PRIMARY STRUCTURE OF PROTEIN L33 FROM THE LARGE SUBUNIT OF THE *ESCHERICHIA COLI* RIBOSOME

Brigitte WITTMANN-LIEBOLD and Renate PANNENBECKER
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, Germany

Received 28 June 1976

1. Introduction

Protein L33 is one of the very basic proteins in the *E. coli* 50S ribosomal subunit [1]. Like protein L34 whose primary structure has been recently established [2] it is one of the smallest proteins of the *E. coli* ribosome. There is evidence that L33 is one of the ribosomal protein components involved in the elongation process; various authors, using affinity labelling techniques, have investigated those proteins which are involved in the ribosomal environment of the peptidyl residue of peptidyl tRNA bound to the puromycin-sensitive site [3–5]. Similar groups of proteins were found in each case, namely L27, L2, L13, L14, L15, L16 and L32 or L33. Among these proteins, the primary structures have already been established for proteins L16 [6], L27 [7] and L32 [8].

In this paper we report the complete amino acid sequence of protein L33. This protein consists of 54 amino acid residues; and its N-terminal residue is an unusual amino acid, namely *N*-methylalanine. Sequence similarities with other ribosomal proteins are summarized and predictions for the secondary structure of L33 are presented.

2. Materials and methods

Protein L33 isolated from *E. coli* strain K as described earlier [9] was provided by Dr H. G. Wittmann. The identity and purity of the protein was determined by two-dimensional polyacrylamide gel electrophoresis [10] and by sequenator degradations of the intact protein. Enzymatic digestions employing trypsin (Merck, Darmstadt) treated with 1-chloro-4-phenyl-3-tosylamidobutan-2-one (TPCK) and thermolysin from Serva (Heidelberg) were performed at pH 8.1 in 0.1 M *N*-methyl-morpholine acetate buffer at 37°C for 4 and 1 h, respectively. Digestion with *Staphylococcus aureus* protease [11] from Miles Laboratories, Slough, UK, was carried out in 0.05 M ammonium acetate buffer, pH 4.0, for 48 h and 37°C. The isolation of the tryptic peptides, and those derived from thermolysin digestion, was achieved by fingerprinting on cellulose thin-layer plates with 10 nmol enzymatic hydrolysate per plate. In addition, these peptides were isolated on a micro column of Dowex 50 × 7 (0.3 × 10 cm) with 300 nmol of protein hydrolysate. The isolation and purification of peptides obtained by *Staphylococcal* protease digestion of protein L33 were performed on a Sephadex G25 (fine) column (180 × 1.5 cm) at room temperature in 5% acetic acid, followed by thin-layer chromatography.

Carboxypeptidase A and B from Boehringer (Mannheim) was used for exopeptidase treatment of the intact protein, in 0.1 M *N*-methyl-morpholine acetate buffer, pH 8.1, for 10 to 120 min and 37°C. Amino acid analyses were performed with the single column procedure on a Durrum amino acid analyzer (model D-500) in the 1–2 nmol range and on a LKB analyzer (model 3201) in the 2.5 nmol range. Total protein samples were hydrolyzed for 20, 48 and 72 h and analyzed on a Durrum analyzer in the 10 nmol range.

Sequence determinations on the intact protein were made using the degradation procedure [12] in an improved Beckman sequenator [13,14] with 0.5–1 mg protein amounts. Sequencing of the

peptides isolated from the C-terminal region was performed either with the solid phase sequencing technique according to Laursen [15] under the conditions published recently [16], or by the combined Dansyl-Edman method [17,18].

Mass spectrometry was performed in a Varian CH-7 apparatus with an electron impact ionic source and with direct sample injection as described elsewhere [14].

3. Results and discussion

Protein L33 has the following amino acid composition: Asp_1, Asn_1, Thr_5, Ser_2, Glu_4, Gln_1, Pro_2, Gly_3, Ala_2, Val_4, Ile_4, Leu_3, Tyr_2, Phe_2, His_2, Lys_{12}, Arg_3 and *N*-methylalanine$_1$ (NMA). Appropriate methods for the detection of methionine, tryptophan and cysteine showed that these three amino acids are all absent. Protein L33 contains 54 amino acid residues and has a mol. wt. of 6255. Its primary structure is shown in fig.1. The number of amino acid residues is in excellent agreement with the results calculated from the total hydrolysis of the intact protein. The content of 17 basic amino acids which comprize 32% of all residues in this protein and of only five acidic amino acids is in very good agreement with the high isoelectric point of L33 [19].

The presence of *N*-methylalanine as the N-terminal

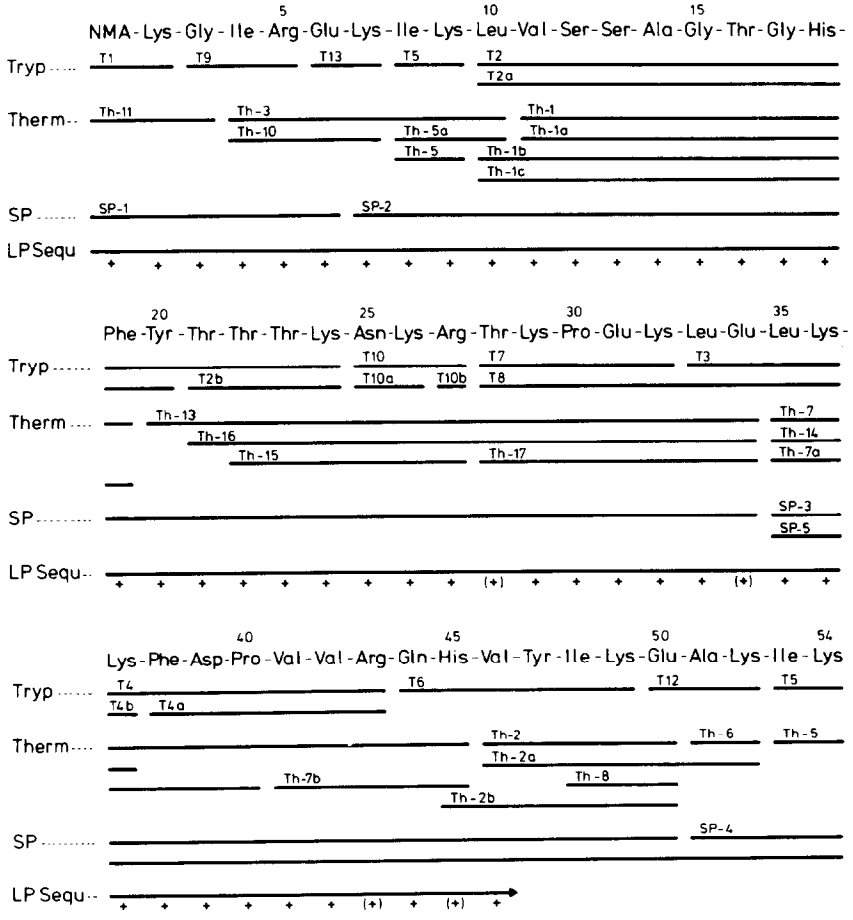

Fig.1. Primary structure of ribosomal protein L33. Tryp = tryptic digestion; Therm = thermolysin digestion; SP = hydrolysis with *Staphylococcus aureus* protease at pH 4.0; Liquid P. Sequ. = Edman degradation of intact protein in an improved Beckman sequenator including automatic conversion reaction; + means unambiguous identification of the released phenylthiohydantoin derivatives by thin-layer chromatography.

residue was proven by mass spectrometry of its phenylthiohydantoin derivative after Edman degradation of the intact protein as well as by amino acid analysis of the tryptic peptide T_1 (positions 1–2) and of the thermolysin peptide Th-11 (positions 1–3). Finally, dansylation of the protein and of peptides T_1 and Th-11 gave rise to an unusual dansyl amino acid which moved with dansyl *N*-methylalanine. Unusual N-methylated end groups in ribosomal proteins have so far been found for proteins S11 (R. Chen, unpublished) and for protein L16 [6]. The experimental details for the identification of these N-methylated amino acids will be reported elsewhere (Chen, R., Brosius, J., Wittmann-Liebold, B., and Schäfer, W., to be published). It is probable that the unidentified methylated amino acid in L33 reported previously [20] is the methylated alanine at the N-terminus.

The amino acid sequence was derived mainly from automatic degradations performed with unmodified protein samples and this established the sequence up to position 46. Furthermore, the sequence of this region was independently proven by isolation of peptides derived from trypsin and thermolysin hydrolysis; the N-terminal sequences of these peptides were obtained by the combined Dansyl-Edman technique.

The sum of the amino acid residues of the three peptides Sp-1, Sp-2, and Sp-5 isolated from the *Staphylococcal* protease digest of protein L33 (see fig.1) was exactly the same as the total amino acid composition of this protein. Whereas SP-1 and Sp-2 could be correlated with the N-terminal region up to position 34, SP-5 was shown by further digestion with trypsin to contain a dipeptide (Leu-Lys) from T_3, and the peptides T_4, T_6, T_{12} and T_5. Peptides Sp-3 and Sp-4, which together comprize the amino acid residues of SP-5, have been isolated and sequenced. The alignment of T_3–T_4–T_6–T_{12}–T_5 was in agreement with the isolation of various thermolysin peptides whose end groups were examined as well (see fig.1). Tryptic peptide T_5 occurs two times within the sequence of L33, namely in positions 8–9 and 53–54. Exopeptidase A and B resulted in the release of lysine and isoleucine. A summary of the sequence determinations is presented in fig.1. The localization of the acidic amino acids and their amides was determined by the electrophoretic behavior of the corresponding peptides and from the sequenator results; these results are in accordance with the specificity of the *Staphylococcal* protease at pH 4.0. Full details of the applied methods and of the results will be described elsewhere.

A comparison of the amino acid sequence of L33 with the other ribosomal proteins of known primary structure, cf. [21] revealed a few sequence similarities which are summarized in table 1.The secondary structure of L33 was predicted using the method of Burgess et al. [22]; α-helix regions were calculated for positons 5–10 and positions 31–36. No β-sheet structure was predicted, but bends should be present in positions 13–22; 26–27; 29–30; 37–38 and 40–41. Alternatively, following the rules of Chou and Fasman [23,24] the following secondary structure is predicted: α-helix for the positions 6–11 and 31–36; β-sheet structure for positions 38–46 and 16–20. Although according to Chou and Fasman's rules a helical region for positions 41–51 is also possible, both methods agree very well in their prediction of the two strong α-helical regions.

Table 1
Sequence correlations of ribosomal protein L33 with other proteins from *E. coli* ribosomes

	Protein	Positions	Protein	Positions
Tyr–Thr–Thr–Thr	L33	20–23	S12	37–40
Lys–Arg–Thr–Lys	L33	26–29	L16	5– 8
Thr–Lys–Pro–Glu	L33	28–31	S8	54–57
Val–Val–Arg–Gln	L33	41–44	S9	46–49
Ile–Lys–Glu–Ala	L33	48–51	L14	43–46
Lys–Glu–Ala–Lys	L33	49–52	L7/L12	81–84
Ile–Arg–Glu–Lys	L33	4– 7	L1	10–13
			S4	142–145

Acknowledgements

We are grateful to Dr H. A. Scheraga for kindly providing his computer program. The excellent technical assistance of Mr Horst Graffunder and of Mr Alfred Franz is greatly appreciated. We also thank Dr Richard Brimacombe for revising the English version of the manuscript.

References

[1] Wittmann, H. G. and Wittmann-Liebold, B. (1974) in: Ribosomes, Cold Spring Harbor Monograph Series, New York, p. 115–140.
[2] Chen, R. and Ehrke, G. (1976) FEBS Lett. 63, 215–217.
[3] Eilat, D., Pellegrini, M., Oen, H., Lapidot, Y. and Cantor, C. R. (1974) J. Mol. Biol. 88, 831–840.
[4] Hsiung, N., Reines, S. A. and Cantor, C. R. (1974) J. Mol. Biol. 88, 841–855.
[5] Czernilofsky, A. P., Collatz, E. E., Stöffler, G. and Kuechler, E. (1974) Proc. Nat. Acad. Sci. USA 71, 230–234.
[6] Brosius, J. and Chen, R. (1976) FEBS Lett. in press.
[7] Chen, R., Mende, L. and Arfsten, U. (1975) FEBS Lett. 59, 96–98.
[8] Wittmann-Liebold, B., Greuer, B. and Pannenbecker, R. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1977–1979.
[9] Hindennach, I., Kaltschmidt, E. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 12–16.
[10] Kaltschmidt, E. and Wittmann, H. G. (1970) Anal. Biochem. 36, 401–412.
[11] Houmard, J. and Drapeau, G. R. (1972) Proc. Nat. Acad. Sci. USA 69, 3506–3509.
[12] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[13] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[14] Wittmann-Liebold, B., Geissler, A. W. and Marzinzig, E. (1975) J. Supramol. Struct. 3, 426–447.
[15] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[16] Wittmann-Liebold, B. and Lehmann, A. (1975) in: Solid phase methods in protein sequence analysis (R. A. Laursen ed.) Pierce Chemical Company Rockford, USA p. 81–90.
[17] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[18] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.
[19] Kaltschmidt, E. (1971) Analytical Biochem. 43, 25–31.
[20] Chang, C. N. and Chang, F. N. (1975) Biochemistry 14, 468–477.
[21] Wittmann-Liebold, B. and Dzionara, M. (1976) FEBS Lett. 61, 14–19.
[22] Burgess, A. W., Ponnuswamy, P. K. and Scheraga, H. A. (1974) Israel J. Chem. 12, 239–286.
[23] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 211–222.
[24] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 222–245.