THE PRIMARY STRUCTURE OF L11, THE MOST HEAVILY METHYLATED PROTEIN FROM *ESCHERICHIA COLI* RIBOSOMES

Mai J. DOGNIN and Brigitte WITTMANN-LIEBOLD
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, Germany

Received 28 October 1977

1. Introduction

Protein L11 from the large ribosomal subunit of *E. coli* was reported to be part of the peptidyltransferase centre [1–4]. However, a 50 S derived core lacking L11 was described to be fully active in peptidyltransferase activity [5,6] and L16 was found to be essential for this activity [7]. Recently, this controversy was settled by showing that L11 is important for the assembly of L16 when all the components are present in stoichiometric amounts. The importance of L11 can be reduced by a ten-fold excess of L16 [8].

Protein L11, which binds specifically to 23 S ribosomal RNA when prepared under non-denaturing conditions [9], can be cross-linked to proteins L7/12 and L10 [10,11]. L11 has been identified by photoaffinity labelling as one of the proteins involved in EF-G-dependent GDP binding [12].

Immune electron microscopy revealed two antibody binding sites for L11 on the 50 S subunit and therefore indicates a slightly elongated shape for this protein [13]. This conclusion is in agreement with physical studies on the isolated protein resulting in an axial ratio of 5–6:1 [14].

Several mutants with altered L11 have been isolated [15–17] and two of them have a relaxed control of RNA synthesis [18]. This indicates that L11 is a participant in the synthesis of ppGpp and pppGpp.

Protein L11 is the most extensively methylated of the *E. coli* ribosomal proteins. The methyl groups were found in one residue of lysine as N_ϵ-trimethyllysine, and in one unidentified neutral amino acid [19,20]. Recently, *N*-trimethylalanine was identified in the mixture of 50 S proteins and was assumed to be the unidentified neutral amino acid in L11 [21]. Furthermore, two other amino acids, N_ϵ-monomethyllysine and dimethylarginine, were reported to be also present in L11 [17,20,22].

This paper summarizes the determination of the complete amino acid sequence of protein L11 including the nature, quantity and location of the different methylated derivatives. Protein L11 consists of 141 amino acids and the blocked N-terminal residue was shown to be *N*-trimethylalanine. Furthermore, two N_ϵ-trimethyllysines were found in this protein. No other methylated amino acid residues could be found in the sequence determination of protein L11.

2. Materials and methods

Protein L11 was isolated from *E. coli* strain K by a procedure to be described elsewhere [23a]. Identity and purity of the protein were checked by two-dimensional polyacrylamide gel electrophoresis [24].

Digestions were performed with TPCK-trypsin (Worthington Biochemicals Corp., Freehold, NJ) at 37°C for 4 h, with TLCK-chymotrypsin (Merck, Darmstadt) at 37°C for 1 h, with thermolysin (Serva, Heidelberg) at 50°C for 1 h, and with *Staphylococcus aureus* protease [25] (Miles Biochemicals, Elkhart, IN) at 37°C for 18 h and 40 h. Details of the enzymatic methods have been described [26–28]. Digestion with carboxypeptidase C (Boehringer, Mannheim) was essentially done according to [29] at pH 5.3 and 50°C for periods up to 4 h. Specific cleavage was performed:

North-Holland Publishing Company – Amsterdam

(i) At arginine residues by modification of the lysine residues with ETPA (exo-*cis*-3,6-endoxo-Δ^4-tetrahydrophthalic anhydride) [30] followed by digestion with trypsin.
(ii) At lysine residues by modification of the arginine residues with 1,2-cyclohexanedione using a slight variation of the procedure described [31] followed by tryptic cleavage of the modified protein.
(iii) At methionine residues by treatment with CNBr for 20 h at room temperature in the dark [32].

Fingerprints of peptides after digestions with different enzymes were made on cellulose thin-layer plates [26–28] and the peptides were extracted with 5.7 N HCl containing 0.02% mercaptoethanol for analysis of their amino acid composition, or with 30–50% acetic acid for their sequence analysis. The separation of peptides on a preparative scale was made by gel filtration on Sephadex G-50 or G-25 superfine (140 × 1 cm) in 10% acetic acid [28] followed by ion-exchange chromatography on a microcolumn of Dowex M71 (0.3 × 10 cm) at 55°C using pyridine formate gradients [28,33]. Alternatively, the peptide separation was accomplished by preparative electrophoresis or chromatography on thin-layer cellulose plates. In a few cases combination of all the methods was necessary.

Amino acid analyses of protein and peptides were performed on a Durrum D-500 and on a LKB-Biocal 3201 analyser as described previously [28,34]. For determination of cysteine the protein and peptides were oxidized with performic acid [35a] before acid hydrolysis. The presence of tryptophan was determined by amino acid analysis after hydrolysis with methane sulfonic acid [36] and by spraying of fingerprint plates with *p*-aminobenzaldehyde [35b].

Amino acid sequences of peptides were determined by the combined dansyl-Edman technique [37,38] and by automatic Edman degradation [39] either in an improved Beckman sequenator [40] or in a solid phase sequenator with attachment of the C-terminal groups to amino-polystyrene resin by 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide [41,42].

Identification of aspartic acid, glutamic acid and their amides was achieved by converting the released 2-anilino-5-thiazolinone derivatives to the phenylthiohydantoin derivatives followed by identification with thin-layer chromatography [40]. When identification after dansylation was uncertain, the free amino acids were liberated from their 2-anilino-5-thiazolinone derivatives by hydrolysis and identified on an amino acid analyser.

N_ϵ-Trimethyllysine was identified by amino acid analysis and by the dansyl-Edman technique, and *N*-trimethylalanine by field-desorption mass spectrometry.

3. Results and discussion

Protein L11 has an amino acid composition of Asp_6 Asn_4 Thr_9 Ser_8 Glu_6 Gln_7 Pro_9 Gly_{12} Ala_{19} Val_{13} Met_5 Ile_9 Leu_7 Tyr_2 Phe_4 TML‡ Lys_{13} Arg_4 Cys_1 TMA†. Histidine and tryptophan are absent. This number of amino acids obtained from the sequence (fig.1) agrees very well with the amino acid composition of the protein. The mol. wt 14 874 calculated from the sequence of L11 is in very good agreement with the value of 15 300 determined by SDS–gel electrophoresis [43].

The strategy of the sequence determination of protein L11 was based on three digestions of the protein: with trypsin, with Staphylococcal protease and with trypsin after modification of the lysine residues with ETPA. The main procedure used to determine the sequences of the peptides obtained by these digestions was the manual dansyl-Edman technique. This method, although time-consuming, made it possible to sequence difficult regions, i.e., regions with accumulation of prolines and hydrophobic residues (e.g., positions 19–25, 52–57, 73–79), unambiguously. Furthermore, the method is efficient: e.g., peptide T2 could be sequenced up to 28 steps with 60 nmol peptide. More material was necessary when long peptides, e.g., T2, SP2, SP4 and SP8, had to be degraded further by other enzymes.

In addition to the manual dansyl-Edman technique an automatic Edman degradation of the peptides was performed by two methods:
(i) With a solid phase sequenator which analyzed most of the tryptic peptides and also peptides ET4 and SP10.
(ii) With an improved Beckman sequenator which analyzed three peptides (T2, T4 and SP8) after reaction with Braunitzer's reagent IV [44].

*TMA, *N*-trimethylalanine; TML, N_ϵ-trimethyllysine

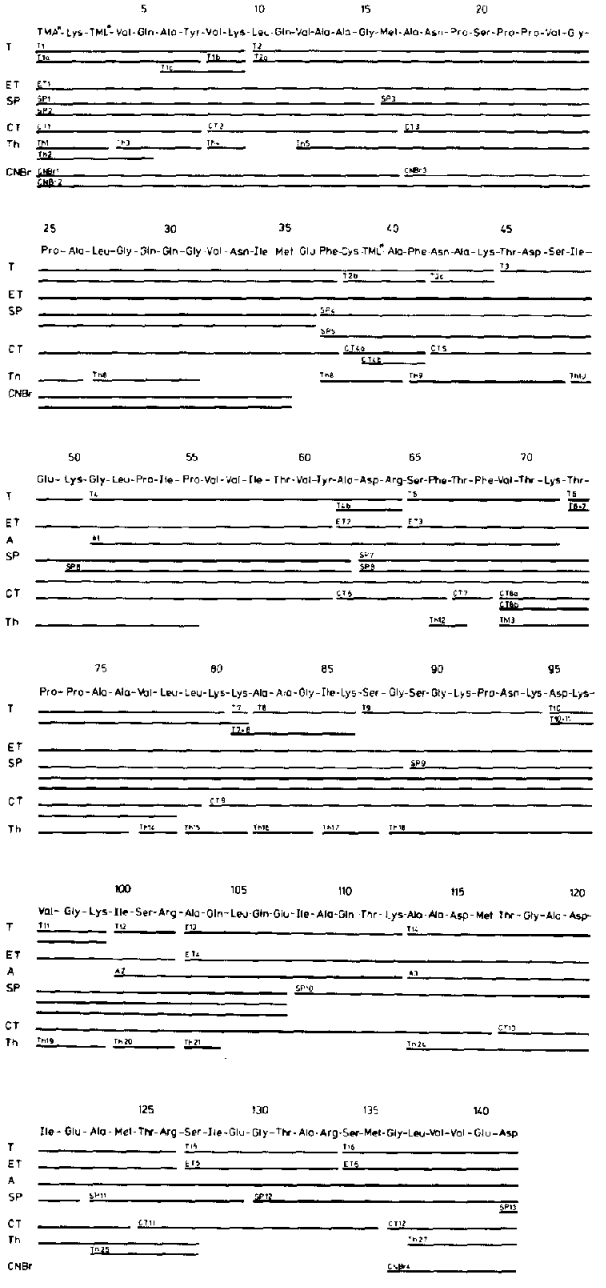

Fig.1. The primary structure of protein L11 from *Escherichia coli*. T, Tryptic peptides; ET, tryptic peptides after blocking the lysines by ETPA; A, tryptic peptides after blocking the arginines by 1,2-cyclohexanedione; SP, peptides from digestion with *Staphylococcus aureus* protease; CT, peptides from digestion with chymotrypsin; Th, peptides from digestion with thermolysin; CNBr, peptides resulting from CNBr cleavage.

The determination of the C-terminal sequence proved difficult as it could not be identified by treatments with carboxypeptidases A, B or C. This is explained by the presence of Glu-Asp at the C-terminal, which was unambiguously identified by correlation of results from all digestions (details will be given elsewhere [23a]).

Determination of the N-terminal sequence was difficult because protein L11 could not be degraded, either with automatic or with manual techniques. The blocked amino acid at the N-terminus does not show any reaction with ninhydrin or with any of the usual amino acid reagents. Furthermore, hydrolysis with hydrochloric acid did not unblock this amino acid so that it was never detected on amino acid analysers. However, we succeeded in determining the nature of this amino acid and its blocking group by mass spectrometry. More details on the determination of the N-terminal sequence of L11 will be published elsewhere [23b].

Only four points will be mentioned in this context:

1. The presence of N_ϵ-monomethyllysine or dimethylarginine in protein L11 [17,20,22] could not be confirmed by the sequence determination. However, a content of less than 10% of such amino acids would not be detected by the sequencing methods applied here.

2. Two N_ϵ-trimethyllysines were found in L11 in position 3 and 39. Previous reports [17,19,22] have indicated the presence of only one N_ϵ-trimethyllysine residue.

3. The blocked N-terminal amino acid of protein L11 was identified as *N*-trimethylalanine confirming that the new methylated amino acid found in the 50 S subunit [21] belongs to L11. This however disagrees with the view that the unknown methylated amino acid of L11 is a dimethyl compound not derived from alanine [45].

4. The total number of methyl groups in L11 is nine: three in the N-terminal trimethylalanine and six in the two trimethyllysine residues. Only six methyl groups have previously been found to be incorporated in an unmethylated mutant of protein L11 [17].

The secondary structure was predicted using computer programmes based on four different methods [46–48]. According to these calculations protein L11 contains about 34–37% α-helical regions, 11–14% β-structure, and 17–22% turns (or loops).

The primary structure of L11 and of other ribosomal proteins of known sequence were compared by a computer programme. Seven tetrapeptides of L11 were found to be identical with peptides belonging to proteins S7, S13, S19, L6 and L7/12. One pentapeptide of L11 was identical with one peptide of protein L27, and one hexapeptide with one of protein L16. This degree of homology can be explained on a random basis [49].

Acknowledgements

We thank Mr Alfred Franz and Mr Arnold Lehmann for excellent technical assistance and Dr Paul Woolley for kindly reading the manuscript.

References

[1] Nierhaus, K. H. and Montejo, V. (1973) Proc. Natl. Acad. Sci. USA 70, 1931–1935.
[2] Hsiung, N., Reines, S. A. and Cantor, C. R. (1974) J. Mol. Biol. 88, 841–855.
[3] Tate, W. P., Caskey, C. T. and Stöffler, G. (1975) J. Mol. Biol. 93, 375–389.
[4] Dietrich, S., Schrandt, I. and Nierhaus, K. H. (1974) FEBS Lett. 47, 136–139.
[5] Ballesta, J. P. G. and Vazquez, D. (1974) FEBS Lett. 48, 266–270.
[6] Howard, G. A. and Gordon, J. (1974) FEBS Lett. 48, 271–274.
[7] Moore, V. G., Atchison, R. E., Thomas, G., Moran, M. and Noller, H. F. (1975) Proc. Natl. Acad. Sci. USA 72, 844–848.
[8] Schulze, H., Hampl, H. and Nierhaus, K. H., manuscript in preparation.
[9] Littlechild, J., Dijk, J. and Garrett, R. A. (1977) FEBS Lett. 74, 292–294.
[10] Expert-Bezançon, A., Barritault, D., Millet, M., Khouvine, Y. and Hayes, D. H. (1975) FEBS Lett. 59, 64–69.
[11] Expert-Bezançon, A., Barritault, D., Millet, M. and Hayes, D. H. (1976) J. Mol. Biol. 108, 781–787.
[12] Maassen, J. A., Möller, W. (1974) Proc. Natl. Acad. Sci. USA 71, 1277–1280.
[13] Tischendorf, G. W., Zeichhardt, H. and Stöffler, G. (1975) Proc. Natl. Acad. Sci. USA 72, 4820–4824.
[14] Giri, L., Dijk, J., Labischinski, H., Bradaczek, H. (1977) Biochemistry in press.
[15] Isono, K., Krauss, J. and Hirota, Y. (1976) Mol. Gen. Genet. 149, 297–302.
[16] Isono, K., Cumberlidge, A. G. and Isono, S. (1977) Mol. Gen. Genet. 152, 239–243.
[17] Colson, C. and Smith, H. O. (1977) Mol. Gen. Genet. 154, 167–173.
[18] Parker, J., Watson, R. J., Friesen, J. D., Fiil, N. P. (1976) Mol. Gen. Genet. 144, 111–114.
[19] Alix, J. H. and Hayes, D. (1974) J. Mol. Biol. 86, 139–159.
[20] Chang, F. N., Chang, C. N. and Paik W. K. (1974) J. Bacteriol. 120, 651–656.
[21] Lederer, F., Alix, J. H and Hayes D. (1977) Biochem. Biophys. Res. Commun. 77, 470–480.
[22] Chang, C. N. and Chang, F. N. (1975) Biochemistry 14, 468–477.
[23a] Dognin, M. J., Wittmann-Liebold, B., manuscript in preparation.
[23b] Dognin, M. J., Wittmann-Liebold, B., manuscript in preparation.
[24] Kaltschmidt, E. and Wittmann, H. G. (1970) Analyt. Biochem. 36, 401–412.
[25] Houmard, J. and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[26] Bitar, K. G. and Wittmann-Liebold, B. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1343–1352.
[27] Stadler, H. and Wittmann-Liebold, B. (1976) Eur. J. Biochem. 66, 49–56.
[28] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.
[29] Tscheche, H. and Kupfer, S. (1972) Eur. J. Biochem. 26, 33–36.
[30] Riley, M. and Perham, S. (1970) Biochem. J. 118, 733–739.
[31] Toi, K., Bynum, E., Norris, E. and Itano, H. (1967) J. Biol. Chem. 242, 1036–1043.
[32] Gross, E. and Witkop, B. (1961) J. Am. Chem. Soc. 83, 1510.
[33] Chen, R. and Wittmann-Liebold, B. (1975) FEBS Lett. 52, 139–140.
[34] Hitz, H., Schäfer, D. and Wittmann-Liebold, B. (1977) Eur. J. Biochem. 75, 496–512.
[35a] Needleman, S. B. (1975) in: Protein Sequence Determination, Molecular Biology, Biochemistry and Biophysics 8, p. 129, Springer-Verlag, Berlin, Heidelberg, New York.
[35b] Ibid. p. 201.
[36] Liu, T. Y. and Chang, Y. H. (1971) J. Biol. Chem. 246, 2842–2848.
[37] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[38] Bruton, C. J. and Hartley, B. S. (1970) J. Mol. Biol. 52, 165–178.
[39] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.

[40] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[41] Previero, A., Derancourt, J., Coletti-Previero, M. A. and Laursen, R. A. (1975) FEBS Lett. 33, 135–138.
[42] Wittmann-Liebold, B. and Lehmann, A. (1975) in: Solid Phase Methods in Protein Sequence Analysis (Laursen, R. A. ed) pp. 81–90, Pierce Chemical Co., Rockford USA.
[43] Zimmermann, R. A. and Stöffler, G. (1976) Biochemistry 15, 2007–2017.
[44] Braunitzer, G., Schrank, B., Ruhfus, A., Petersen, S. and Petersen, U. (1971) Hoppe-Seyler's Z. Physiol. Chem. 352, 1730–1732.
[45] Chang, F. N. and Budzilowicz, C. (1977) J. Bacteriol. 131, 105–110.
[46] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.
[47] Wittmann-Liebold, B., Robinson, S. M. L. and Dzionara, M. (1977) FEBS Lett. 81, 204–213.
[48] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) J. Supramol. Struct. in press.
[49] Wittmann-Liebold, B. and Dzionara M. (1976) FEBS. Lett. 65, 281–283.