Wittmann-Liebold, B. (1973) FEBS Lett. 36, 247–249.
[16] Wittmann-Liebold, B. (1973) Hoppe Seyler's Z. Physiol. Chem. 354, 1415–1431.
[17] Chen, R. Brosius, J., Wittmann-Liebold, B. and Schäfer, W. (1977) J. Mol. Biol. 111, 173–181.
[18] Chang, J. Y. (1978) FEBS Lett. 91, 63–68.
[19] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe Seyler's Z. Physiol. Chem. 358, 1003–1019.