THE PRIMARY STRUCTURE OF PROTEIN L20 FROM THE LARGE SUBUNIT OF THE *ESCHERICHIA COLI* RIBOSOME

Brigitte WITTMANN-LIEBOLD and Christine SEIB
Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 63–73, D-1000 Berlin 33 (Dahlem), Germany

Received 25 April 1979

1. Introduction

Protein L20 is a very basic protein of the large subunit of the *E. coli* ribosome [1]. As shown by immune electron microscopy, the protein has two antibody binding sites which are at different sites on the 50 S ribosomal subunit surface, indicating that in situ protein L20 has an elongated shape. Site A is located in the 'seat' area of the 50 S subunit close to the corner of the particle in the vicinity of binding sites L18 B, L11 B, and L3; whereas the second site, L20B, is situated at the lower pole of the subunit (reviewed in [2]).

Protein L20 belongs to a group of 12 proteins which bind independently and specifically to the 23 S RNA (reviewed in [3]). By nuclease digestion studies of single protein–RNA complexes and by reconstitution experiments with RNA fragments, L20 was shown to rebind to the 13 S RNA fragment in the 5′-proximal region of the 23 S RNA [4,5].

Protein L20 is one of the five proteins which are essential for the formation of the RI^{*}_{50} [1] particle [6]. This particle is a necessary intermediate of the in vitro assembly process of the 50 S subunit. Recent analysis revealed that L20 is only an assembly protein, i.e., it is involved in neither the late assembly nor the protein synthesizing functions of the 50 S subunit [7].

In this paper the complete primary structure of protein L20 is presented and predictions for its secondary structure are made. Furthermore, a comparison of its primary structure with the sequences of 45 other ribosomal proteins from *E. coli* and two sequences from yeast ribosomal proteins was performed.

2. Materials and methods

Protein L20 was isolated from *E. coli* K12 strain A19 as described in [8] and provided by Dr H. G. Wittmann. The identity and purity of the protein were checked by two-dimensional polyacrylamide gel electrophoresis [9].

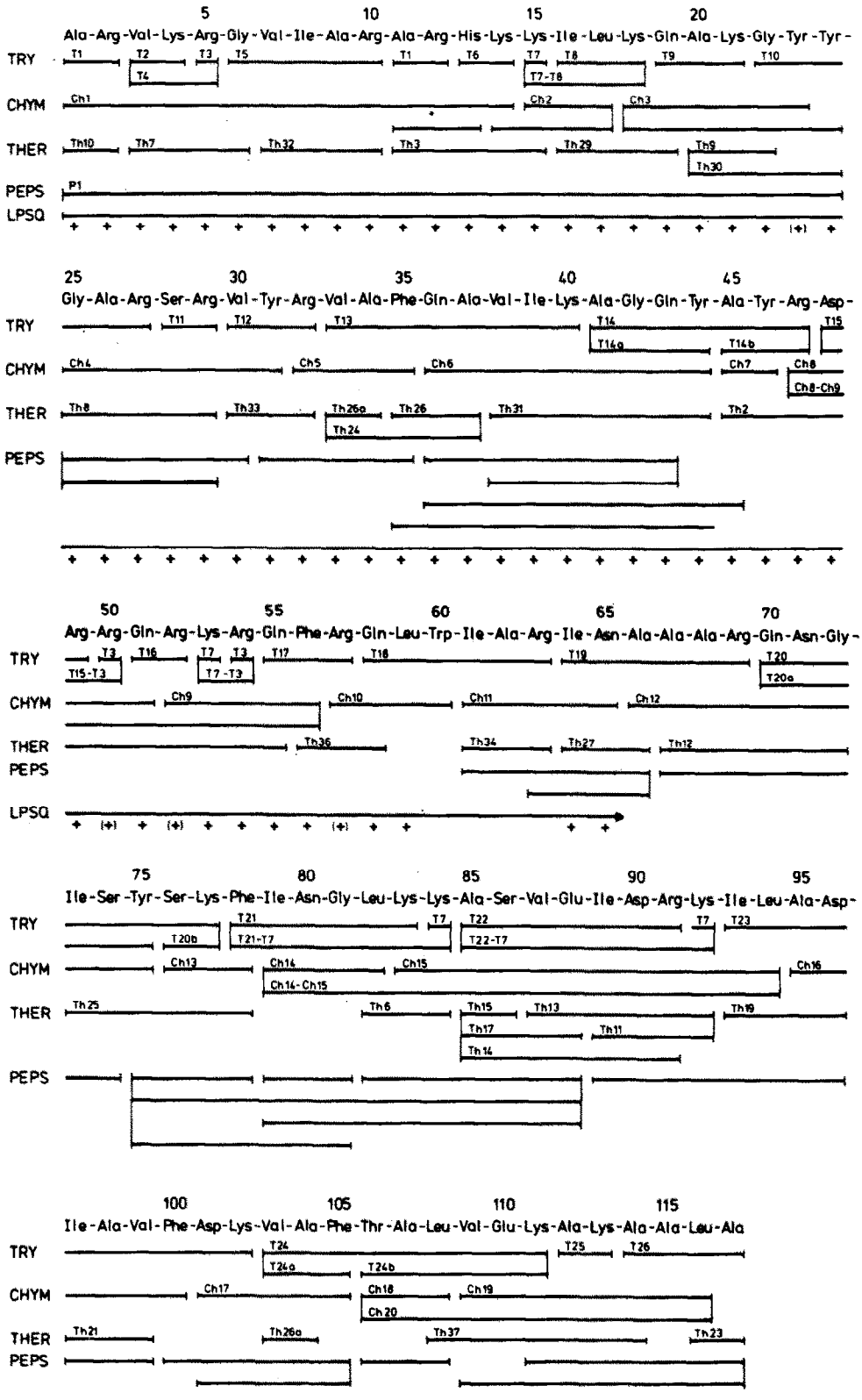
Enzymatic digestions of the protein were performed with trypsin and α-chymotrypsin at pH 8.1 in 0.1–0.2 M *N*-methylmorpholine acetate buffer at 37°C for 2 h and 5 h, respectively. Digestions with thermolysin were made under the same conditions for 2 h at 55°C. Hydrolysis with pepsin was in 0.05 M acetic acid for 2 h at 37°C.

Separation of the numerous peptides was achieved by:

(i) The fingerprint technique using thin-layer cellulose sheets [10];
(ii) Chromatography on Dowex 50 X8 (M71) micro columns [11];
(iii) By gel filtration on Sephadex G-50 (superfine) columns in 10% acetic acid [11].

Amino acid analyses were performed after hydrolysis in 5.7 N HCl (with the addition of 0.02% 2-mercaptoethanol) with the aid of a Durrum D-500 analyzer as described in [10]. Tryptophan was detected by spraying Ehrlich reagent [12] on fingerprints made after digestion of L20 with trypsin, thermolysin and chymotrypsin. Tryptophan was determined quantitatively by hydrolysis with 4 N methane sulfonic acid [13].

Sequencing of the various peptides was carried out using the combined micro-dansyl-Edman degradation


5 10 15 20
Ala-Arg-Val-Lys-Arg-Gly-Val-Ile-Ala-Arg-Ala-Arg-His-Lys-Lys-Ile-Leu-Lys-Gln-Ala-Lys-Gly-Tyr-Tyr-
TRY
CHYM
THER
PEPS
LPSQ
25 30 35 40 45
Gly-Ala-Arg-Ser-Arg-Val-Tyr-Arg-Val-Ala-Phe-Gln-Ala-Val-Ile-Lys-Ala-Gly-Gln-Tyr-Ala-Tyr-Arg-Asp-
50 55 60 65 70
Arg-Arg-Gln-Arg-Lys-Arg-Gln-Phe-Arg-Gln-Leu-Trp-Ile-Ala-Arg-Ile-Asn-Ala-Ala-Ala-Arg-Gln-Asn-Gly-
75 80 85 90 95
Ile-Ser-Tyr-Ser-Lys-Phe-Ile-Asn-Gly-Leu-Lys-Lys-Ala-Ser-Val-Glu-Ile-Asp-Arg-Lys-Ile-Leu-Ala-Asp-
100 105 110 115
Ile-Ala-Val-Phe-Asp-Lys-Val-Ala-Phe-Thr-Ala-Leu-Val-Glu-Lys-Ala-Lys-Ala-Ala-Leu-Ala

with 20–50 nmol/sample [14]. Some peptides were sequenced by using the solid-phase method [15] with 30–50 nmol of sample. In this case the attachment of the C-terminal carboxyl groups to amino-polystyrene resin was carried out as described in [16]. More recently, the new DABITC–PITC double coupling method [17] was applied with 3–10 nmol of peptide material. The micro-methods for the sequencing studies are summarized in [18]. Details of the experimental procedures used for the sequence analysis of protein L20 will be presented elsewhere.

3. Results and discussion

3.1. *Sequence determination*

The primary structure of protein L20 was determined by automatic degradation of the intact protein [19] and from sequencing peptides obtained by digestions with trypsin, α-chymotrypsin, thermolysin and pepsin. Peptides which could be isolated in a purified form by a combination of column chromatography (on either Dowex 50 micro colums or Sephadex) and standard thin-layer techniques (summarized in [18]) are listed in fig.1. Sequencing of these peptides was performed by the micro-dansyl-Edman technique [14], by solid-phase degradation [15,16] and by the more sensitive combined DABITC–PITC method [17].

3.2. *Alignment of isolated peptides*

Cleavage with *Staphylococcus aureus* protease or treatment with either cyanogen bromide or 2-(2-nitrophenylsulphenyl)-3-methyl-3′-bromoindolenine (BNPS-skatole) or *N*-bromosuccinimide gave incomplete hydrolysis of peptide bonds, resulting in a variety of peptides of similar size which could not be isolated in purified form on a micro-scale. Some of the fragments obtained were heterogeneous at their N- or C-termini and therefore not suitable for sequence or alignment studies.

The order of the peptides in the N-terminal half of the protein (up to position 65) was derived from automatic liquid-phase Edman degradation of the intact protein in an improved Beckman sequencer [20] with an automatic conversion device [21]. The sequence of the 50 N-terminal residues of L20 which had been published before isolation of the peptides [19], was found to be exactly the same as the sequence resulting from the recent peptide studies, as shown in fig.1.

In the C-terminal half of the molecule the chymotryptic peptides Ch11 (pos. 61–65), Ch12 (66–75), Ch13 (76–78), the peptic peptides (pos. 61–66, 67–74, 75–78, 75–81) and the thermolytic peptides Th12 (67–72), Th25 (73–78) served as 'bridges' for aligning the tryptic peptides T18–T19–T20–T21. The alignment of the tryptic peptides T21–T7–T22–T7–T23 was cross-checked from the sequence of the peptides Ch15 (83–94), Ch14–Ch15 (79–94) and the peptic peptides (pos. 82–88, 75–88, 79–88 and 89–96). Finally, the alignment of T23–T24–T25–T26 was obtained from the peptides Ch17 (101–105), Ch19 (109–116), Ch20 (106–116) and from the pepsin fragments (pos. 100–105, 101–105, 109–117 and 111–117). The C-terminal sequence Lys–Ala–Lys–Ala–Ala–Leu–Ala was confirmed by sequencing a peptide isolated after digestion with *Staphylococcus aureus* protease.

The C-terminal half of the protein chain (after position 61) is completely covered by peptic peptides which have all been sequenced. The whole protein chain of protein L20 is spanned by tryptic peptides and by fragments resulting from chymotryptic hydrolysis. From position 33 onwards all tryptic peptides were completely sequenced and partial sequence determinations were performed on several of the chymotryptic peptides.

The main problems in sequencing this ribosomal protein were caused by:

(i) The high content of alanine (18.8%) and basic amino acids (26.5%);

Fig.1. The primary structure of protein L20 from *Escherichia coli* ribosomes. *Abbreviations:* TRY, trypsin digestion (peptides T); CHYM, digestion with α-chymotrypsin (peptides Ch); THER, digestion with thermolysin (peptides Th); PEPS, digestion with pepsin; LPSQ, degradation of intact protein in an improved Beckman sequencer with an automatic conversion device; + denotes clearly identified by thin-layer technique and by mass spectrometry; (+) indicates where the observed spots on thin-layer sheets were weak.

(ii) The unspecific cleavages at tyrosine residues during tryptic digestion;
(iii) The high number of tryptic peptides with N-terminal glutamine and the low recovery of peptide T18 (58–63);
(iv) The very basic region, positions 49–57;
(v) Low background contamination of proline and methionine in amino acid analyses of the intact protein [22]; however both amino acids not being present in the partial peptides listed in fig.1.

3.3. *The primary structure of protein L20*

Combination of all results discussed above led to a sequence of protein L20 of *E. coli* as shown in fig.1. According to this primary structure, protein L20 consists of 117 amino acid residues with the following composition: Asp 4, Asn 3, Thr 1, Ser 4, Glu 2, Gln 7, Pro 0, Gly 6, Ala 22, Val 9, Met 0, Ile 10, Leu 6, Tyr 6, Phe 5, His 1, Lys 14, Arg 16, Trp 1, Cys 0. The molecular weight calculated on the basis of the sequence is 13 366. The N-terminal region of this protein up to position 40 consists of only basic and neutral amino acids, whereas 5 out of the 6 acidic residues are found in the C-terminal part of the chain. Basic amino acids are clustered in the following regions: positions 2–5, 10–18, 27–32 and 49–57. The 12 aromatic amino acids are located almost pairwise along the chain.

3.4. *Secondary structure of protein L20*

Secondary structure predictions are presented for protein L20 in fig.2. They were made according to four predictive methods as detailed in [23]. Regions of helix, extended structure and β-turns are indicated in the line 'PRE' of fig.2 provided that at least three out of four methods resulted in the same prediction. Based on this average, protein L20 contains at least 49% helix, 2% extended structure and 19% β-turns. Strong helix-regions are predicted for positions 9–19, 56–68 and 84–89/116. Three pronounced turn areas are calculated for positions 21–28, 46–50 and 71–76. Extended structure might occur in the region about positions 29–32/39.

3.5. *Comparison with other ribosomal proteins*

The comparison of the L20 amino acid sequence with other known ribosomal protein sequences demonstrates the unique character of protein L20. It does not share many identical or homologous tetrapeptides with the others. The pentapeptide Ala–Val–Ile–Lys–Ala (pos. 37–41) and the sequence Leu–Lys–Lys–Ala (pos. 82–85) of L20 occur also in the *E. coli* ribo-

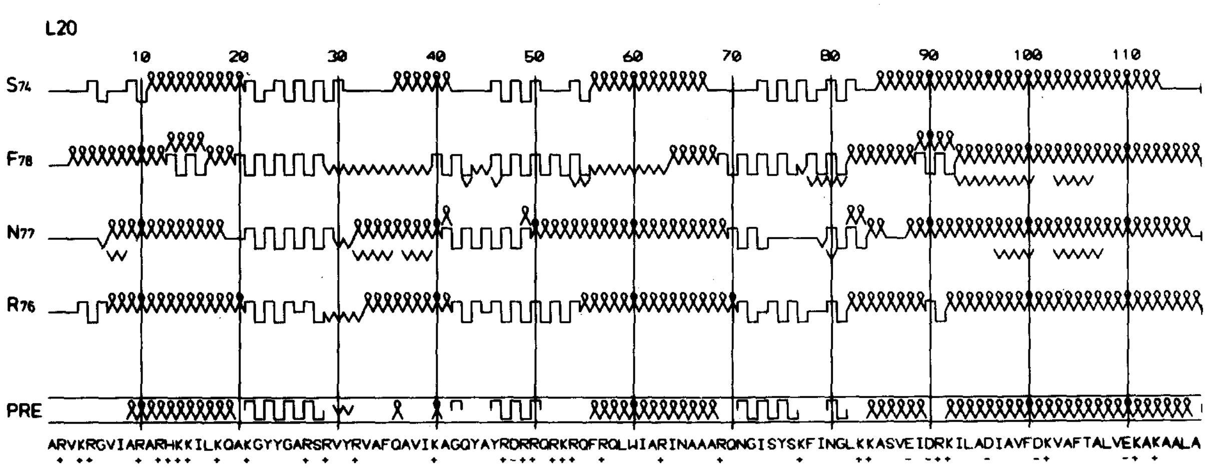


Fig.2. Secondary structure prediction of ribosomal protein L20 according to four methods (detailed in [23]). The symbols represent residues in a helical region (ꝏ), β-turns or loop (⊓⊔), extended structure (W) and random coil conformation (———), respectively. The line 'PRE' summarizes the secondary structure obtained when three or all of the four predictions are in agreement. The prediction methods are according to Burgess, Ponnuswamy and Scheraga (1974) (S); to Chou and Fasman (1977, 1978) (F); to Nagano (1977) (N); and to Robson and Suzuki (1976) (R) as detailed in [23].

somal protein L12 (pos. 67–71 and pos. 106–109). With *E. coli* proteins S4 and L1 protein L20 shares three and four identical tetrapeptides, respectively, and several homologous tetrasequences. Ribosomal protein Y-44 from yeast [24] has the sequence Glu–Lys–Lys–Gln–Lys (pos. 93–97) which is homologous to the sequence Asp–Arg–Arg–Gln–Arg of protein L20 (pos. 48–52).

References

[1] Kaltschmidt, E., Dzionara, M. and Wittmann, H. G. (1970) Mol. Gen. Genet. 109, 292–297.

[2] Stöffler, G. and Wittmann, H. G. (1977) in: Molecular Mechanism of Protein Biosynthesis, ed. (Weissbach, H., Pestka, S. eds) pp. 117–202, Academic Press, New York.

[3] Brimacombe, R., Stöffler, G. and Wittmann, H. G. (1978) Ann. Rev. Biochem. 47, 217–249.

[4] Spierer, P., Zimmermann, R. A. and Mackie, G. A. (1975) Eur. J. Biochem. 52, 459–468.

[5] Chen-Schmeisser, U. and Garrett, R. A. (1976) Eur. J. Biochem. 69, 401–410.

[6] Spillmann, S., Dohme, F. and Nierhaus, K. H. (1977) J. Mol. Biol. 115, 513–523.

[7] Nowotny, V., Schulze, H. and Nierhaus, K. H. (1979) Hoppe-Seyler's Z. Physiol. Chemie 360, 335–336.

[8] Hindennach, J., Kaltschmidt, E. and Wittmann, H. G. (1971) Biochemistry 23, 12–16.

[9] Kaltschmidt, E. and Wittmann, H. G. (1970) Analyt. Biochem. 36, 401–412.

[10] Hitz, H., Schäfer, D. and Wittmann-Liebold, B. (1977) Eur. J. Biochem. 75, 497–512.

[11] Heiland, J., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.

[12] Winkler, S. (1934) Hoppe-Seyler's Z. Physiol. Chem. 228, 50.

[13] Liu, T. Y. and Chang, Y. H. (1971) J. Biol. Chem. 246, 2842–2848.

[14] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.

[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 (Laursen, R. A. ed) pp. 83–90, Pierce Chem. Co., Rockford.

[17] Chang, J. Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.

[18] Wittmann-Liebold, B., Brauer, D. and Dognin, J. M. (1977) in: Solid Phase Methods in Protein Sequence Analysis, (Previero, A. and Coletti-Previero, M. A. eds) Elsevier/North-Holland, Amsterdam, New York.

[19] Wittmann-Liebold, B., Geissler, A. W. and Marzinzig, E. (1975) J. Supramol. Struct. 3, 426–447.

[20] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.

[21] Wittmann-Liebold, B., Graffunder, H. and Kohls, H. (1976) Analyt. Biochem. 75, 621–633.

[22] Kaltschmidt, E., Dzionara, M. and Wittmann, H. G. (1970) Molec. Gen. Genet. 109, 292–297.

[23] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.

[24] Itoh, T. and Wittmann-Liebold, B. (1978) FEBS Lett. 96, 399–402.