The primary structure of protein L2 from the *Escherichia coli* ribosome

Makoto Kimura, Liane Mende and Brigitte Wittmann-Liebold

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33 (Dahlem), Germany

Received 14 October 1982

The complete primary structure of protein L2 which is the largest protein component of the *E. coli* 50S subunit, has been established. A combination of enzymatic and chemical cleavages has been employed to isolate peptides, which were sequenced by the micro-DABITC/PITC double-coupling method [FEBS Lett. (1978) 93, 205–214]. The sequence determined shows protein L2 to consist of 272 amino acid residues with $M_r = 29730$. Secondary structure predictions were made based on the primary structure. Further, sequence regions homologous to other ribosomal proteins are presented. These results suggest protein L2, which binds specifically to the 23S RNA, to show homologous sequence stretches to the other RNA-binding proteins.

E. coli *ribosome* *Protein L2* *Primary structure determination* *Secondary structure prediction* *Homology to other ribosomal proteins*

1. INTRODUCTION

Protein L2 is the largest protein of the *Escherichia coli* 50S ribosomal subunit. Its M_r is 28000–32000 according to sedimentation equilibrium and SDS gel electrophoresis [1]. An open and unfolded structure has been postulated for protein L2 since it is highly susceptible to proteolytic attack under mild conditions and in PMR studies it gave no indication for a compact folding of the protein chain [2]. It binds specifically to the 23S RNA [3–5], and its binding site is within the middle region of the RNA chain [6,7].

Protein L2 can be cross-linked in situ to proteins L5, L6, L7/L12, L9, L10, L11 and L17 [8], and it has been found together with L15 or L17 in protein complexes [9]. Protein L2 is among the few proteins which are involved in the binding of 5S RNA to the 23S RNA [10]. It is important for the peptidyltransferase activity of the 50S subunit as shown by affinity labeling and reconstitution studies [11].

Here, the complete primary structure of protein L2 is reported. It consists of 272 amino acid residues and has $M_r = 29730$. Based on the amino sequence, the secondary structure of the protein is predicted. Furthermore, the primary structure of protein L2 is compared to that of all other ribosomal proteins whose amino acid sequences are known.

2. MATERIALS AND METHODS

2.1. *Protein isolation*

Protein L2 was isolated from 50S subunits of *E. coli* K12 (strain A19) ribosomes in the presence of urea as in [12] and was kindly provided by Dr H.G. Wittmann. Another sample was purified without urea as in [13] and was a gift of Dr J. Dijk.

2.2. *Limited trypsin digestion*

Cleavage of the protein after blocking the lysines with ETPA, exo-*cis*-3,6-endoxo-∇^4-tetrahydrophthalic acid anhydride [14], synthesized according to [15], was performed as in [16]: 8mg protein were treated in 4ml 0.2M borate buffer at pH 8.5 for 1h with 10mg reagent at 0°C, then dialysed against dilute ammonia at pH 8.0 and digested with

Published by Elsevier Biomedical Press
001457593/82/0000—0000/$2.75

TPCK-trypsin (from Worthington, Freehold NJ) for 4 h at 37°C (enzyme/substrate, 1:50). Deblocking was done in 30% acetic acid overnight at room temperature, and the resulting peptides (LB-peptides) were separated by Sephadex G-50 s.f. (150 × 1 cm column) in 10% acetic acid, followed by thin-layer fingerprinting as in [17,18]. Cleavage of the protein at the lysines after modification of the arginines with 1,2-cyclohexanedione (from Pierce Chemical Co., Rockford IL) was as in [19]. 6 mg protein were treated with 12 mg reagent in 2 ml 0.2 M NaOH for 3.5 h at room temperature, then neutralized and digested with trypsin (as above). The resulting peptides (CHD-peptides) were separated as above.

2.3. *Digestion with* Staphylococcus aureus *protease [20]*

Digestion of 10 mg protein was with 250 μg enzyme (from Miles Biochemicals, Elkhart IL) in 0.05% ammonium acetate at pH 3.9 for 48 h at 37°C. The enzyme was added in two portions. Separation of the resulting SP-peptides was as above. Some peptide fractions containing larger fragments were further digested with trypsin.

2.4. *Digestion with* Armillaria mellea *protease [21]*

2 mg protein L2 were digested with 4 μg enzyme (obtained from Dr V. Barkholt Pedersen, Copenhagen) in 400 μl 0.2 M *N*-methylmorpholine acetate buffer (pH 8.1) for 6 h at 37°C. The resulting AP-peptides were separated on Sephadex G-50 s.f. (140 × 1 cm column) in 10% acetic acid, followed by thin-layer fingerprinting.

2.5. *Digestion with trypsin*

Samples (1 mg each) of untreated and of oxidized protein were digested with 20 μg TPCK-trypsin in 100 μl buffer (pH 8.1, as above) for 4 h at 37°C, and the digest was separated by thin-layer fingerprinting.

2.6. *Cleavage with CNBr [22]*

3 mg protein were treated with 10 mg CNBr in 300 μl 70% formic acid for 20 h at room temperature in the dark. The peptides were separated on Sephadex G-75 s.f. (210 × 1 cm column) in 10% acetic acid. Peptide CBI was further digested with trypsin (with 10 μg in 400 μl pH 8.1 buffer) and with *Armillaria mellea* protease (with 2 μg in 400 μl of same buffer) at 37°C for 3 h and 6 h, respectively. Peptides CBII, CBIV and CBV were cleaved with thermolysin at 50°C (with 2 μg in 200 μl of same buffer) for 12 h; CBVI was digested with pepsin (with 4 μg in 300 μl of 5% formic acid) for 6 h at 37°C. All resulting peptides were purified by thin-layer fingerprinting.

2.7. *Cleavage with BNPS-skatole [23]*

2 mg protein were reacted with 3 mg reagent, 2-(2-nitrophenyl-sulfenyl)-3-methyl-3′-bromoindolenine (from Pierce Chemical Co., Rockford IL) in 200 μl 60% acetic acid for 28 h at room temperature. The BNPS-peptides were separated as described for CNBr-peptides.

2.8. *Sequence analysis of peptides and the intact protein*

Liquid phase sequencing of the intact protein was made with a modified Beckman sequencer as detailed [24,25]. The released PTH-amino acid derivatives were identified on silica thin-layer sheets in three successive solvent mixtures and further by mass spectrometry as in [26]. All peptide fragments were sequenced manually by the DABITC/PITC (4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate) double coupling method [27] as given in [28].

Amino acid analyses were performed in Durrum D-500 analyzers in the sensitivity range of 1–2.5 nmol/amino acid. Details of the experimental procedures are reviewed in [28].

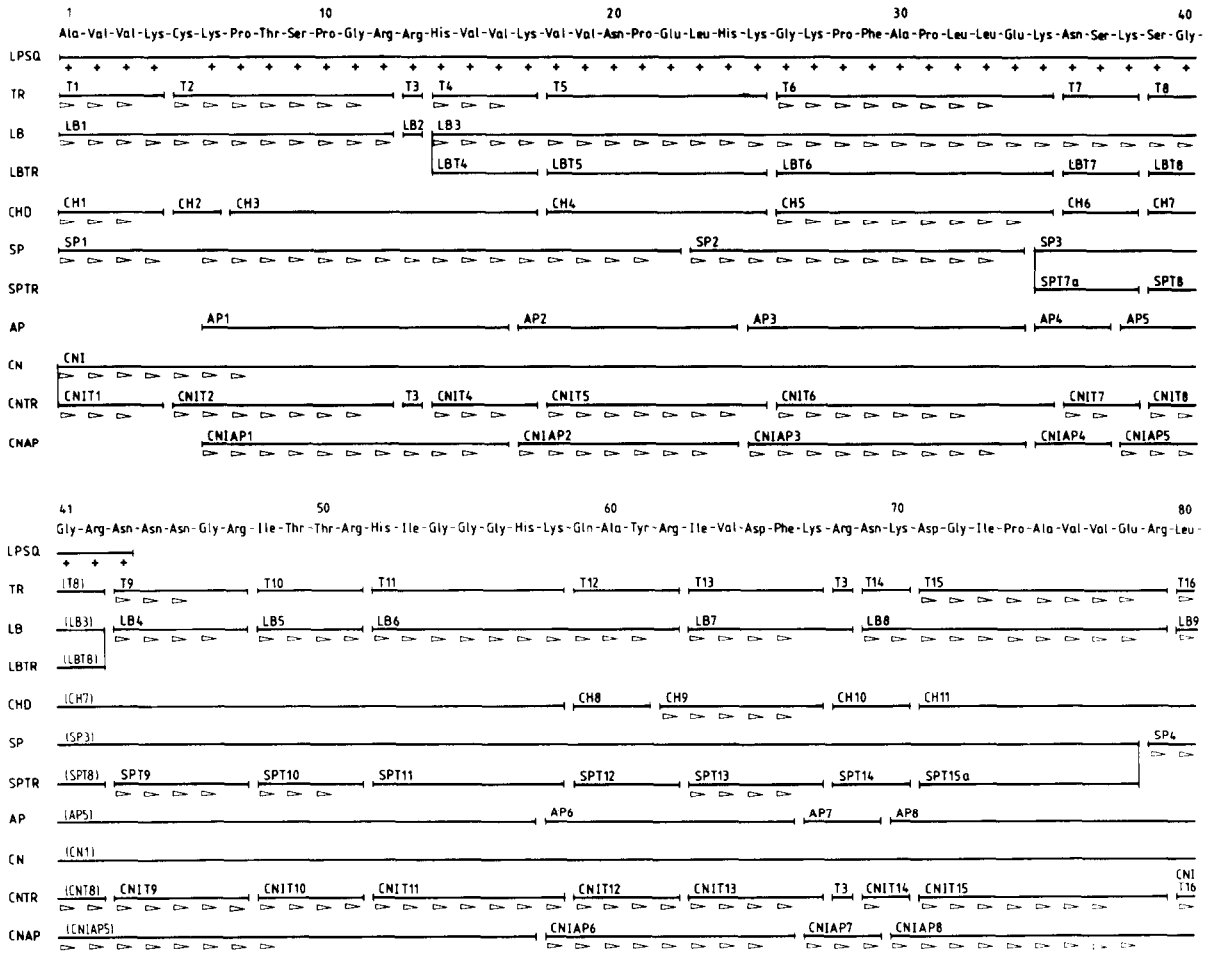
3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

The N-terminal sequence of protein L2 up to position 43 was derived from liquid phase sequencing of the intact protein in a modified Beckman sequencer as reported [26]. The other sequences came from sequence analysis of several sets of fragments obtained after enzymatic digestions of the protein with trypsin, *Staphylococcus aureus* and *Armillaria mellea* protease and pepsin. In order to facilitate the alignment, larger sized peptides were obtained by treatment with cyanogen bromide.BNPS-skatole,2-(2-nitrophenyl-sulfenyl)-3-methyl-3-bromoindolenine, and by limited tryptic digestion after blocking of the ε-amino groups of the lysines with ETPA, exo-*cis*-3,6-endoxo-∇^4-

tetrahydropnthalic acid anhydride or vice versa after reaction of the arginines with 1,2-cyclohexanedione. All peptides were isolated by a combination of gel filtration on Sephadex columns with thin-layer methods or by direct thin-layer fingerprinting. Their sequences were obtained manually with the sensitive DABITC/PITC, 4-(*N,N*-dimethylamino)-azobenzene-4′-isothiocyanate/phenylisothiocyanate, double-coupling method.

The combination of the sequence information obtained from the various cleavages allowed us to establish the complete primary structure of protein L2 unambiguously as presented in fig. 1. The nucleotide sequence of the gene coding for protein L2 is not available so far. Therefore, no comparison with such data can be made.



(continued)

Fig. 1. Primary structure of protein L2. Sequence data on individual peptides are indicated as follows: LPSQ, sequenced by an improved Beckman sequencer; (∓) positively identified; (▷) sequenced with the DABITC/PITC double coupling method; TR, SP, AP, PE indicate peptides derived from cleavage with trypsin, *S. aureus* protease, *A. mellea* protease and pepsin, respectively. LB and CHD denote cleavages with trypsin after blocking of the lysines with EPTA and reaction of the arginines with cyclohexanedione, respectively. CN means cleavage with cyanogen bromide and LBTR, SPTR, CNTR, CNTH and CNAP subcleavages of the respective fragment with trypsin, thermolysin, and *A. mellea* protease; BNPS, cleavage with BNPS-skatole.

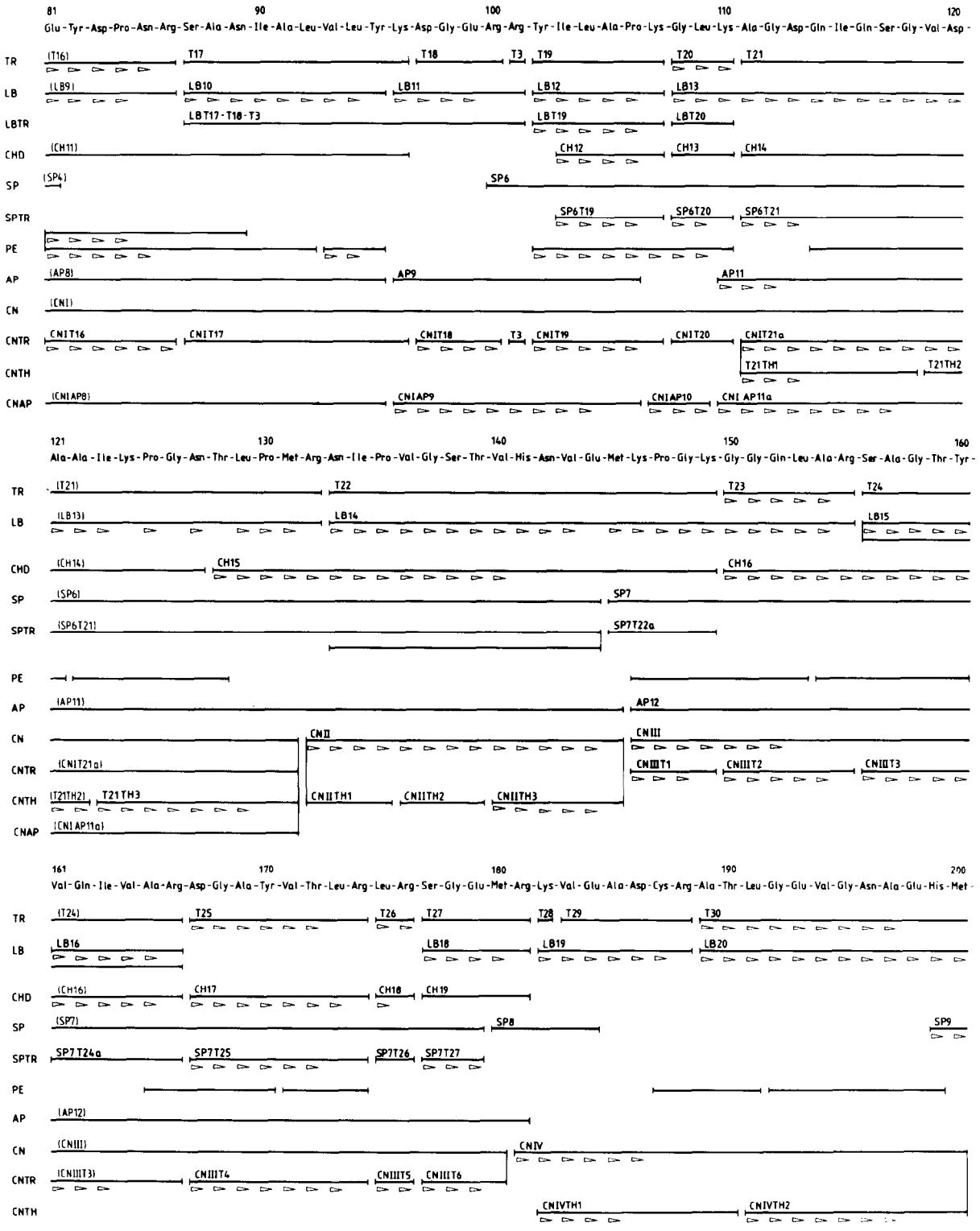


Fig. 1. continued.

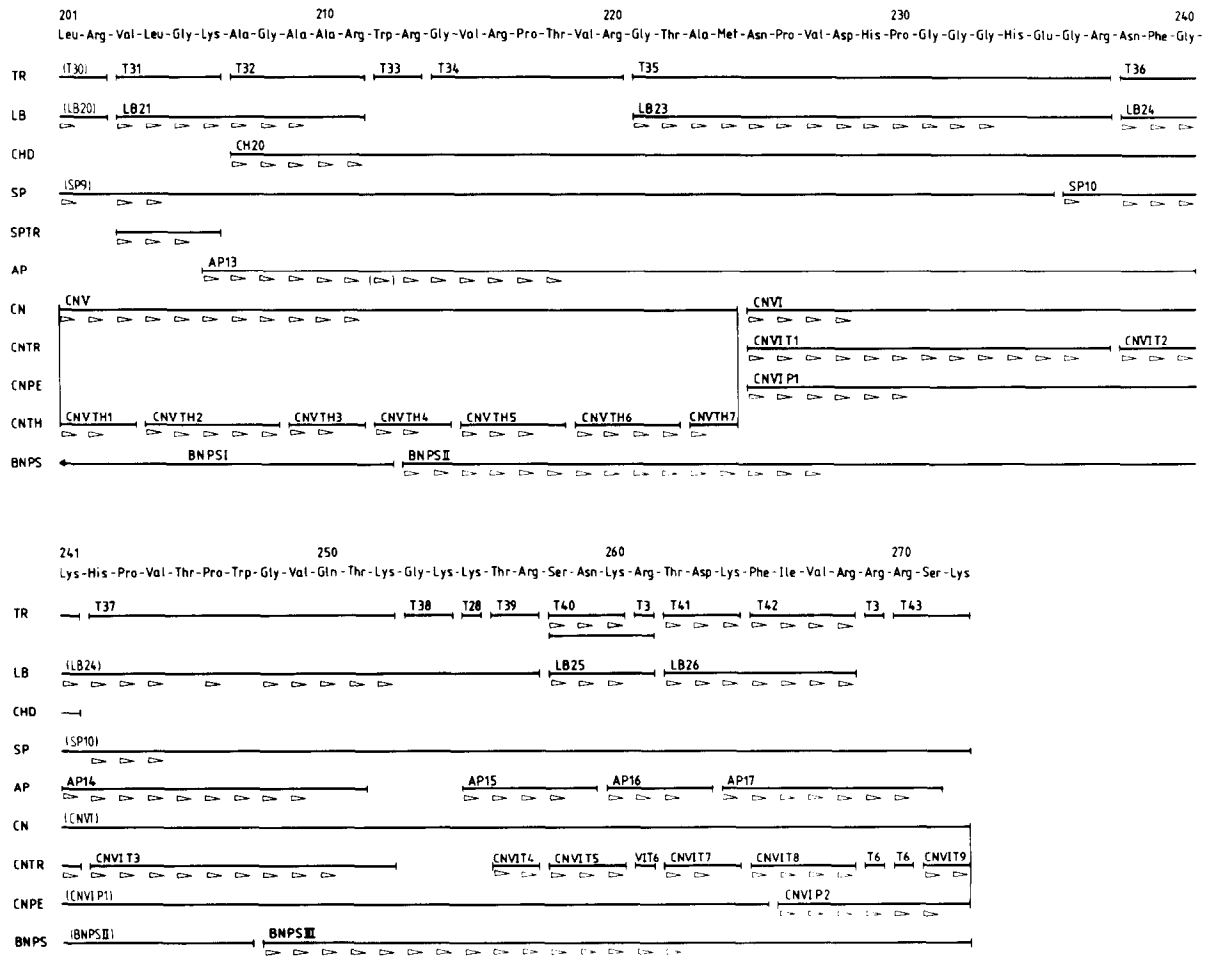


Fig. 1. concluded.

3.2. *Characterization of the protein*

According to the sequence given in fig. 1, protein L2 has 272 amino acid residues. This amounts to $M_r = 29730$. Protein L2 has a content of 23.2% basic residues (Lys, Arg and His) opposed to only 7.7% acidic residues (Asp and Glu). Notable is the high content of glycine (12.5%) and valine (9.6%). Its amino acid composition is as follows:

Asp_{10}, Asn_{15}, Thr_{14}, Ser_{10}, Glu_{11}, Gln_6, Pro_{17}, Gly_{34}, Ala_{21}, Val_{26}, Met_5, Ile_{11}, Leu_{15}, Tyr_6, Phe_4, His_9, Lys_{25}, Arg_{29}, Trp_2, Cys_2.

The distribution of the amino acids within the polypeptide chain is not uniform. Basic residues are clustered in both the N-terminal (pos. 3–70) and C-terminal (pos. 199–272) regions which also contain 8 of the histidines accumulated in 3 sequence areas. Each of the two histidines are interspaced by the sequence Ile–Gly–Gly–Gly– and Pro–Gly–Gly–Gly–, respectively. The C-terminal region contains both tryptophans and 2 of the 4 phenylalanines. The N-terminal region contains one of the 2 cysteines and the remaining 2 phenylalanines. The central part of protein L2 (pos. 60–224) contains all methionines and all tyrosines. Based on this distribution of seldomly occurring residues and on some repetitions in sequence stretches, internal repeats within the sequence of protein L2 can be given, as presented in fig. 2. Internal repetitions have also been observed for other ribosomal proteins ([29]; in preparation).

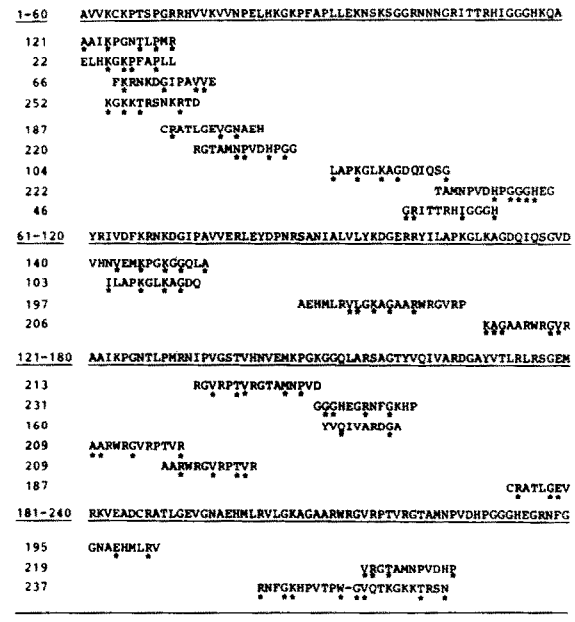


Fig. 2. Internal repeats in the sequence of protein L2.

3.3. *Secondary structure predictions of protein L2*

According to the algorithms in [30–34], secondary structural elements in protein L2 have been calculated on the basis of the amino acid sequence, as detailed in [35,36]. The predictions give a higher content of β-sheet structure in comparison to most of the other ribosomal proteins. These β-sheets are interrupted by strongly predicted turns (fig. 3).

Table 1

Percentage values of predicted conformations in protein L2

Prediction according to:	α-Helix (%)	β-Sheet (%)	Turn (%)	Unpredicted (%)
[30]	7	11	37	45
[31,32][a]	9–15	27–34	41	16
[33]	18–22	14–27	49	18
[34]	30	19	40	11

[a] Computed by a programme according to [36]

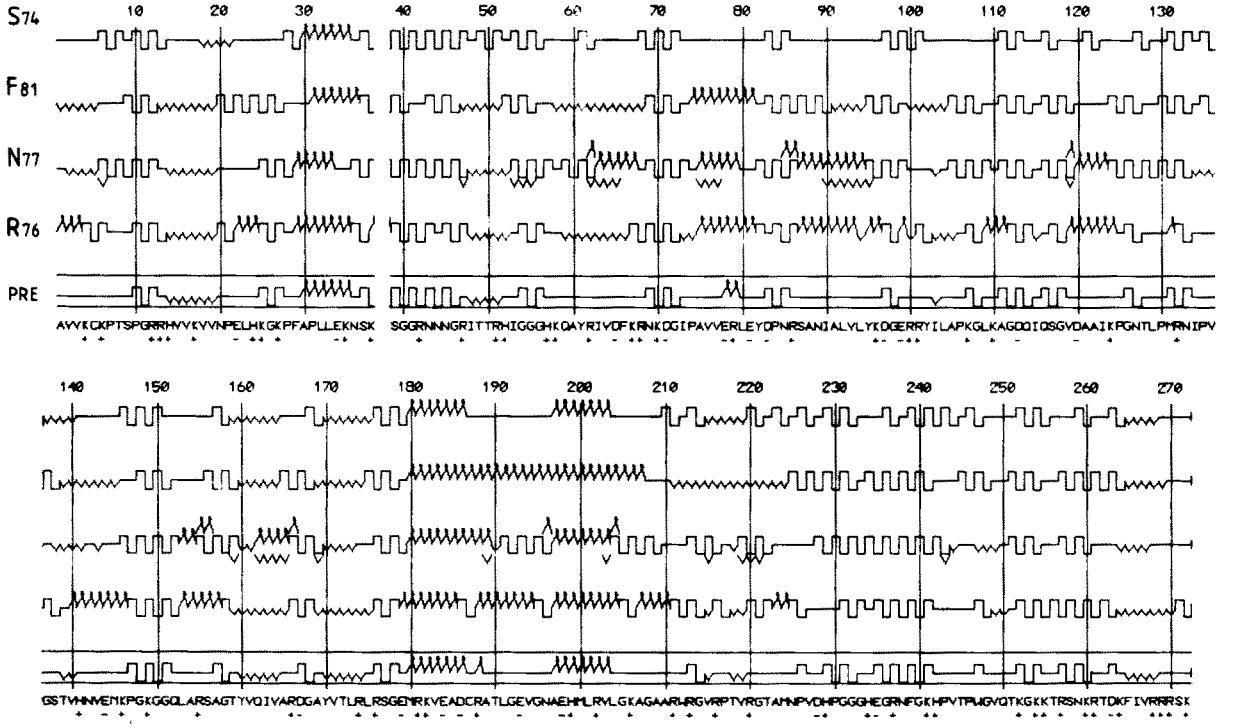


Fig. 3. Secondary structure predictions of protein L2. The calculations were made according to: S, [30]; F, [31,32]; N, [33]; R, [34]. For the predictions according to Chou and Fasman a computerized version has been employed [36]: (ΛΛΛ) α-helix; (VVV) β-sheet; (⊓⊔⊓) β-turn structure; (——) random coil. In the line PRE, three of four calculations gave predictions in agreement.

Table 2

Homologous sequence regions in protein L2 and other ribosomal proteins of *E coli*

Protein	Positions	Sequence
L2	39- 85	SGGRNNNGRITTRHIGGGHKQAYRIVDFKRNKDGIPAVVERLEYDPN
S12	52- 70	CRVRLTNGFEVTSYIGGEG
S5	101-117	GTGIIAGGAMRAVLEVA
S8	34- 60	AIANVLKEEGFIEDFKVEGDTKPELEL
S2	198-217	VIPGNDDAIRAVTLYLGAVA
S3	110-126	LDAKLVADSITSQLERR
S3	201-218	IFKGEILGGMAAVEQPEK
S21	20- 39	RSCEKAGVLAEVRRREFYEK
L2	85-111	NRSANIALVLYKDGERRYI-LAPKGLKA
S1	325-348	GDVVEVMVLDIDEERRRISLGLKQCKA
S1	411-430	KGDEIAAVVLQVDAERERIS
L2	111-147	AGDQIQSGVDAAIKPGNTLPMRNIPVGSTVHNVEMKP
S9	8- 30	TGRRKSSAARVPIKPGNGKIVIN
S5	107-126	GGAMRAVLEVAGVHNVLAKA
S19	37- 54	TIFPDRMIGLTIAVHNGR
L2	148-166	GKGGQLARSAGTYVQIVAR
S10	85-103	DALMRLDLAAGVDVQISLG
L2	163-186	IVARDGAYVTLRLRSGEMRKVEAD
S7	39- 60	ESIVYSALETLAQRSGKSELEA
S9	100-118	AGFVTRDARQVERKKVGLR
L2	199-224	HMLRVLGKAGAARWRGVRPTVRGTAM
S7	95-120	NALAMRWIVEAARKRGDKSMALRLAN
S3	44- 61	KELAKASVSRIVIERPAK
L2	229-248	HPGGGHEGRNFGKHPVTPWG
S12	63- 82	TSYIGGEGHNLQEHSVILIR
L2	247-268	WGVQTKGKKTRSNKRTDKFIVR
S13	92-113	RGLPVRGQRTKTNARTRKGPRK
L2	12- 31	RRHVVKVVNPELHKGKPFAP
L3	103-119	DVKKVDVTGTSKGKGFA
L23	51- 70	FEVEVEVVNTLVVKGKVKRH
L2	35- 61	KNSKSGGRNNNGRITTRHIGGGHKQAY
L9	110-128	EVAKSEVRLPNGVLRTHGE
L9	91-107	FGSIGTRDIADAVTAAD
L2	73-101	IPAVVERLEYDPNRSANIALVLYKDGERR
L18	55- 75	EKAIAEQLKYTGNKDAAAAVG
L23	74- 92	IGRRSDKKAYVTLKEGQNL
L2	91-127	ALVLYKDGERRYILAPKGLKAGDQIQSGVDAAIKPGN
L4	32- 48	VVAYAAGARQGTRAQKT

Protein	Positions	Sequence
L6	3- 20	VAKAPVVVPAGVDVKING
L9	17- 37	DQVNVKAGYARNFLVPQGKAV
L11	73- 90	PPAAVLLKKAAGIKSGSG
L14	35- 51	VGDIIKITIKEAIPRGK
L2	128-145	TLPNRNIPVGSTVHNVEM
L1	133-150	RGLMPNPKVGTVTPNVAE
L2	141-187	HNVEMKPGKGGQLARSAGTYVQIVARDGAYVTLRLRSGEMRKVEADC
L6	20- 37	GQVITIKGKNGELTRTLN
L4	62- 78	QKGTGRARSGSIKSPIW
L4	66- 83	GRARSGSIKSPIWRSGGV
L9	80- 99	IASKAGDEGKLFGSIGTRDI
L17	82- 99	ELGPRFASRAGGYTRILK
L22	6- 24	KHRHARSSAQKVRLVADLI
L3	176-192	DVVRVDAERNLLLVKGA
L14	93-118	QPIGTRIFGPVTRELRSEKFMKIISL
L28	41- 61	SEKRFVTLRVSAKGMRVIDKK
L29	42- 58	LLKQVRRDVARVKTLLN
L34	27- 44	GRQVLARRRAKGRARLTV
IF-3	1- 17	MKGGKRVQTARPNRING
IF-3	118-144	RFLEEGDKAKITLRFRGREMAHQQIGM
IF-1	26- 43	ENGHVVTAHISGKMRKNY
L2	179-222	EMRKVEADCRATLGEVGNAEHMLRVLGKAGAARWRGVRPTVRGT
L16	42- 59	TARQIEAARRAMTRAVKR
L9	51- 67	RAELEAKLAEVLAAANA
L9	21- 41	VKAGYARNFLVPQGKAVPATK
L12	58- 75	LKAAGANKVAVIKAVRGA
L12	100-110	KDDAEALKKALEEAGAEVE
L13	23- 40	KTLGRLATELARRLRGKH
L14	3- 22	QEQTMLNVADNSGARRVMCI
L15	120-136	VTVRGLRVTKGARAAIE
L1	186-207	ENLEALLVALKKAKPTQAKGVY
L18	96-117	GFQYHGRVQALADAAREAGLQF
L2	243-272	PVTRWGVQTKGKKTRSNKRTDKFIVRRRSK
L16	51- 69	RAMTRAVKRQGKIWIRVFP
L4	50- 67	AEVTGSGKKPWRQKGTGR
L15	27- 45	LGKTGGRGHKGQKSRSGGG
L13	62- 80	VAVTGNKRTDKVYYHHTGH
L22	88-105	RAKGRADRILKRTSHITV
L32	2- 18	VQQNKPTRSKRGMRRSH

Searching of homologous sequence stretches was made by the aid of a computer program under conditions explained in the text: (*) identical residues; (+) amino acids whose codons differ by one nucleotide; (−) amino acids whose codons differ by two nucleotides

Helices were calculated for pos. 180–203/207 and short ones at pos. 30–36 and 74–81. However, the Chou and Fasman predictions show that these regions also have similar probabilities for the occurrence of β-sheet. Thus, this protein may mainly be constructed by β-sheet and turn structural elements. The average values for the calculated secondary structures in protein L2 are listed in table 1.

3.4. *Sequence comparison with other ribosomal proteins*

A comparison of the sequence of protein L2 with the sequences of all other ribosomal proteins from *E. coli* and of about 20 ribosomal proteins deriving from other organisms has been performed. Strong homologies which span the complete chains of these proteins were not observed. However, shorter stretches of homologous structures were found, especially for protein L2 and L1 (see table 2). This table also lists some other homologies, found by a newly developed computer assisted program. The search was performed for protein regions of different size (10–35 residues long) under the following conditions:

(i) At least 30–50% of the residues in the regions had to be identical amino acids in the proteins compared;

(ii) Of the remaining non-identical amino acids at least 70% had to be those which differ by only one or two nucleotides within the corresponding amino acid codons.

When the comparison was made under these conditions, almost no homologous regions were observed to correspond with the N-terminal portion of protein L2. On the other hand, similarities were found in the regions between pos. 40–120, 130–180 and 190–225. Proteins deriving from both, the small and the large subunit, contribute to these homologies. It is remarkable that protein L2 has a region homologous to the repeated structure in protein S1 which is located within the RNA-binding domain [29]. Homologies were also found with repeats in the structure of other proteins, such as S7, S9, and L9 (not shown in table 2). These will be presented elsewhere. Among the L-proteins having homologous stretches to protein L2 are predominantly proteins which bind to the 23 S RNA. As the same function is also assigned to protein L2 [3–5] the similarities in the sequence stretches may reflect a common functional feature of these proteins. This will be discussed in detail elsewhere.

ACKNOWLEDGEMENTS

We thank Dr V. Barkholt Pedersen, Copenhagen for a gift of *Armillaria mellea* protease and Drs H.G. Wittmann and J. Dijk for providing protein L2. The technical assistance of Ms Mabel Pereira-Montero and the help in designing the computer programs of Dr M. Dzionara and K. Ashman is gratefully acknowledged.

REFERENCES

[1] Wittmann, H.G. (1974) in: Ribosomes: (Nomura, M. et al. eds) pp. 93–114, Cold Spring Harbor Laboratory Press, Long Island NY.

[2] Littlechild, J., Malcolm, A.L., Paterakis, K., Ackermann, J. and Dijk, J. (1982) J. Biol. Chem. in press.

[3] Stöffler, G., Daya, L., Rak, K.H. and Garrett, R.A. (1971) Mol. Gen. Genet. 114, 125–133.

[4] Garrett, R.A., Müller, S., Spierer, P. and Zimmermann, R.A. (1974) J. Mol. Biol. 88, 553–557.

[5] Marquardt, O., Roth, H.E., Wystup, G. and Nierhaus, K.H. (1979) Nucleic Acids Res. 6, 3641–3650.

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

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

[8] Traut, R.R., Lambert, J.M., Boileau, G. and Kenny, J.W. (1979) in: Ribosomes: Structure, Function and Genetics (Chambliss, G. et al. eds) pp. 89–110, Univ. Park Press, Baltimore MD.

[9] Wystup, G., Teraoka, W., Schulze, H., Hampl, H. and Nierhaus, K.H. (1979) Eur. J. Biochem. 100, 101–113.

[10] Röhl, R. and Nierhaus, K.H. (1982) Proc. Natl. Acad. Sci. USA 79, 729–733.

[11] Nierhaus, K.H. (1979) in: Ribosomes: Structure, Function and Genetics (Chambliss, G. et al. eds) pp. 267–294, Univ. Park Press, Baltimore MD.

[12] Hindennach, I., Kaltschmidt, E. and Wittmann, H.G. (1971) Eur. J. Biochem. 23, 12–16.

[13] Dijk, J. and Littlechild, J. (1979) Methods Enzymol. 59, 481–502.

[14] Riley, M. and Perham, R.N. (1970) Biochem. J. 118, 733–739.

[15] Diels, O. and Alder, K. (1929) Ber. Dtsch. Chem. Ges. 62, 554–562.
[16] Reinbolt, J., Tritsch, D. and Wittmann-Liebold, B. (1979) Biochimie 61, 501–522.
[17] Heiland, I., Brauer, D. and Wittmann-Liebold (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.
[18] Hitz, H., Schäfer, D. and Wittmann-Liebold, B. (1977) Eur. J. Biochem. 75, 497–512.
[19] Toi, K., Brynum, F., Norris, E. and Hano, H. (1967) J. Biol. Chem. 242, 1036–1043.
[20] Houmard, J. and Drapeau, G.R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[21] Lewis, W.G., Basford, J.M. and Walton, P.L. (1978) Biochim. Biophys. Acta 522, 551–560.
[22] Gross, E. and Witkop, B. (1962) J. Biol. Chem. 237, 1856–1860.
[23] Fontana, A. (1972) Methods Enzymol. 25, 419–423.
[24] Wittmann-Liebold, B. (1980) in: Polypeptide Hormones (Beers, R.F. jr and Bassett, E.G. eds) pp. 87–120, Raven, New York.
[25] Wittmann-Liebold, B. (1981) in: Chemical Synthesis and Sequencing of Peptides and Proteins (Liu, T.Y. et al. eds) pp. 75–110, Elsevier Biomedical, Amsterdam, New York.
[26] Wittmann-Liebold, B., Geissler, A.W. and Marzinzig, E. (1975) J. Supramol. Struct. 3, 426–447.
[27] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[28] Wittmann-Liebold, B. and Lehmann, A. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, Ch. ed) pp. 49–72, Elsevier Biomedical, Amsterdam, New York.
[29] Kimura, M., Foulaki, K., Subramanian, A.R. and Wittmann-Liebold, B. (1982) Eur. J. Biochem. 123, 37–53.
[30] Burgess, A.W., Ponnuswamy, P.K. and Scheraga, H.A. (1974) Israel J. Chem. 12, 239–286.
[31] Chou, P.Y. and Fasman, G.D. (1977) J. Mol. Biol. 115, 135–145.
[32] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[33] Nagano, K. (1977) J. Mol. Biol. 109, 251–274.
[34] Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327–356.
[35] Dzionara, M., Robinson, S.M.L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.
[36] Rawlings, N., Ashman, K. and Wittmann-Liebold, B. (1983) submitted.