AMINO ACID SEQUENCE OF PROTEIN L22 FROM THE LARGE SUBUNIT OF THE *ESCHERICHIA COLI* RIBOSOME

B. WITTMANN-LIEBOLD and B. GREUER

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

Received 15 September 1980

1. Introduction

Although protein L22 is a primary binding protein which binds specifically to the 23 S RNA, its binding is stimulated by other proteins, e.g., L4, L17 and L20, as shown by the assembly map [1]. It belongs to a group of proteins which are important during the early stages of 50 S reconstitution [2].

Immune electron microscopy has shown protein L22 to be located on the back of the 50 S 'armchair' [3]. Chemical crosslinks between L22 and L32 have been found [4] indicating that these two proteins are neighbours within the 50 S subunit. The isolation of a protein complex consisting of L22 and L19 [5] points to a strong interaction between the two proteins.

Protein L22 can be affinity labeled with a puromycin derivative [6], and this technique has also identified L22 to be near the 3'-end of tRNA bound to the ribosome (cited in [7]). Among the mutants with an altered L22 [8–10] there are two which are resistant to erythromycin. One of them has been studied genetically and biochemically in detail [8].

Here we report the complete amino acid sequence of protein L22 which consists of 110 amino acids and has M_r 12 227. We depict the secondary structure of this protein derived from 4 different prediction programmes. The results of a computer search for regions of homology in protein L22 and other ribosomal proteins, are given.

2. Materials and methods

Protein L22 was isolated from 50 S subunits of *E. coli* strain K12 according to [11] and was kindly provided by Dr H. G. Wittmann.

Sequence studies were performed on peptides prepared by the following cleavages of the intact protein:

(i) Digestion with trypsin (in 0.1 M *N*-methylmorpholine acetate buffer at pH 8.1, 37°C, 4–8 h, enzyme substrate ratio 1:50) and with trypsin after blocking the lysine residues with ETPA (exo-*cis*-3,6-endoxo-Δ^4-tetrahydrophthalic acid) [12,13];
(ii) Digestion with a protease from *Armillaria mellea* [14] which cleaves the N-terminal peptide bonds of lysines (same buffer at pH 8.1, 37°C, 6 h, enzyme substrate ratio 1:1000);
(iii) Digestion with chymotrypsin (same buffer at pH 8.1, 37°C, 1 h, enzyme substrate ratio 1:200);
(iv) Digestion with thermolysin (same buffer at pH 8.1, 50°C, 2–4 h, enzyme substrate ratio 1:100);
(v) Digestion with *Staphylococcus aureus* protease (same buffer at pH 8.1, 37°C, 20 h, enzyme substrate ratio 1:30);
(vi) Partial acid hydrolysis at peptide bonds adjacent to aspartic acid residues (in 2% acetic acid for 15 h at 110°C);
(vii) Cleavage with cyanogen bromide at methionine residues (in 70% formic acid, reagent to protein ratio 1:1, 24–48 h at room temperature).

The resulting peptides were isolated (see table 1) by:

(i) Thin-layer fingerprint technique;
(ii) Column chromatography on Dowex 50 (2 × 90 mm, 50°C), followed by one-dimensional preparative thin-layer purification;
(iii) Gel filtration on Sephadex columns (1 × 180 cm) in 10% acetic acid or dilute ammonia (pH 9.0) followed by separation on fingerprints or one-dimensional chromatography on thin-layer sheets as in [15,16].

Table 1
Isolation of peptides derived from protein L22

Cleavage[a]	Isolation procedure
Trypsin (TR)	(a) Fingerprint technique on thin-layer sheets (b) Dowex 50 micro-column in pyridine formate gradients, pH 2.7–6 as detailed in [17] (c) Gel filtration on Sephadex G-50 sf. in 0.07% ammonia in H_2O
Armillaria mellea protease (AMP)	Fingerprint technique on thin-layer sheets
ETPA	Gel filtration on Sephadex G-50 sf. in 10% acetic acid
ETPA-TR	Purified ETPA peptides after deblocking were further cleaved with trypsin at lysine; these tryptic fragments were isolated by fingerprinting on thin-layer sheets
Cyanogen bromide (CNBr)	Gel filtration on Sephadex G-50 sf., followed by rechromatography on Sephadex G-75 sf. for bigger fragments
CNBr-TR and CNBr-HAc	Purified cyanogen bromide fragments were further cleaved by trypsin or dilute acid hydrolysis and these fragments isolated by thin-layer fingerprints
Dilute acid hydrolysis (HAc)	Gel filtration on Sephadex G-50 sf. in 10% acetic acid
Chymotrypsin (CHY)	(a) Fingerprint technique on thin-layer sheets (b) Dowex 50 micro column in pyridine formate gradients, pH 2.7 to 6 as detailed in [17]
Thermolysin (TH)	(a) Fingerprint technique on thin-layer sheets (b) Dowex 50 micro column in pyridine formate gradients, pH 2.7 to 6 as detailed in [17]
Staphylococcus aureus protease (SP)	Fingerprint technique on thin-layer sheets

[a] *Abbreviations:* as in legend to fig.1

Sequence analysis of the intact protein was carried out in a modified Beckman sequencer programmed to use double coupling and cleavage reactions, as well as an automatic conversion device, as summarized in [18]. Two degradations were performed with 1.5–2 mg protein and the released PTH-amino acid derivatives were identified by a thin-layer technique and by mass spectrometry [19].

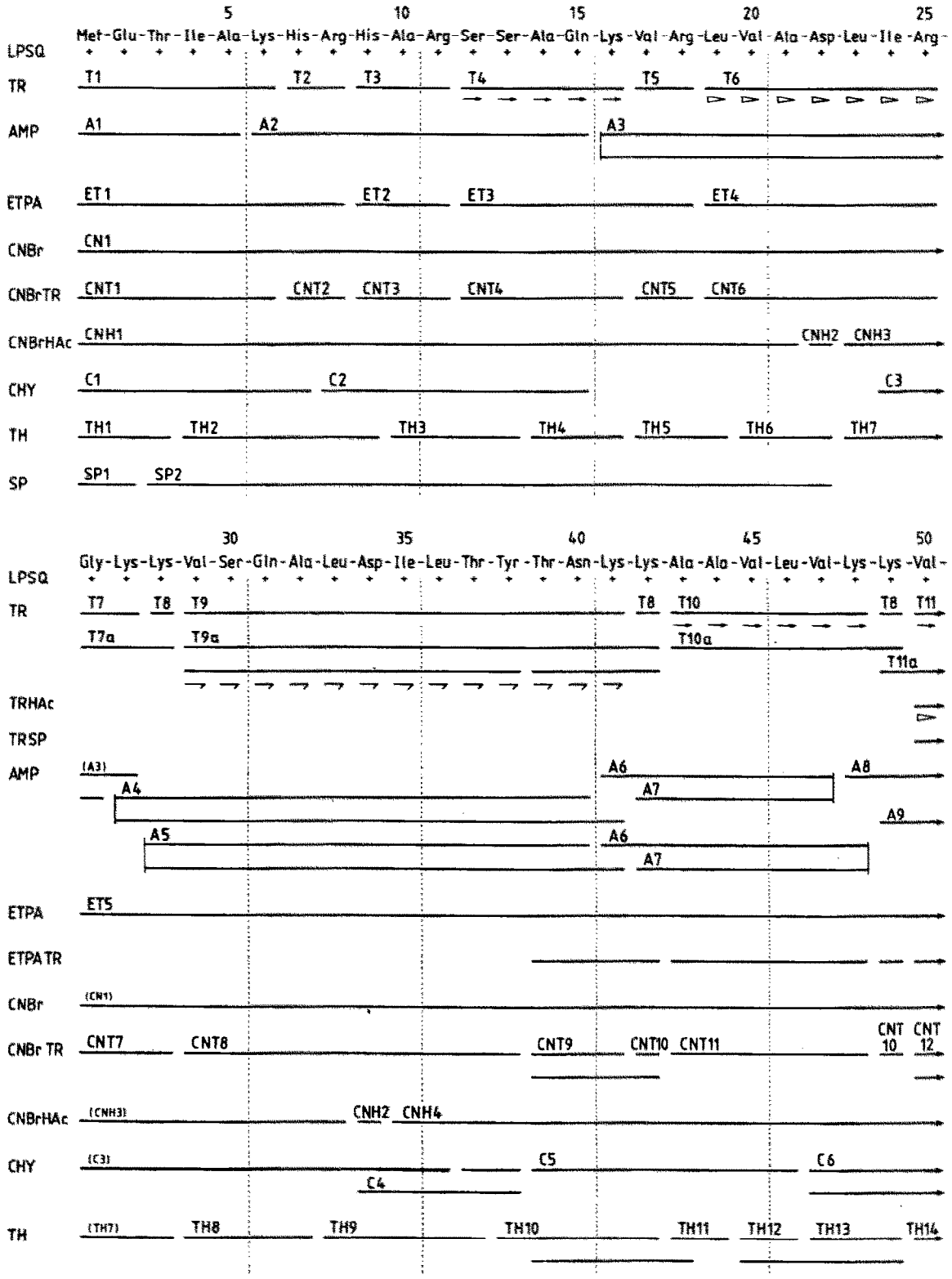
The diverse peptides were sequenced manually by the combined dansyl-Edman technique [20], or by the DABITC/PITC (4′-*N*,*N*′-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate) double coupling method [21]. Further the DABITC/PITC technique was manually applied to peptides linked covalently to diisothiocyanate glass support [22] as in [23]. Solid-phase sequencing was performed automatically (in a sequencer selfmade according to [24]) after attaching the carboxyl groups of the C-terminal end of the peptides with water-soluble carbodiimide to aminopolystyrene resin as in [25]. The released PTH-amino acid derivatives were identified by thin-layer techniques [19].

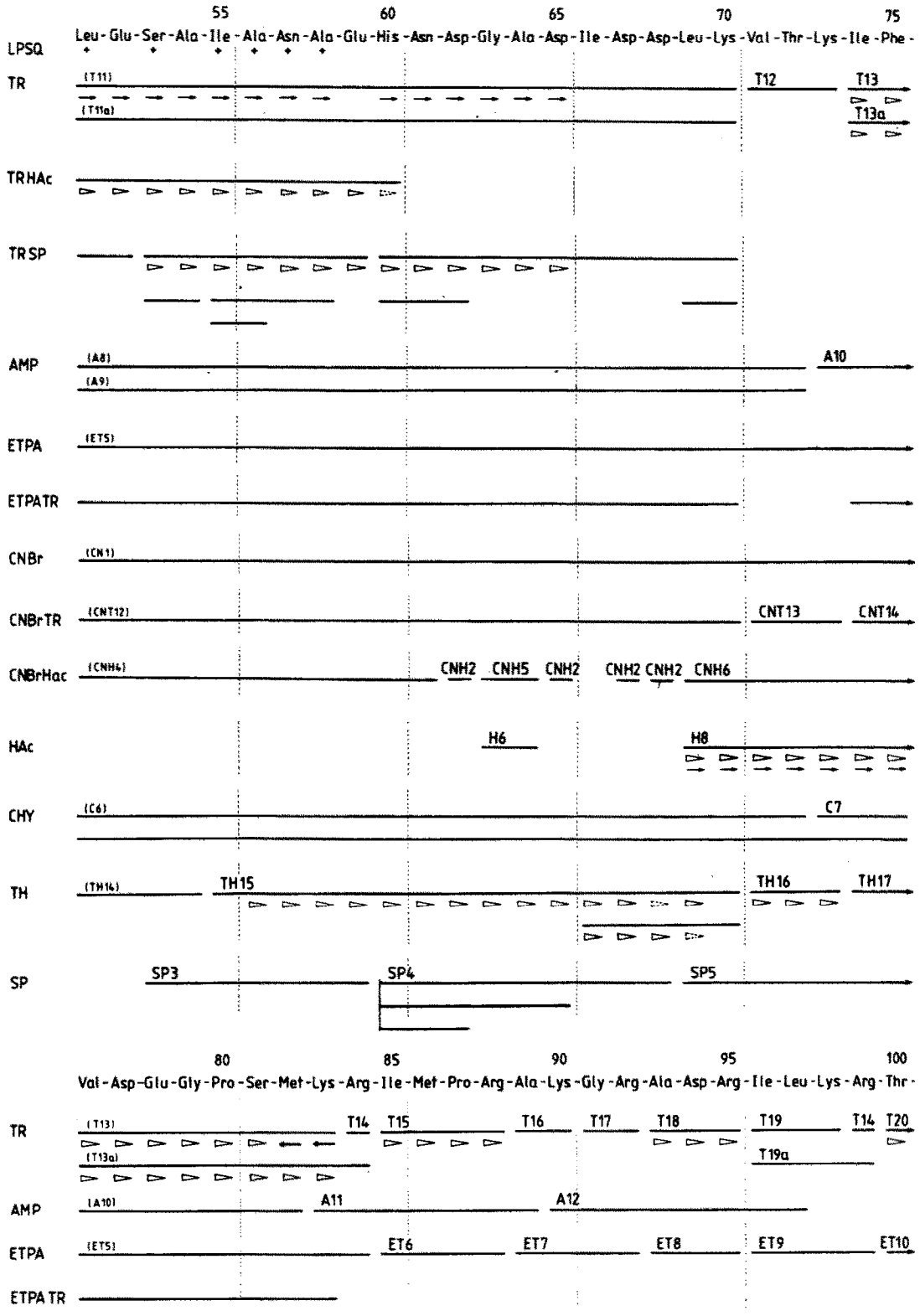
3. Results and discussion

3.1. *Sequence determination*

The N-terminal sequence of protein L22, up to position 58 was determined [19] by liquid-phase Edman degradation performed in a modified Beckman sequencer as in [18,19]. The sequence obtained was confirmed by sequencing peptides isolated from digestions of the protein with trypsin, chymotrypsin, thermolysin, *S. aureus* protease [26] and *A. mellea* protease [14]. Sequence analysis was performed on all tryptic peptides by the manual dansyl-Edman technique [20] and by either solid-phase [23–35] or the DABITC/PITC degradation method [21], as presented in fig.1.

The alignment of the tryptic peptides T11/T12/T13 (pos. 50–83) from the centre of the protein, was derived from the following fragments: (i) chymotrypsin peptides C6 (pos. 47–72), C7 (pos. 73–75), C8 (pos. 76–83); (ii) cyanogen bromide peptide CN1 (pos. 1–82) which on further treatment with trypsin released all the N-terminal tryptic peptides including T11 (minus lysine); (iii) ETPA peptide ET-5 (pos. 26–84) which after deblocking and trypsin treatment


5
10
15
20
25
Met-Glu-Thr-Ile-Ala-Lys-His-Arg-His-Ala-Arg-Ser-Ser-Ala-Gln-Lys-Val-Arg-Leu-Val-Ala-Asp-Leu-Ile-Arg-
LPSQ
TR
T1
T2
T3
T4
T5
T6
AMP
A1
A2
A3
ETPA
ET1
ET2
ET3
ET4
CNBr
CN1
CNBrTR
CNT1
CNT2
CNT3
CNT4
CNT5
CNT6
CNBrHAc
CNH1
CNH2
CNH3
CHY
C1
C2
C3
TH
TH1
TH2
TH3
TH4
TH5
TH6
TH7
SP
SP1
SP2
30
35
40
45
50
Gly-Lys-Lys-Val-Ser-Gln-Ala-Leu-Asp-Ile-Leu-Thr-Tyr-Thr-Asn-Lys-Lys-Ala-Ala-Val-Leu-Val-Lys-Lys-Val-
LPSQ
TR
T7
T8
T9
T8
T10
T8
T11
T7a
T9a
T10a
T11a
TRHAc
TRSP
AMP
(A3)
A4
A5
A6
A7
A8
A9
ETPA
ET5
ETPA TR
CNBr
(CN1)
CNBr TR
CNT7
CNT8
CNT9
CNT10
CNT11
CNT 10
CNT 12
CNBrHAc
(CNH3)
CNH2
CNH4
CHY
(C3)
C4
C5
C6
TH
(TH7)
TH8
TH9
TH10
TH11
TH12
TH13
TH14


55
60
65
70
75
LPSQ
Leu-Glu-Ser-Ala-Ile-Ala-Asn-Ala-Glu-His-Asn-Asp-Gly-Ala-Asp-Ile-Asp-Asp-Leu-Lys-Val-Thr-Lys-Ile-Phe-
TR
T11
T11a
T12
T13
T13a
TRHAc
TRSP
AMP
A8
A9
A10
ETPA
ET5
ETPATR
CNBr
CN1
CNBrTR
CNT12
CNT13
CNT14
CNBrHac
CNH6
CNH2
CNH5
CNH2
CNH2
CNH2
CNH6
HAc
H6
H8
CHY
C6
C7
TH
TH16
TH15
TH16
TH17
SP
SP3
SP4
SP5
80
85
90
95
100
Val-Asp-Glu-Gly-Pro-Ser-Met-Lys-Arg-Ile-Met-Pro-Arg-Ala-Lys-Gly-Arg-Ala-Asp-Arg-Ile-Leu-Lys-Arg-Thr-
TR
T13
T13a
T14
T15
T16
T17
T18
T19
T19a
T14
T20
AMP
A10
A11
A12
ETPA
ET5
ET6
ET7
ET8
ET9
ET10
ETPA TR

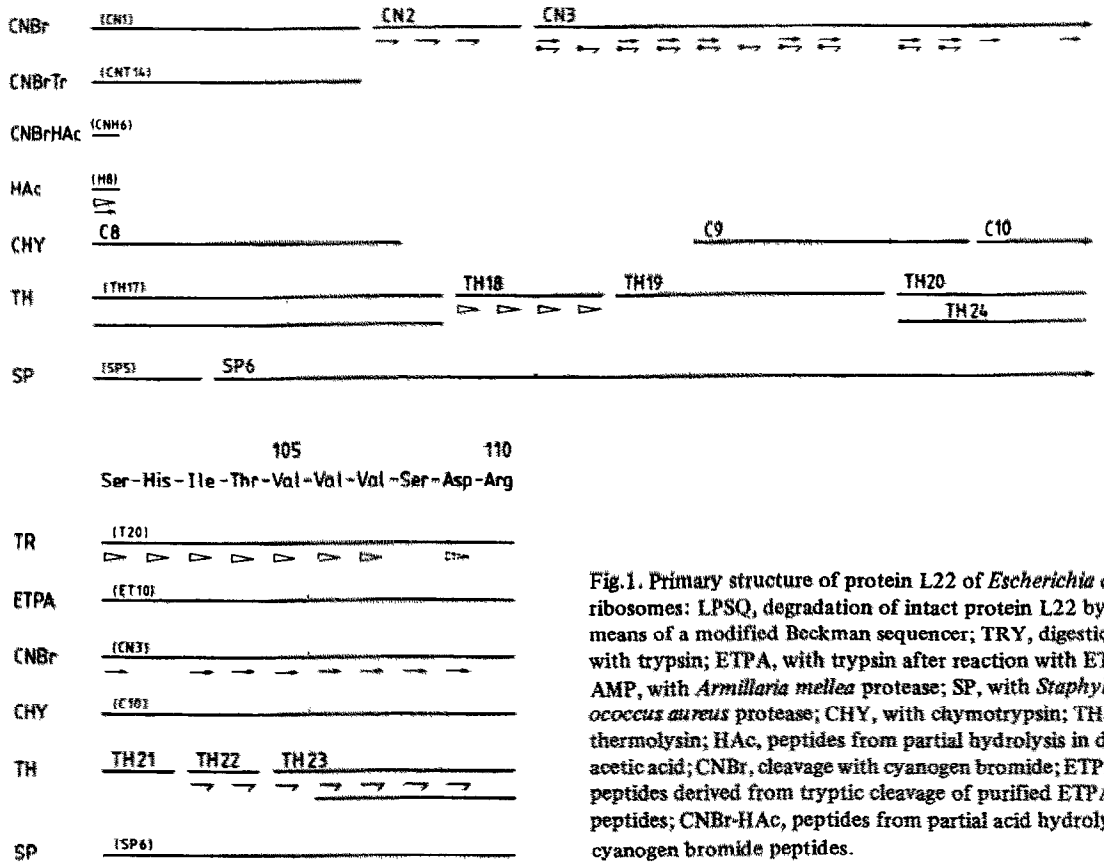


Fig.1. Primary structure of protein L22 of *Escherichia coli* ribosomes: LPSQ, degradation of intact protein L22 by means of a modified Beckman sequencer; TRY, digestion with trypsin; ETPA, with trypsin after reaction with ETPA; AMP, with *Armillaria mellea* protease; SP, with *Staphylococcus aureus* protease; CHY, with chymotrypsin; TH, with thermolysin; HAc, peptides from partial hydrolysis in dilute acetic acid; CNBr, cleavage with cyanogen bromide; ETPA-TR, peptides derived from tryptic cleavage of purified ETPA peptides; CNBr-HAc, peptides from partial acid hydrolysis of cyanogen bromide peptides.

gave peptides T7, T8, T9, T10, T11, T12, T13, and due to 2 lysine residues being adjacent additional peptides (T7a, T9a, T10a) with one more lysine residue (at pos. 28, 42, 49) were isolated (fig.1). The alignment agreed with the sequence of a peptide obtained by partial acid hydrolysis of the protein (peptide H8, pos. 69–76). It was difficult to establish the sequence of tryptic peptide T11, especially the second half, because:

(i) The peptide prepared by gel filtration was contaminated with peptides T13/T13a (pos. 74–84). This difficulty was overcome by either using long thin columns loaded with <2 mg tryptic digest dissolved in a small volume of 8 M urea, or isolating the chymotryptic peptide C6 or the thermolytic peptide TH15 (pos. 55–70) which correspond to the same region of the amino acid chain;

(ii) Incomplete tryptic cleavage at pos. 28 and 49 which led to bigger fragments containing peptide T11 but with heterogenous N-terminal sequences;

(iii) The large percentage of aspartic acid residues in peptide T11 which made it susceptible to cleavage during column runs with acidic solutions, during chromatography in the usual butanol–acetic acid–water–pyridine system and after repetitive Edman degradations which utilize strong acid at the cleavage stages. The latter problems were solved by employing dilute ammonia at pH 9 for the isolation of T11 or the thermolytic peptide TH15, and by employing the DABITC/PITC double coupling solid-phase method [23] but with short cleavage times (e.g., 10 min).

The C-terminal sequence of protein L22 was derived from sequencing the remaining tryptic peptides and the small cyanogen peptides CN2 (pos. 83–86) and CN3 (pos. 87–110) by the dansyl-Edman and solid-phase techniques. The other peptides obtained from this region concurred with the sequence derived (see fig.1).

Four independent complete sets of peptide fragments were employed to determine the sequence of protein L22, as presented in fig.1. The peptides were

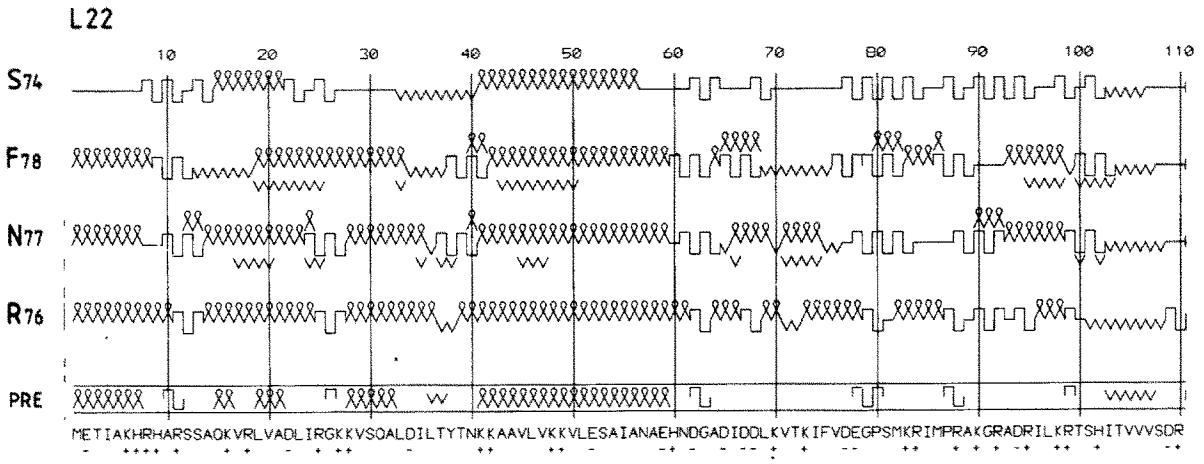


Fig.2. Predictions of the secondary structure of protein L22 according to 4 different methods (details in text and [27]).

obtained by cleaving the protein with trypsin, cyanogen bromide, thermolysin, and after blocking the lysines cleaving specifically at arginine.

The newly-employed enzyme *A. mellea* protease [14] gave all expected cleavages at the N-terminal peptide bonds of the lysines, and cleaved also at lysyl–lysine peptide bonds. The peptide yields were sufficient at the enzyme substrate ratio of 1:1000, 6 h cleavage time at 37°C in salt-free buffer conditions. Thus, the enzyme provides an easy means of generating tryptic bridging peptides, useful for sequencing or functional studies of proteins.

3.2. *Characteristics of the sequence*

Protein L22 has 110 amino acid residues and M_r 12 227. Its amino acid composition, as derived from the sequence given in fig.1, is Asp_9, Asn_3, Thr_6, Ser_7, Glu_4, Gln_2, Pro_2, Gly_4, Ala_{13}, Val_{11}, Met_3, Ile_9, Leu_8, Tyr_1, Phe_1, His_4, Lys_{13}, Arg_{10}, Trp_0, Cys_0. This is in good agreement with the results derived from amino acid analyses of the entire protein.

Although the basic amino acids clearly dominate in this protein, there is a distinct acidic region at pos. 52–68 which contains one of the histidine residues. The basic amino acids are not evenly distributed throughout the protein, they are clustered at the N-terminal end, pos. 6–18, 25–28, 41–49 and near at the C-terminus of the chain. The proline and inner methionine residues are close to each other at pos. 80–87. Repetitive sequence stretches are formed by valines and alanines at pos. 43–47, 54–58 and 105–107.

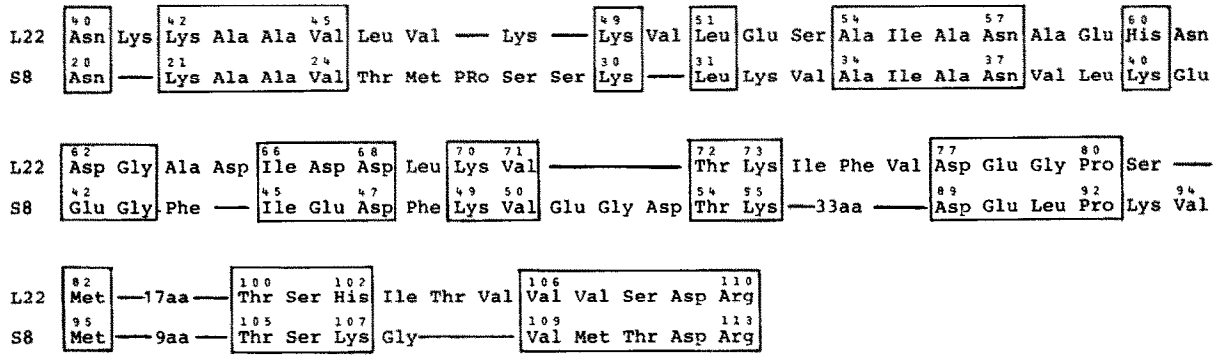


Fig.3. Identical or homologous sequence regions found for ribosomal proteins L22 and S8. Boxed are identical residues or conservative replacements, e.g., aspartic acid by glutamic acid, serine by threonine.

Table 2
Homologous regions of protein L22 and other ribosomal proteins

Protein	Positions	Homologous sequences
L22	18 - 22	Arg-Leu-Val-Ala-Asp
S3	113 - 117	Lys-Leu-Val-Ala-Asp
L22	63 - 67	Gly-Ala-Asp-Ile-Asp
S3	158 - 161	Gly-Ala-Glu-Ile
	179 - 182	Ala-Asp-Ile-Asp
L22	15 - 19	Gln-Lys-Val-Arg————Leu
S10	4 - 10	Gln-Arg-Ile-Arg-Ile-Arg-Leu
L22	16 - 19	Lys-Val-Arg-Leu
S12	53 - 56	Arg-Val-Arg-Leu
L22	45 - 48	Val-Leu-Val-Lys
S12	79 - 82	Ile-Leu-Ile-Arg
L22	106 - 110	Val-Val-Ser-Asp-Arg
S17	11 - 15	Val-Val-Ser-Asp-Lys
L22	63 - 70	Gly-Ala-Asp-Ile-Asp-Asp-Leu-Lys
L7/L12	114 - 120	Gly-Ala-Glu-Val-Glu———Val-Lys
L22	41 - 45	Lys-Lys-Ala-Ala-Val
L15	69 - 73	Arg-Lys-Ala-Ala-Ile
L22	25 - 28	Arg-Gly-Lys-Lys
L24	18 - 21	Lys-Gly-Lys-Arg
L22	46 - 49	Leu-Val-Lys-Lys
L24	40 - 43	Leu-Val-Lys-Lys
L22	69 - 72	Leu-Lys-Val-Thr
L28	48 - 51	Leu-Arg-Val-Ser
L22	88 - 91	Arg-Ala-Lys-Gly
L28	71 - 74	Arg-Ala-Arg-Gly
L22	25 - 29	Arg-Gly-Lys-Lys-Val
L32	49 - 53	Arg-Gly-Arg-Lys-Val
L22	16 - 20	Lys-Val-Arg-Leu-Val
L33	7 - 11	Lys-Ile-Lys-Leu-Val
L22	88 - 93	Arg-Ala-Lys-Gly-Arg-Ala
L34	35 - 40	Arg-Ala-Lys-Gly-Arg-Ala
L22	47 - 53	Val-Lys-Lys-Val-Leu-Glu-Ser
RL-P3	23 - 29	Ile-Lys-Lys-Ile-Leu-Asp-Ser
L22	63 - 68	Gly-Ala-Asp-Ile-Asp-Asp
SC-YP-A1	31 - 36	Gly-Ala-Glu-Val-Asp-Glu

3.3. *Secondary structure predictions of protein L22*

Four different predictive methods (described in [27]) were employed to calculate the secondary structure of protein L22, as presented in fig.2. The recent parameters were used for the Chou and Fasman prediction [28,29]. The final line of fig.2 represents agreement between 3 out of the 4 predictions, and the percentage of each type of structure is as follows: 33% helix, 10% turns or loops and ⩾6% extended structure (see line 'PRE' in fig.2).

3.4. *Comparison with sequences of other ribosomal proteins*

The sequence of protein L22 was compared with 50 sequenced *E. coli* ribosomal proteins (reviewed in [30]) and with 7 complete protein structures from other organisms (yeast, *Bacillus subtilis, Artemia salina* and rat liver; reviewed in [31]). The results are listed in table 2, where identical or similar sequence regions are given. The greatest degree of homology was found between L22 and S8, but there were no long regions (see fig.3). Interestingly both proteins belong to the group of ribosomal proteins found to be rRNA primary binding proteins. At the beginning of the 'homologous' sequence both proteins have a strongly predicted α-helix region, and at the C-terminal end alternate β-sheet and turn regions are predicted.

Acknowledgements

We wish to thank Mrs J. Krauss and Mr W. Kühnau for excellent technical assistance during the early stage of the sequencing work performed on this ribosomal protein. Dr Michael Dzionara developed the computer search programme for homologous sequence regions. We thank Dr V. Barkholt Pedersen, Institut før Biokemisk Genetik, København for kindly supplying us with a sample of *Armillaria mellea* protease.

References

[1] Roth, H. E. and Nierhaus, K. H. (1980) Eur. J. Biochem. 103, 95–98.
[2] Spillmann, S., Dohme, F. and Nierhaus, K. H. (1977) J. Mol. Biol. 115, 513–523.
[3] Stöffler, G., Bald, R., Kastner, B., Lührmann, R., Stöffler-Meilicke, M. and Tischendorf, G. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 171–205, University Park Press, Baltimore, MD.
[4] Kenny, J. W. and Traut, R. R. (1976) J. Mol. Biol. 127, 243–263.

[5] Wystup, G., Teraoka, H., Schulze, H., Hampl, H. and Nierhaus, K. H. (1979) Eur. J. Biochem. 100, 101–113.
[6] Krassnigg, F., Erdmann, V. A. and Fasold, H. (1978) Eur. J. Biochem. 87, 439–443.
[7] Cooperman, B. S. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 531–554, University Park Press, Baltimore, MD.
[8] Wittmann, H. G., Stöffler, G., Apirion, A., Rosen, L., Tanaka, K., Tamaki, M., Takata, R., Dekio, S., Otaka, E. and Osawa, S. (1973) Mol. Gen. Genet. 127, 175–189.
[9] Isono, K., Krauss, J. and Hirota, Y. (1976) Mol. Gen. Genet. 149, 297–302.
[10] Dabbs, E. R. and Wittmann, H. G. (1976) Mol. Gen. Genet. 149, 303–309.
[11] Hindennach, I., Kaltschmidt, E. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 12–16.
[12] Hitz, H., Schäfer, D. and Wittmann-Liebold, B. (1977) Eur. J. Biochem. 75, 497–512.
[13] Wittmann-Liebold, B. and Greuer, B. (1979) FEBS Lett. 108, 69–74.
[14] William, G. L., Basford, J. M. and Walton, P. L. (1978) Biochim. Biophys. Acta 522, 551–560.
[15] 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) pp. 219–232, Elsevier/North-Holland, Amsterdam, New York.
[16] Wittmann-Liebold, B. and Lehmann, A. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, Ch. ed) pp. 49–72, Elsevier/North-Holland, Amsterdam, New York.
[17] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe Seyler's Z. Physiol. Chem. 375, 1751–1770.
[18] Wittmann-Liebold, B. (1980) in: Polypeptide Hormones (Beers, R. F. et al. eds) Raven Press, New York, in press.
[19] Wittman-Liebold, B., Geissler, A. W. and Marzinzing, E. (1975) J. Supramol. Struct. 3, 426–447.
[20] Chen, R. (1976) Hoppe Seyler's Z. Physiol. Chem. 375, 873–886.
[21] Chang, J. Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[22] Wachter, E., Machleidt, W., Hofner, H. and Otto, J. (1973) FEBS Lett. 35, 97–102.
[23] Chang, J. Y. (1979) Biochim. Biophys. Acta 578, 188–195.
[24] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[25] Wittmann-Liebold, B. and Lehmann, A. (1975) in: Solid-Phase Methods in Protein Sequence Analysis (Laursen, R. A., ed.) pp. 81–90, Pierce Chem. Co., Rockford, IL.
[26] Houmard, J. and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[27] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe Seyler's Z. Physiol. Chem. 358, 1003–1019.
[28] Chou, P. Y. and Fasman, G. D. (1977) J. Mol. Biol. 115, 135–175.
[29] Chou, P. Y. and Fasman, G. D. (1978) Ann. Rev. Biochem. 47, 251–276.
[30] Wittmann, H. G., Littlechild, J. and Wittmann-Liebold, B. (1980) in: Ribosomes (Chambliss, G. et al. eds) pp. 51–88, University Park Press, Baltimore, MD.
[31] Wittmann-Liebold, B. (1980) in: RNA-Polymerase, tRNA and Ribosomes (Osawa, S. ed) University of Tokyo Press, Hongo, in press.