THE PRIMARY STRUCTURE OF PROTEIN L4 FROM THE LARGE SUBUNIT OF THE *ESCHERICHIA COLI* RIBOSOME

Makoto KIMURA and Brigitte WITTMANN-LIEBOLD
Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33 (Dahlem), Germany

Received 6 October 1980

1. Introduction

Protein L4 binds directly and specifically to 23 S RNA [1]. Its binding site is located within the 13 S RNA which is a fragment at the 5'-end of the 23 S RNA strand [2]. Recently, it became possible to localize a contact point between the 23 S RNA and protein L4: nucleotide U-615 has been crosslinked by UV-irradiation of 50 S subunits to Tyr-35 of L4 [3]. Studies on protein–protein crosslinks by bifunctional reagents revealed that proteins L11 and L14 are neighbors of L4 [4], and isolation of small 50 S fragments showed that proteins L3, L4 and L24 are present within a small fragment [5]. This finding is in good agreement with the results from immune electron microscopy according to which protein L4 is located at the right-hand side of the 50 S 'armchair' in direct neighborhood to L24 and relatively close to L3 [6].

Reconstitution experiments demonstrated protein L4 to be an important member of the 'early assembly group' of proteins [7] and to be a ribosomal component which highly stimulates the peptidyltransferase activity as revealed by single protein omission tests [8]. Among mutants with an altered L4 protein [9–11] some have a high level of resistance to the antibiotic erythromycin [9,10]. The amino acid exchanges in protein L4 of several erythromycin-resistant mutants probably occur always within the same peptide [10,12].

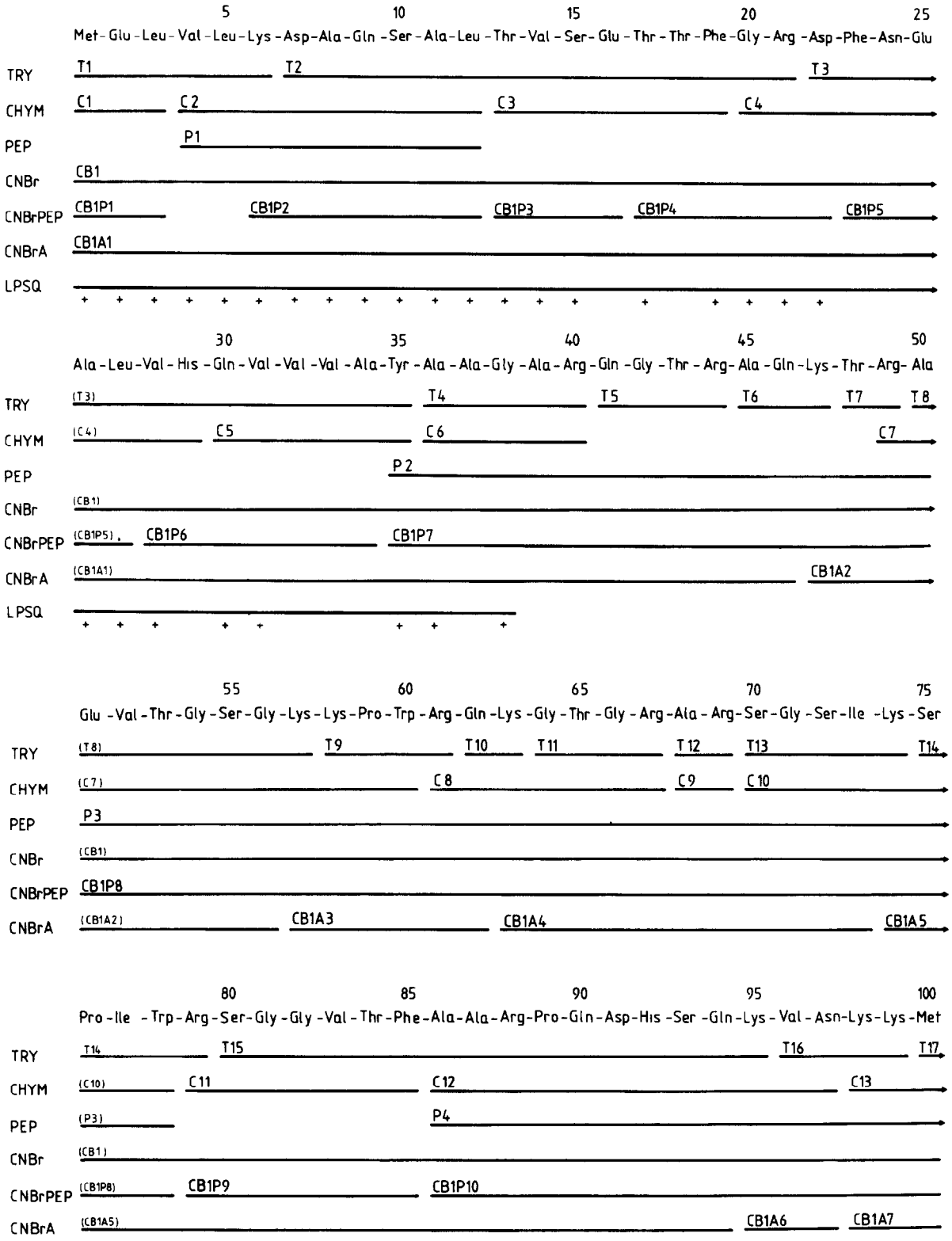
Here, we report the complete primary structure of L4 with 201 amino acids. We also depict the occurrence of secondary structure elements, e.g., α-helix and β-sheets, along the protein chain by using 4 prediction programmes. Finally, a comparison is made between the primary structure of protein L4 and that of ribosomal proteins from *E. coli* and other organisms.

2. Materials and methods

Protein L4 had been isolated as in [13] or [14] and was kindly provided by Drs H. G. Wittmann or J. Dijk, respectively. The identity and purity of the protein were checked by two-dimensional polyacrylamide gel electrophoresis [15].

Tryptic digestion of protein L4 was performed at pH 8.1 in 0.2 M *N*-methylmorpholine acetate buffer at 37°C for 4–6 h. Water was added to the lyophilized digests, and after centrifugation the soluble peptides were removed and separated by the fingerprint technique using thin-layer cellulose sheets [16]. The precipitate was washed with 10% acetic acid to remove any trapped soluble peptides and to obtain peptide T-26 (pos. 140–162) in purified form. Digestion of protein L4 with α-chymotrypsin was carried out at pH 8.1 in 0.2 M *N*-methylmorpholine acetate buffer at 37°C for 3 h. The peptides were separated by thin-layer fingerprinting. Protein L4 was also digested with pepsin in 5% formic acid at 37°C for 12 h. The peptides were separated by gel filtration on Sephadex G25, superfine (1 × 180 cm) in 10% acetic acid and purified by one-dimensional thin-layer chromatography.

Protein L4 was cleaved with CNBr in 70% formic acid at room temperature for 20 h, and the peptides were separated by gel filtration on Sephadex G75, superfine (1 × 210 cm) in 10% acetic acid. The large peptides resulting from CNBr cleavage (CB-1, CB-2 and CB-3) were further digested with pepsin or *Armillaria mellea* protease [17]. The procedure for peptic digestion of the peptide CB-1 and CB-3 was identical with that used for peptic digestion of protein L4. Digestions of peptides CB-1 and CB-2 with *A. mellea* protease were carried out at pH 8.1 in 0.2 M


5 10 15 20 25
Met-Glu-Leu-Val-Leu-Lys-Asp-Ala-Gln-Ser-Ala-Leu-Thr-Val-Ser-Glu-Thr-Thr-Phe-Gly-Arg-Asp-Phe-Asn-Glu
TRY T1 T2 T3
CHYM C1 C2 C3 C4
PEP P1
CNBr CB1
CNBrPEP CB1P1 CB1P2 CB1P3 CB1P4 CB1P5
CNBrA CB1A1
LPSQ
30 35 40 45 50
Ala-Leu-Val-His-Gln-Val-Val-Val-Ala-Tyr-Ala-Ala-Gly-Ala-Arg-Gln-Gly-Thr-Arg-Ala-Gln-Lys-Thr-Arg-Ala
TRY (T3) T4 T5 T6 T7 T8
CHYM (C4) C5 C6 C7
PEP P2
CNBr (CB1)
CNBrPEP (CB1P5), CB1P6 CB1P7
CNBrA (CB1A1) CB1A2
LPSQ
55 60 65 70 75
Glu-Val-Thr-Gly-Ser-Gly-Lys-Lys-Pro-Trp-Arg-Gln-Lys-Gly-Thr-Gly-Arg-Ala-Arg-Ser-Gly-Ser-Ile-Lys-Ser
TRY (T8) T9 T10 T11 T12 T13 T14
CHYM (C7) C8 C9 C10
PEP P3
CNBr (CB1)
CNBrPEP CB1P8
CNBrA (CB1A2) CB1A3 CB1A4 CB1A5
80 85 90 95 100
Pro-Ile-Trp-Arg-Ser-Gly-Gly-Val-Thr-Phe-Ala-Ala-Arg-Pro-Gln-Asp-His-Ser-Gln-Lys-Val-Asn-Lys-Lys-Met
TRY T14 T15 T16 T17
CHYM (C10) C11 C12 C13
PEP (P3) P4
CNBr (CB1)
CNBrPEP (CB1P8) CB1P9 CB1P10
CNBrA (CB1A5) CB1A6 CB1A7

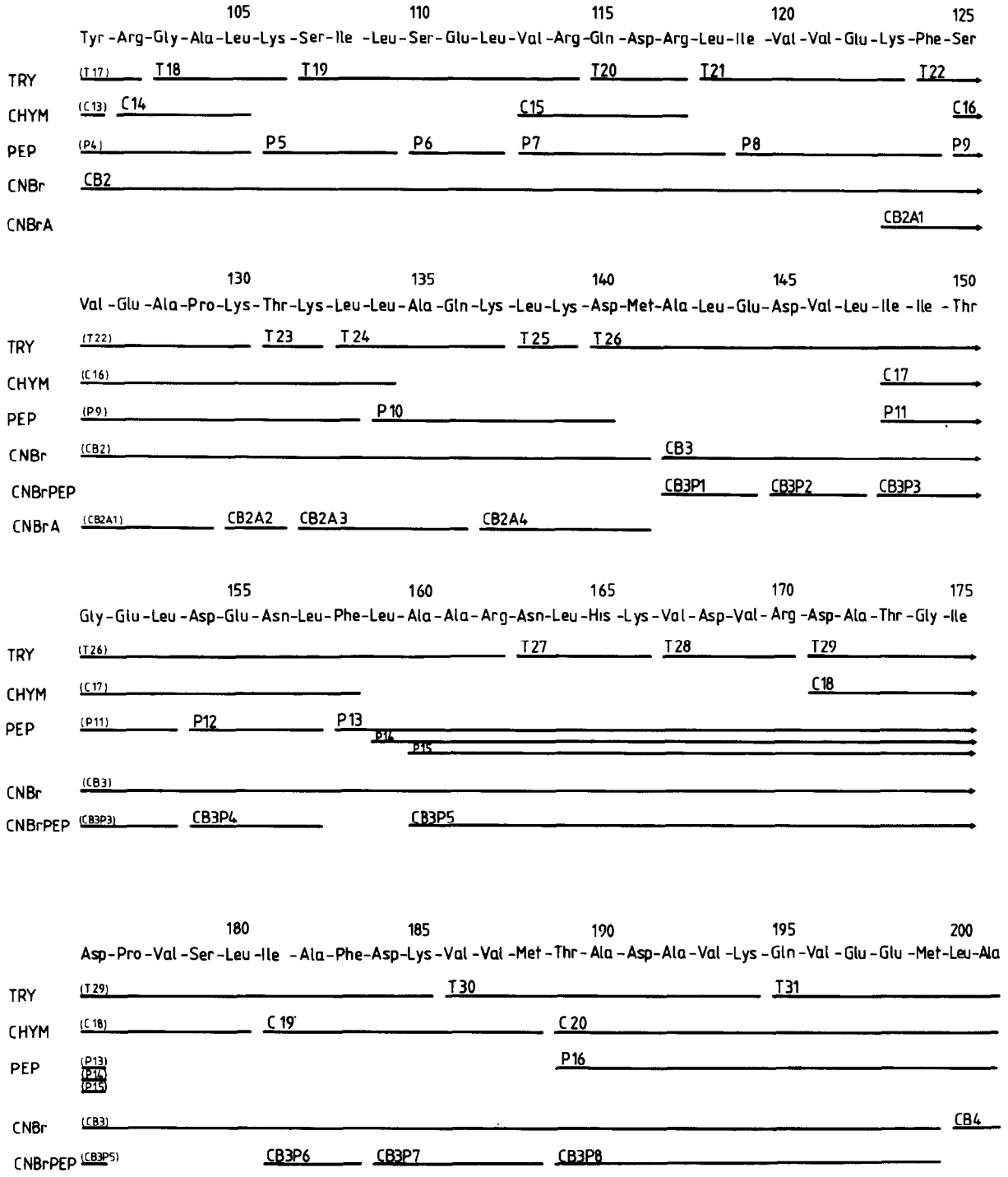


Fig.1. The primary structure of protein L4. *Abbreviations:* TRY, trypsin digestion (peptides T); CHYM, digestion with α-chymotrypsin (peptides C); PEPS, digestion with pepsin (peptides P); CNBr, cleavage with CNBr (peptides CB); CNBr-PEPS, digestion of CB peptides with pepsin (peptides CB-P); CNBr-A, digestion of CB peptides with *Armillaria mellea* protease (peptides CB-A); LPSQ, degradation of intact protein in an improved Beckman sequencer with an automatic conversion device; +, indicates where the observed spots on thin-layer sheets were unambiguously identified; (+), weakly identified.

N-methylmorpholine acetate buffer using 1:500 enzyme:peptide weight ratio at 37°C for 6 h. The peptides were separated by thin-layer fingerprinting.

The amino acid sequences of the various peptides were determined by the DABITC/PITC double-coupling method [18] using 1–5 nmol peptides. DABITH-Ile and -Leu were determined by one-dimensional thin-layer chromatography on polyamide plates [19]. Amino acid analyses were performed on Durrum D-500 analysers as in [16]. Tryptophan was determined photometrically (by Dr P. Woolley), and it was also detected by spraying fingerprint plates with Ehrlich reagent [20]. Cysteine was estimated as cysteic acid after performic acid oxidation of the native protein [21]. The Edman degradation on the intact protein was made as in [22].

3. Results and discussion

3.1. *Sequence determination*

The N-terminal sequence of L4 up to position 38 has been determined by automatic degradation of the intact protein. The remaining amino acid residues of protein L4 were completely sequenced by peptides resulting from tryptic, chymotryptic, peptic and *A. mellea* protease digestion of L4 or CNBr cleavage peptides.

Tryptic digestion of L4 gave 30 water-soluble and 1 water-insoluble peptide. The complete sequences of all peptides (except T-3 and T-26) were determined by the DABITC/PITC double-coupling method. The sequence of peptide T-3 (pos. 22–35) was obtained from the sequence of the chymotryptic peptides C-4

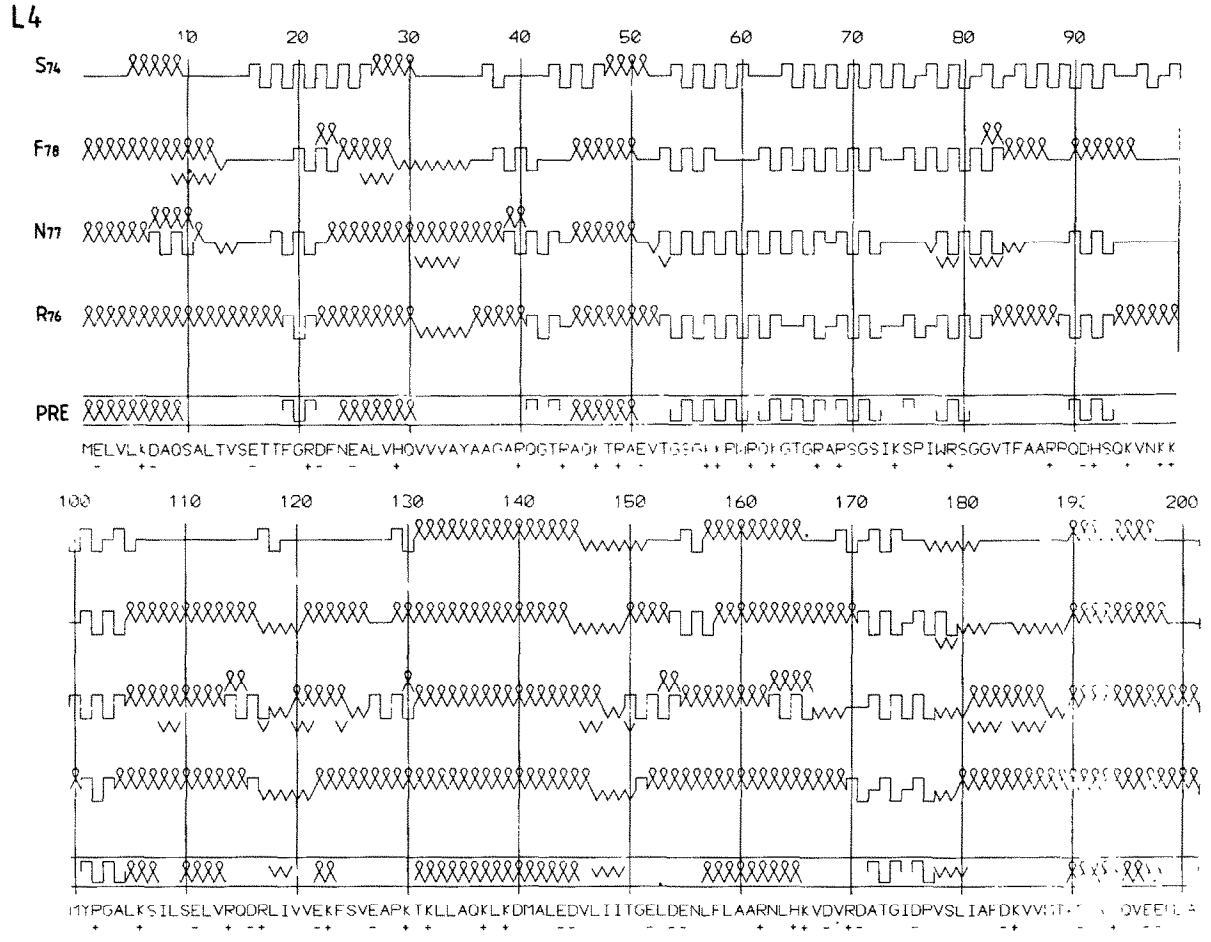


Fig.2. Secondary structure prediction of ribosomal protein L4 according to 4 methods (details in [23]). The symbols represent residues in a helical region (ΧΧΧΧΧΧΧ), β-turns or loop (⌐¬⌐¬⌐¬), extended structure (∿∿∿) and random coil conformation (——). The line 'PRE' summarizes the secondary structure obtained when at least 3 of the 4 predictions (S, F, N and R) are in agreement.

(pos. 20–29) and C-5 (pos. 30–35). The sequence of the peptide T-26 (pos. 140–162) was obtained from the peptic peptides CB-3P-1 (pos. 142–144), CB-3P-2 (pos. 145–147), CB-3P-3 (pos. 148–153), CB-3P-4 (pos. 154–157) and P-13 (pos. 158–176).

Twenty peptides (C-1–C-20) were obtained from the chymotryptic digestion and 16 peptides (P-1–P-16) from the peptic digestion of protein L4. They were sequenced if necessary.

To confirm the alignment of the tryptic peptides, L4 was treated with CNBr, and the resulting peptides were separated by gel filtration on Sephadex G75. Three of the 5 peptide bonds involving methionine residues (pos. 100, 141 and 199) were cleaved by CNBr, yielding 4 peptides (CB-1, CB-2, CB-3 and CB-4). The large peptides CB-1, CB-2 and CB-3 were further digested with pepsin or *A. mellea* protease. The peptides thus formed were separated by thin-layer fingerprinting and then sequenced. In this way the amino acid sequence of protein L4 was completely determined as shown in fig.1.

3.2. *Characteristics of the sequence*

The amino acid composition derived from the sequence of protein L4 is: Asp_{12}, Asn_4, Thr_{12}, Ser_{12}, Glu_{12}, Gln_{10}, Pro_5, Gly_{13}, Ala_{23}, Val_{21}, Met_5, Ile_8, Leu_{19}, Tyr_2, Phe_6, His_3, Lys_{18}, Arg_{14} and Trp_2. This composition is in agreement with the results obtained from the amino acid analysis of the protein. The M_r based on the sequence is 22 087.

The two tyrosine residues occur at pos. 35 and 101, and the tryptophan residues are located at pos. 60 and 78. Protein L4 contains 35 basic and 24 acidic amino acids. There are clusters of basic amino acids at pos. 40–49 and 57–88. Long regions with hydrophobic residues occur in the C-terminal half at positions 105–121 and 133–149.

3.3. *Secondary structure predictions of protein L4*

The prediction of secondary structure of L4 is presented in fig.2. It was made according to 4 predictive methods as detailed in [23]. Based on this prediction, helix regions occur at positions 1–9, 45–50, 105–113, 131–145, 157–165 and 190–198. A pronounced turn area was calculated at positions 54–72/80. The amount of secondary structure calculated for protein L4 is 32% helix, 19% β-turn and 4% extended structure (average values, see line 'PRE' in fig.2).

3.4. *Comparison with other ribosomal proteins*

The sequence of protein L4 was by computer search programme compared with other ribosomal protein sequences [24,25]. Protein L4 shares the pentapeptides Ala–Asp–Ala–Val–Lys (pos. 190–194), Gln–Val–Val–Val–Ala (pos. 30–34) and Ala–Leu–Glu–Asp–Val (pos. 142–146) with *E. coli* proteins S11 (pos. 70–74), protein S16 (pos. 18–22) and *H. cutirubrum* protein L20 (pos. 44–48), respectively. However, no identical regions with >5 adjacent amino acids were found. When 'conservative' replacements, such as glutamic acid for aspartic acid, isoleucine for valine, arginine for lysine and serine for threonine or vice versa were allowed, the N-terminal region (pos. 40–70) of L4 is similar to regions in several other ribosomal proteins as shown in fig.3.

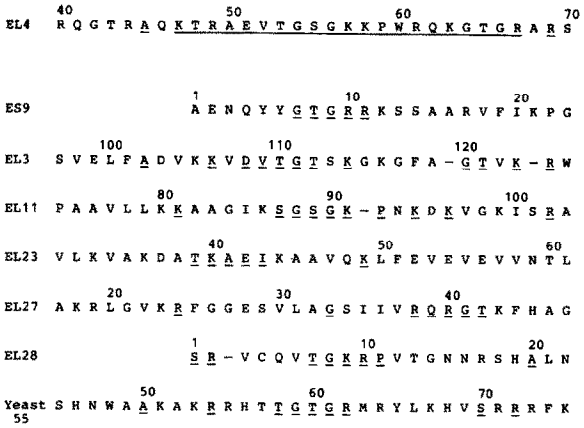


Fig.3. Comparison of homologous regions from protein L4 and other ribosomal proteins.

Acknowledgements

We thank Dr H. G. Wittmann for many stimulating discussions; Dr V. Barkholt Pederson, Institute for Biokemisk Genetik, Copenhagen, for kindly supplying us with a sample of *Armillaria mellea* protease and Mr K. Ashman and Mr N. D. Rawlings for performing computer analyses.

References

[1] Stöffler, G., Daya, L., Rak, K. H. and Garrett, R. A. (1971) J. Mol. Biol. 62, 411–414.
[2] Zimmermann, R. A. (1974) in: Ribosomes (Lengyel, P. et al. eds) pp. 225–269, Cold Spring Harbor Lab. Press, Long Island NY.

[3] Maly, P., Rinke, J., Ulmer, E., Zwieb, C. and Brimacombe, R. (1980) Biochemistry 19, 4179–4190.
[4] Traut, R. R., Lambert, J. M., Boileau, G. and Kenny, J. W. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 89–110, Univ. Park Press, Baltimore MD.
[5] Roth, H. E. and Nierhaus, K. H. (1975) J. Mol. Biol. 94, 111–121.
[6] Stöffler, G., Bald, R., Kastner, B., Lührmann, R., Stöffler-Meilicke, M., Tischendorf, G. and Tesche, B. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 171–205, Univ. Park Press, Baltimore MD.
[7] Spillmann, S., Dohme, F. and Nierhaus, K. H. (1977) J. Mol. Biol. 115, 513–523.
[8] Nierhaus, K. H. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 267–294, Univ. Park Press, Baltimore MD.
[9] Otaka, E., Teraoka, H., Tamaki, M., Tanaka, K. and Osawa, S. (1970) J. Mol. Biol. 48, 499–510.
[10] Wittmann, H. G., Stöffler, G., Apirion, D., Rosen, L., Tanaka, K., Tamaki, M., Takata, R., Dekio, S., Otaka, E. and Osawa, S. (1973) Mol. Gen. Genet. 127, 175–189.
[11] Dabbs, E. R. (1978) Mol. Gen. Genet. 165, 73–78.
[12] Otaka, E., Itoh, T., Osawa, S., Tanaka, K. and Tamaki, M. (1971) Mol. Gen. Genet. 114, 14–22.
[13] Hindennach, I., Kaltschmidt, E. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 12–16.
[14] Dijk, J. and Littlechild, J. A. (1979) Methods Enzymol 59, 481–502.
[15] Kaltschmidt, E. and Wittmann, H. G. (1970) Anal. Biochem. 36, 401–412.
[16] Hitz, H., Schäfer, D. and Wittmann-Liebold, B. (1977) Eur. J. Biochem. 75, 497–512.
[17] Lewis, W. G., Basford, J. M. and Walton, P. L. (1978) Biochim. Biophys. Acta 522, 551–560.
[18] Chang, Y. J., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[19] Yang, Y. C. (1979) Hoppe-Seyler's Z. Physiol. Chem. 360, 1673–1675.
[20] Needleman, S. B. (1975) in: Protein Sequence Determination. Molecular Biology, Biochemistry and Biophysics, vol. 8, p. 201, Springer-Verlag, Berlin, Heidelberg, New York.
[21] Needleman, S. B. (1975) in: Protein Sequence Determination. Molecular Biology, Biochemistry and Biophysics, vol. 8, p. 129, Springer-Verlag, Berlin, Heidelberg, New York.
[22] Wittmann-Liebold, B., Geissler, A. W. and Marzinzig, E. (1975) J. Supramol. Struct. 3, 426–447.
[23] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.
[24] Wittmann, H. G., Littlechild, J. and Wittmann-Liebold, B. (1980) in: Ribosomes (Chambliss, G. et al. eds), pp. 51–88, Univ. Park Press, Baltimore MD.
[25] Matheson, A. T., Möller, W., Amons, R. and Yaguchi, M. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 297–332, Univ. Park Press, Baltimore MD.