# Primary Structure of *Escherichia coli* Ribosomal Protein L31†

Jürgen Brosius‡

ABSTRACT: Protein L31 from the 50S ribosomal subunit of *Escherichia coli* was manually sequenced by the dansyl-Edman method. Owing to the availability of only small quantities of purified L31, sequencing methods were scaled down such that the entire primary structure could be determined with 700 μg of protein. The techniques employed are described in detail. The protein consists of a single chain of 62 amino acids, with a calculated molecular weight of 6967. Four half-cystine residues were identified at positions 16, 18, 37, and 40. Evidence is presented that suggests that these residues form two disulfide bridges in the protein, as isolated.

For a detailed understanding of the structure and function of the ribosome it is crucial to know the primary structure of its components. Elucidation of the primary structure of both ribosomal RNA and the 54 ribosomal proteins is currently underway (for a review see Brimacombe et al., 1976). Protein sequence information is of great help in the interpretation of a wide range of structure–function studies, such as x-ray diffraction, genetic analysis, and evolutionary studies. Furthermore, knowledge of the amino acid sequences of ribosomal proteins will facilitate a more precise location of protein–protein cross-links, binding sites of antibiotics and antibodies against ribosomal proteins, and sites of interaction with ribosomal RNA.

Protein L31 is one of a group of very small (46 to 63 amino acids long) basic proteins from the 50S ribosomal subunit of *Escherichia coli*. It is reported to be a "fractional" protein, occurring only in a fraction of 50S ribosomal subunits, as isolated (Weber, 1972). As shown here, L31 is remarkable for its high content of half-cystine. To my knowledge, no functional studies have been reported for this protein. Here I present the determination of the complete primary structure of protein L31, a single polypeptide chain of 62 amino acids, containing 4 half-cystine residues. Due to the scarcity of material, methods were scaled down to allow the complete structural determination to be carried out on 700 μg of protein. These methods are of general applicability to proteins of biological interest that are available only in small quantities.

## Materials and Methods

If not stated otherwise, chemicals were of analytical grade purchased from Merck. Apamine, an oligopeptide containing half-cystine residues in positions 1, 3, 11, and 15, was from Serva (Heidelberg, West Germany) and used as a reference for sequencing Cys-containing peptides.

*Isolation of Protein L31.* Preparation of 50S subunits from *E. coli* strain B ribosomes was performed as described (Hindennach et al., 1971) and a group of proteins containing L31 was prefractionated according to Schwabe (1972) at pH 3.8. The split proteins were desalted on Sephadex G25 and lyophilized. An aliquot of the protein mixture was examined by two-dimensional polyacrylamide gel electrophoresis (Kaltschmidt and Wittmann, 1970).

The protein mixture was dissolved in 100 mL of 0.01 M sodium acetate (pH 5.6), 6 M urea, and applied to a column (1.5 × 20 cm) of carboxymethylcellulose (Whatman CM 52) and the proteins were eluted with a 2-L linear gradient from 0.01 to 0.4 M sodium acetate (pH 5.6), in 6 M urea. About 130 fractions containing 15 mL were collected at a flow rate of 1 drop per 15 s and the absorption at 260 nm was recorded with a Spectrochrom M (Gilson). Peak fractions were desalted with Sephadex G25 m (eluent, 15% acetic acid) and lyophilized. About 5–10 μg was examined by a mini-two-dimensional polyacrylamide gel electrophoresis (Isono, 1974). The gels were stained with Coomassie blue R 250 (Serva, Heidelberg, West Germany).

*Amino Acid Composition.* Protein (10 μg) was hydrolyzed in vacuo in a glass tube (7 × 78 mm) with 100 μL of constant boiling 5.7 N HCl containing 0.02% 2-mercaptoethanol and a small amount of crystalline phenol (Sanger and Thompson, 1953). After 20-h hydrolysis at 110 °C the sample was dried down in vacuo and analyzed with a Durrum D-500 amino acid analyzer (Palo Alto, Calif.) at a sensitivity of 2.0 A. In the same way 10 μg of oxidized protein was treated except that phenol was omitted. Oxidation of protein L31 was carried out with performic acid according to Hirs (1956). The tryptophan content of the protein was checked by examining all thin-layer fingerprints under ultraviolet (UV) light at 366 nm. Tryptophan-containing peptides show a blue fluorescence, especially after performic acid oxidation. The N-terminal amino acid was determined by dansylation as described for peptides, using 1 μg of L31.

*Enzymatic Digestion of Protein L31 and Isolation of the Resulting Peptides.* L31 was digested enzymatically with trypsin, chymotrypsin, *Staphylococcus aureus* protease (Houmard and Drapeau, 1972), thermolysin, pepsin, and carboxypeptidases A, B, and C (Tschesche and Kupfer, 1972) in small conical glass tubes. All enzymes except for the carboxypeptidases were stored in aqueous stock solutions (1 mg/mL) at −20 °C for several weeks. The cleavage conditions and source of enzymes are summarized in Table I.

Digestions were stopped by quick-freezing and lyophilization. Peptides were separated by mapping on cellulose thin-layer plates as described by Yaguchi et al. (1975), with the modification that 20 × 20 cm Polygram CEL 300 (Machery and Nagel, Düren, West Germany) plates were used, electrophoresis was carried out at 500 V for 110 min, and chromatography was carried out for 5–6 h. Peptides were stained either by spraying lightly with ninhydrin (0.3 mg of ninhydrin; 3.0 mL of 2,4,6-trimethylpyridine, 10 mL of acetic acid, and

† From the Max-Planck Institut für Molekulare Genetik, Abt. Wittman, Berlin-Dahlem, West Germany. *Received September 8, 1977.*

‡ Present address: Thimann Laboratories, Division of Natural Sciences, University of California, Santa Cruz, Calif. 95064.

0006-2960/78/0417-0501$01.00/0 

TABLE I: Enzymatic Digestion of Protein L31.

| Expt | Enzyme | Amt of protein (μg) and state | Amt of enzyme (μg) | Buffer used | Time and temp (°C) | Purpose[f] |
|---|---|---|---|---|---|---|
| 1 | Trypsin-Tos-PheCH$_2$Cl[a] | 75, unoxidized | 10 | 150 μL of buffer I[d] | 4 h, 37 | aac |
| 2 | Trypsin-TosPheCH$_2$Cl[a] | 70, unoxidized | 10 | 150 μL of buffer I[d] | 4 h, 37 | seq |
| 3 | Trypsin-TosPheCH$_2$Cl[a] | 60, oxidized | 5 | 150 μL of buffer I[d] | 4 h, 37 | seq |
| 4 | Trypsin-TosPheCH$_2$Cl[a] | 30, oxidized | 2.5 | 100 μL of buffer I[d] | 4 h, 37 | aac |
| 5 | α-Chymotrypsin[a] | 110, unoxidized | 3 | 150 μL of buffer I[d] | 20 min, 37 | aac (50 μg)<br>seq (60 μg) |
| 6 | *Staph. aureus* protease[b] | 50, unoxidized | 4 | 150 μL of 0.05 M potassium phosphate (pH 7.8) | 16 H, 37 | aac |
| 7 | Thermolysin[c] | 50, unoxidized | 5 | 150 μL of buffer I[d] | 90 min, 45 | aac |
| 8 | Thermolysin[c] | 60, unoxidized | 5 | 150 μL of buffer I[d] | 90 min, 45 | seq |
| 9 | Pepsin 2 × cryst. porcine[c] | 50, unoxidized | 5 | 150 μL of 0.05 N HCl | 1 h, 37 | aac |
| 10 | Pepsin 2 × cryst. porcine | 50, oxidized | 5 | 150 μL of 0.05 N HCl | 1 h, 37 | seq |
| 11 | Pepsin 2 × cryst. porcine | 30, oxidized | 2.5 | 100 μL of 0.05 N HCl | 1 h, 37 | aac |
| 12 | Carboxypeptidase A[e] | 10, unoxidized | 0.5 | 100 μL of buffer I[d] | 20 min, 37 | aac |
| 13 | Carboxypeptidase B[e] | 10, unoxidized | 0.5 | 100 μL of buffer I[d] | 20 min, 37 | aac |
| 14 | Carboxypeptidase C[e] | 10, unoxidized | 20 mU | 200 μL of 0.05 M sodium citrate buffer, pH 5.3 | 20 min, 50 | aac |

[a] Worthington, Freehold, N.J. [b] Miles, Slough, U.K. [c] Serva, Heidelberg, West Germany. [d] Buffer I: 0.2 M *N*-methylmorpholine acetate, pH 8.0. [e] Böhringer, Mannheim, West Germany. [f] Abbreviations used are: aac, amino acid composition; seq, sequence analysis.

TABLE II: Amino Acid Composition of Protein L31.[a]

| | Unoxidized protein | Oxidized protein | Values derived from sequence |
|---|---|---|---|
| Cys | | 3.83[b] | 4 |
| Asn | 7.91 | 7.72 | 3 |
| Asp | | | 5 |
| Thr | 3.74 | 3.55 | 4 |
| Ser | 3.42 | 3.11 | 4 |
| Gln | 2.87 | 3.39 | 1 |
| Glu | | | 2 |
| Pro | 2.11 | 2.24 | 2 |
| Gly | 5.00 | 5.64 | 5 |
| Ala | 2.29 | 2.23 | 2 |
| Val | 4.61 | 5.34 | 5 |
| Met | 1.47 | 1.92[c] | 2 |
| Ile | 2.47 | 3.10 | 3 |
| Leu | 2.05 | 1.97 | 2 |
| Tyr | 0.85 | 0.31 | 1 |
| Phe | 2.80 | 2.75 | 3 |
| His | 2.65 | 2.58 | 3 |
| Lys | 6.77 | 7.07 | 7 |
| Arg | 4.00 | 4.00 | 4 |
| Trp | | | 10 |

[a] Amino acid composition as determined after hydrolysis at 110 °C for 20 h (uncorrected values). [b] Determined as cysteic acid. [c] Determined as methionine sulfone.

95 mL of ethanol) or with 0.0005% fluorescamine (Hoffmann-La Roche, Basel, Switzerland) in acetone after adjusting the pH of the plate by spraying with 5% pyridine in acetone (Vandekerckhove and Van Montagu, 1974).

Peptide spots were scratched off the thin-layer plates with a spatula and extracted with 300 μL of 5.7 N HCl/0.02% 2-mercaptoethanol or with 300 μL of 50% acetic acid. Peptides extracted with hydrochloric acid were hydrolyzed as described above. The sensitivity range on the amino acid analyzer was 0.5 or 1.0 A. Peptides extracted with acetic acid were evaporated to dryness over sodium hydroxide.

*Dansyl-Edman Degradation.* The procedure was carried out according to Gray and Hartley (1963), Hartley (1970), and Bruton and Hartley (1970) with minor modifications.

Aqueous pyridine (50 μL; 50%) (3 times distilled over ninhydrin) was added to the dried peptide extracted from one fingerprint in a glass-stoppered tube (10 × 55 mm). For determination of the N-terminal amino acid 2.5 μL was removed into small glass tubes (3 × 50 mm). Then 1.3 μL of twice-distilled phenyl isothiocyanate (Pierce, Rockfort, Ill.; stored under nitrogen at −20 °C) was added; the tubes were flushed with nitrogen and stoppered. After vortexing, the samples were incubated for 1 h at 45 °C or 0.5 h at 50 °C in a heating block. The incubation mixture was thoroughly dried over sodium hydroxide and phosphorus pentoxide in a desiccator preheated to 60 °C. The residue was dissolved in 100 μL of twice-distilled trifluoroacetic acid (Fluka), vortexed and incubated 15 min at 45 °C. Following evaporation over NaOH, 50 μL of double distilled water was added and the 2-anilino-5-thiazolinones of the respective amino acids were extracted twice with 200 μL of *n*-butyl acetate, and stored in hydrolysis tubes (7 × 70 mm) at −20 °C. The aqueous phase was evaporated, redissolved in 50 μL of 50% pyridine, 2.5 μL was removed for dansylation and the rest submitted to a new degradation cycle.

The tubes containing aliquots from the Edman cycle were centrifuged briefly in a clinical centrifuge and evaporated to dryness. To the tube 5 μL of a 1:1 mixture of 0.2 M sodium bicarbonate and 2.5 mg/mL of dansyl[1] chloride in acetone were added. After sealing with Parafilm incubation was carried out at 45 °C for 10 min. The reaction mixture was taken to dryness and 10 μL of 5.7 N HCl/0.02% 2-mercaptoethanol was added. After sealing the tubes, the peptides were hydrolyzed for 8–15 h at 105 °C. The acid was evaporated over NaOH and the residue dissolved in 5 μL of 95% ethanol. The solution was spotted 7 mm from both sides of a 5 × 5 cm polyamide plate (Schleicher und Schüll, Dassel, West Ger-

[1] Abbreviations and symbols used are: aa, amino acid; Cya, cysteic acid; $MeO_2$, methionine sulfone; Tos-PheCH$_2$Cl, L-(1-tosylamido-2-phenyl)-ethyl chloromethyl ketone; dansyl chloride (Dns-Cl), 5-dimethylamino-1-naphthalenesulfonyl chloride; Dns, 5-dimethylamino-1-naphthalenesulfonyl; Dns-OH, 5-dimethylamino-1-naphthalenesulfonic acid; ATZ, 2-anilino-5-thiazoline; Pth, 3-phenyl-2-thiohydantoin; Ptc, phenylthiocarbamoyl; T, peptides resulting from digestion with trypsin; TL, peptides resulting from digestion with thermolysin; CH, peptides resulting from digestion with α-chymotrypsin; SP, peptides resulting from digestion with *Staphylococcus aureus* protease; P, peptides resulting from digestion with pepsin; →, dansyl-Edman degradation; ·→, dansyl-Edman degradation plus hydrolysis of ATZ derivatives into amino acids and identification by amino acid analysis; |→, dansyl-Edman degradation plus identification of the respective Pth-aa by thin-layer chromatography.

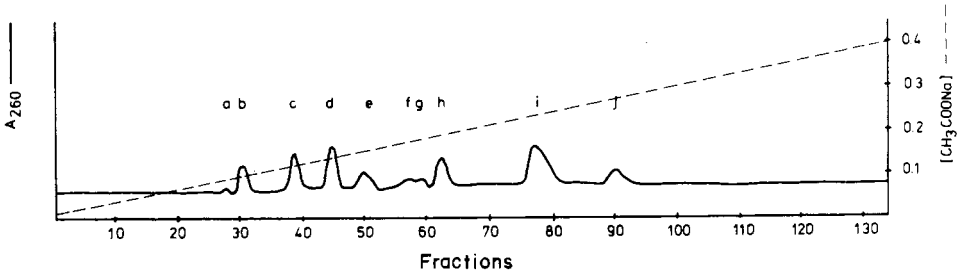



FIGURE 2: Elution profile of the pH 3.8 protein fraction (see Figure 1; supplementary material) on a carboxymethylcellulose column (1.5 × 20 cm); concentration of $NaOOCCH_3$ in moles.

many). A reference mixture was applied to the back side exactly opposite the unknown sample. The chromatography solvents employed were: (A) 1.5% formic acid; (B) toluene–acetic acid (10:1); (C) ethyl acetate–methanol–acetic acid (20:1:1); (D) 0.05 M trisodium phosphate–ethanol (3:1); (E) 1 M ammonia–ethanol (1:1); (F) 0.15 M ammonia; (G) *n*-heptane–1-butanol–formic acid (10:10:1). Usually the Dns-amino acids were first chromatographed in system A and then perpendicular to the first direction in systems B and C. In order to discriminate Dns-Arg from ε-Dns-Lys system D was used in the second dimension in addition to solvents B and C (Weiner et al., 1972). If in oxidized peptides Dns-Cya was detected, system D was replaced by system E. Dns-His could be visualized by first chromatographing in system F and then perpendicular to that direction in system G.

*Assignment of Amides.* Discrimination of aspartic acid and glutamic acid from their respective amides and assignment of Trp was carried out by converting the 2-anilino-5-thiazolinones into the respective Pth-amino acids (Edman and Henschen, 1975) with the following modifications: the organic phase from the Edman degradation was dried down and incubated at 80 °C for 5 min in 100 μL of 1 N HCl. The resulting Pth-aa were extracted 3 times with 100 μL of ethyl acetate, dried down, redissolved in 15 μL of ethylene chloride–ethanol (7:3), and spotted on precleaned (methanol–1-propanol–2-propanol; 98:1:1) silica gel thin-layer plates (20 × 20 cm, Kieselgel 60G 254, Merck). References containing (a) Pth-Glu, Pth-Gln, and Pth-Trp, and (b) Pth-Asp and Pth-Asn were spotted next to the samples. Chromatography of the Pth-aa was performed for 60 min in chloroform–1-propanol–2-propanol (98:1:1) and after drying for 60 min in ethylene chloride–acetic acid (6:1) in the same direction. The spots were detected under UV light at 254 nm or by a ninhydrin color reaction (Roseau and Pantel, 1969; Inagami and Murakami, 1972).

*Hydrolysis of Thiazolinones.* Hydrolysis of the amino acid 2-anilino-5-thiazolinones with HI was carried out according to Smithies et al. (1971). If Cya or Met were expected HI was replaced by 5.7 N HCl or 5.7 N HCl–0.02% 2-mercaptoethanol, respectively. The resulting amino acids were identified on a Durrum D-500 amino acid analyzer at a sensitivity of 0.2 A.

## Results

*Isolation of Protein L31.* Extraction of the 50S subunits at pH 3.8 released proteins L11, L25, and L33 and minor amounts of other proteins as well as a significant amount of L31 (Figure 1) (see Supplementary Material Available paragraph at end of text). Pure (>90%) L31 was obtained by a single carboxymethylcellulose chromatography step as shown in Figure 2. Peak e was identified to contain protein L31 by mini two-dimensional polyacrylamide gel electrophoresis (Figure 3) (Supplementary Material). A total yield of 1.06 mg of the lyophilized protein L31 was obtained from 2.5 g of 50S ribosomal subunits.

*Amino Acid Composition.* The amino acid compositions of the unoxidized and oxidized proteins are summarized in Table II. Nonoxidized protein showed a small peak in the cysteine position, which is also the position of *N*-monomethylmethionine (Brosius, 1976; Chen et al., 1977). After oxidation, however, only cysteic acid, but no *N*-monomethylmethionine sulfone, was observed. The absence of tryptophan was inferred from the lack of any fluorescent spot in any of the peptide maps, and by unambiguous identification of amino acids at all positions by the dansyl-Edman method. The reliability of detection of Trp by fluorescence under long-wavelength UV excitation has been previously established during sequence studies on Trp-containing proteins L4 (Stahl and Brosius, unpublished), L16 (Brosius, 1976; Brosius and Chen, 1976), and L19 (Brosius and Arfsten, 1978), where the presence or absence of Trp was confirmed by the use of Ehrlich's reagent and direct identification of Pth-Trp.

*Sequence Studies.* A compilation of the amount of protein used for the various enzymatic digestions is given in Table I.

*Tryptic Peptides.* All tryptic peptides were isolated and sequenced completely by the described methods. Amino acid compositions of the tryptic peptides are presented in Table III (supplementary material). Tryptic peptide maps of the unoxidized and oxidized protein are shown in Figures 4a and 4b, respectively.

$$\underset{\text{T1}}{\overrightarrow{\text{Met-Lys-Lys}}}$$

As observed with other ribosomal proteins (e.g., peptide T1 in protein L18; Brosius et al., 1975) trypsin cleaved only after the second of two adjacent lysine residues. T1 occurred as two spots on the peptide map due to partial oxidation of methionine to methionine sulfoxide. The peptide containing the oxidized methionine has a lower $R_f$ value in the described chromatography solvent. Peptide T1 is the N-terminal peptide of the protein, as dansylation of intact L31 revealed N-terminal methionine, in agreement with the results of Alakhov et al. (1975), and no other tryptic peptide contains N-terminal methionine.

$$\underset{\text{T2}}{\overrightarrow{\text{Asp-Ile-His-Pro-Lys}}}$$

Peptide T2 was completely sequenced by the dansyl-Edman method. Dns-His was determined in systems F and G, as a spot migrating a little bit faster than Dns-OH in both solvents. Lysine was detected as α-Dns-ε-Ptc-Lys which runs in system B slightly slower than Dns-Leu and in system C like Dns-Phe. Aspartic acid was determined by conversion of the ATZ derivative into Pth-Asp and by chromatography on silica gel

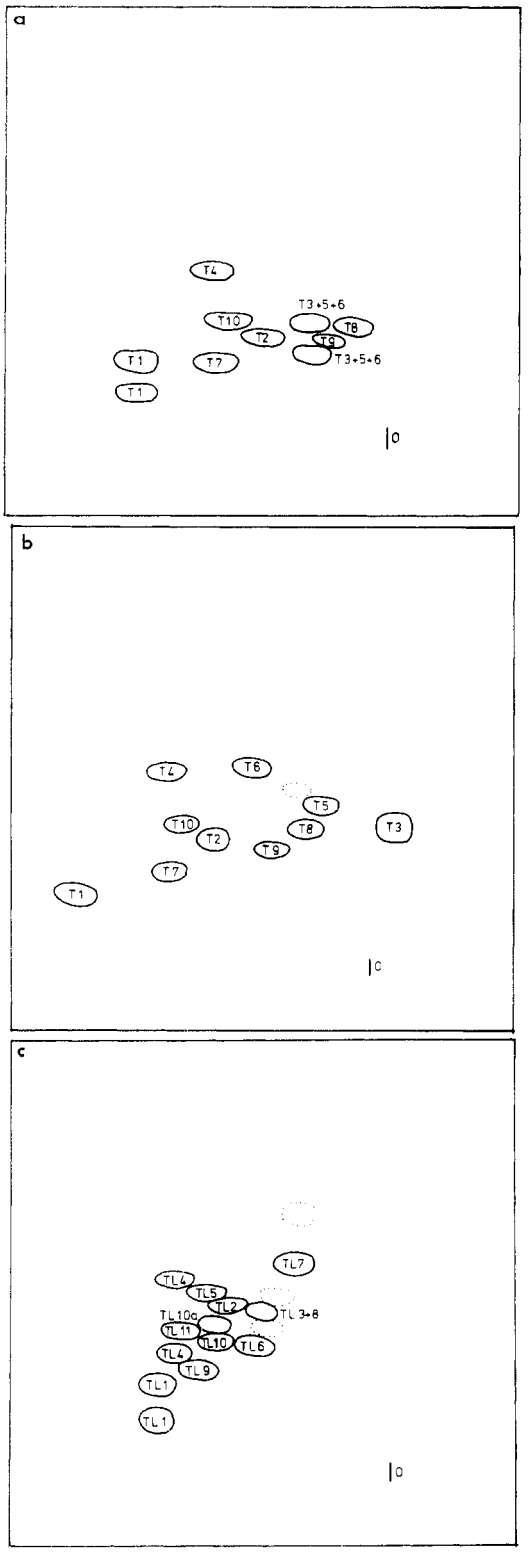



FIGURE 4: Thin-layer peptide maps from enzymatic digests of protein L31; O = origin; dotted lines indicate minor amounts of peptides generated by incomplete or unspecific digestion. In all peptide maps shown the direction of electrophoresis is horizontal (cathode left and anode right) and the direction of chromatography is vertical (bottom to top): (a) tryptic digest of unoxidized protein; (b) tryptic digest of oxidized protein; (c) thermolysin digest of unoxidized protein. Figures 4d–f can be found in the supplementary material.

thin-layer plates. The remaining four residues were confirmed by hydrolysis of the ATZ.

T3 +-5 +-6. On tryptic fingerprints from unoxidized protein there were two spots (Figure 4a) which gave identical amino acid compositions. The different migration in the chromatography system is due to the partial oxidation of a methionine residue. In each Edman degradation step three amino acids appeared, corresponding to peptides T3, T5, and T6. The three peptides were isolated in pure form after oxidizing protein L31 and separating the tryptic peptides by mapping (Figure 4b). Each of them was then separately sequenced. In the nonoxidized peptides a faint spot between Dns-Phe and Dns-Glu was detected in systems B and C. The same spots were also observed in the Cys positions of unoxidized apamine, when sequenced.

Tyr-Glu-Glu-Ile-Thr-Ala-Ser-Cya-Ser-Cya-
Gly-Asn-Val-$MeO_2$-Lys
T3

All 15 residues were determined by the dansyl-Edman method. After systems A, B, and C the Dns derivatives of the two cysteic acid residues were chromatographed in solvent E. Dns-Cya was observed migrating as a sharp line slightly faster than Dns-OH. Positions 1 and 3 of oxidized apamine gave the same result. The two glutamic acid residues and the asparagine residue were determined by thin-layer chromatography of the respective Pth-amino acids. The remaining positions except for the serine residues were confirmed by analysis of the hydrolyzed ATZ derivatives. Using this procedure Thr was detected as $\alpha$-aminobutyric acid.

Ile-Arg
T4

Ser-Thr-Val-Gly-His-Asp-Leu-Asn-Leu-Asp-
Val-Cya-Ser-Lys
T5

In addition to the complete determination of the 14 residues by dansylation the identifications (except for the serine residues) were confirmed by hydrolysis of the ATZ derivatives. The ATZ derivatives of the Asp and Asn residues were identified after conversion into their respective Pth derivatives.

Cya-His-Pro-Phe-Phe-Thr-Gly-Lys
T6

Dns-His was identified in systems F and G, and Dns-Cya was identified in system E. All positions were confirmed by hydrolysis of the ATZ derivatives and identification of the resulting free amino acids on the amino acid analyzer.

Gln-Arg
T7

The N-terminal glutamine residue was not converted to pyrrolidonecarboxylic acid during isolation of the peptide. Thus it remained accessible to Edman degradation. Glutamine was identified as Pth-Gln.

Asp-Val-Ala-Thr-Gly-Gly-Arg
T8

Peptide T8 was sequenced by dansyl-Edman degradation and the ATZ derivatives were hydrolyzed and determined by amino acid analysis. The first position was determined as Pth-Asp by thin-layer chromatography.

Val-Asp-Arg
T9

Phe-Asn-Lys
T10

Peptides T9 and T10 were sequenced by the dansyl-Edman procedure. Their second positions were determined as Pth-Asp

and Pth-Asn, respectively. T10 was identified as the C-terminal tryptic peptide of L31 since carboxypeptidase A released no amino acid, carboxypeptidase B released lysine, in agreement with the results of Alakhov et al. (1975), and carboxypeptidase C released Lys, Asn, Phe, Arg, Asp, Val, and Gly after 20-min incubation of the whole protein.

*Thermolysin Peptides.* The peptides after digestion of unoxidized protein with thermolysin were isolated by mapping (Figure 4c). Their amino acid compositions are given in Table III (supplementary material). They were completely sequenced in the same way as the tryptic peptides. TL3 and TL8 were isolated linked together, presumably connected by disulfide bridges, and were sequenced as a mixture. In the following the applied sequence methods are indicated by arrows (see footnote 1). Regions overlapping tryptic peptides are indicated.

Met-Lys-Lys-Asp
T1
TL1

The N-terminal peptide contains tryptic peptide T1 and an Asp residue which is at the N-terminal position of tryptic peptides T2 and T8.

Ile-His-Pro-Lys-Tyr-Glu-Glu
T2 T3
TL2

Peptide TL2 contains the C terminus of T2 and the N-terminal residues of T3.

Ile-Thr-Ala-Ser-Cys-Ser-Cys-Gly-Asn
T3
TL3

Leu-Asp-Val-Cys-Ser-Lys-Cys-His-Pro-Phe
T5 T6
TL8

The organic phases containing the ATZ derivatives of the Cys residues were oxidized before hydrolysis with 5.7 N HCl (Brauer, 1977), and were thus identified as cysteic acid. In step two the ATZ derivatives were converted into the respective Pth-amino acids. Pth-Asp was identified. After subtracting the sequence of one chain containing a part of tryptic peptide T3 from that of the other chain, a sequence remained which gave the alignment of tryptic peptides T5 and T6. This was later confirmed by peptic peptide P2′ (see below).

Val-Met-Lys
T3
TL4

This tripeptide (TL4) contains the C-terminal portion of tryptic peptide T3.

Ile-Arg-Ser-Thr
T4 T5
TL5

Peptide TL5 connects T4 to the N-terminal part of T5.

Val-Gly-His-Asp
TL6

Leu-Asn
TL7

Peptides TL6 and TL7 are fragments of tryptic peptide T5.

Phe-Thr-Gly-Lys-Gln-Arg-Asp
T6 T7
TL9

Peptide TL9 contains the last 4 residues of tryptic peptide T6, the dipeptide T7, and an aspartic acid residue, which could belong to either tryptic peptide T2 or T8.

Val-Ala-Thr-Gly-Gly-Arg-Val-Asp-Arg
T8 T9
TL10

This peptide (TL10) was obtained by incomplete cleavage with thermolysin and provides the bridge between tryptic peptides T8 and T9.

Val-Ala-Thr-Gly-Gly-Arg
TL10a

TL10a is a fragment of TL10.

Phe-Asn-Lys
TL11

Peptide TL11 is identical with the C-terminal tryptic peptide T10.

*Chymotryptic Peptides.* The chymotryptic peptides (Table III, Figure 4d; see supplementary material) gave the necessary information for further alignment of peptides in protein L31. The enzyme used showed a considerable degree of tryptic activity (see peptides CH2 and CH4a–d, Figure 5).

Met-Lys-Lys-Asp-Ile-His-Pro-Lys-Tyr-Glu-Glu-
T1 T2 T3
CH1

Ser-Thr-Val-Gly-His-Asp-Leu-Asn-Leu-Asp-Val-
T5
CH3

Ile-Thr-Ala-Ser-Cys-Ser-Cys-Gly-Asn-Val-Met
CH1

Cys-Ser-Lys-Cys-His-Pro-Phe
T6
CH3

Peptides CH1 and CH3 are presumably linked by disulfide bridges and occurred in the same spots in fingerprints. They were sequenced together up to step 18 by the dansyl-Edman method. When the sequence of tryptic peptide T5 and the first four residues of T6 were subtracted the alignment of tryptic peptides T1–T3 was obtained. This was later confirmed by sequence analysis of peptic peptide P1′ (see below).

Lys-Ile-Arg
T3 T4
CH2

Peptide CH2 provided the bridge between tryptic peptides T3 and T4. Peptide T3 is the only lysine-containing tryptic peptide which has a cleavage site for chymotrypsin adjacent to the terminal lysine. Another, more remote possibility is that the residual tryptic activity of the chymotrypsin could cleave between the two lysines of T1. This seems very unlikely since (1) trypsin is apparently unable to cleave in this position (see above) and (2) peptide T1 has already been shown to be ad-

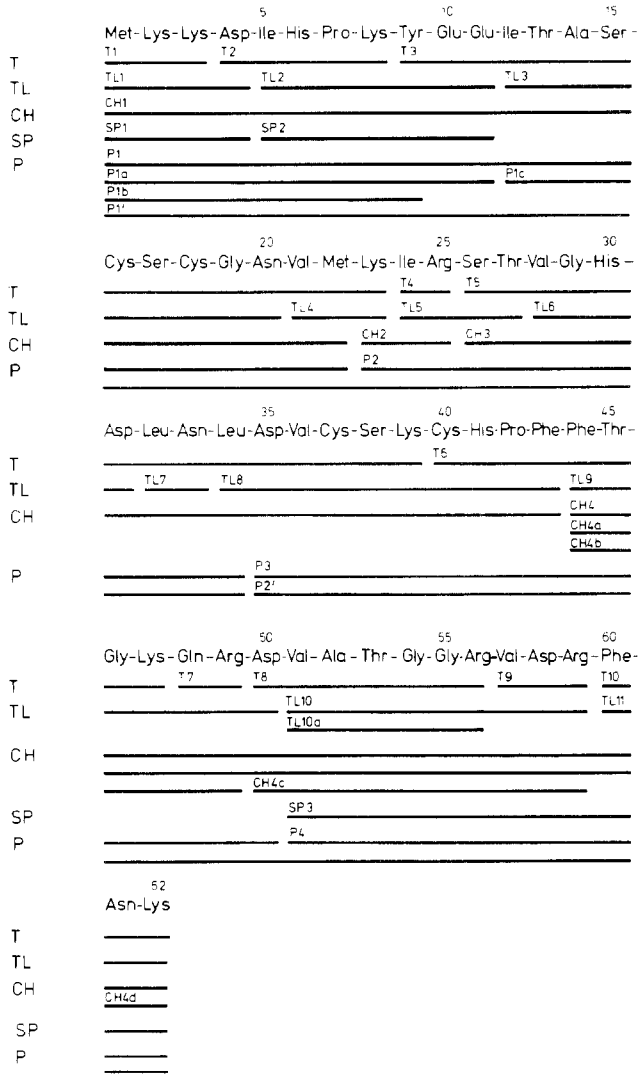
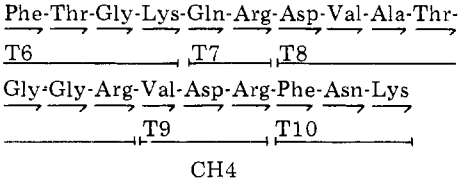



FIGURE 5: Complete amino acid sequence of ribosomal protein L31: T, tryptic peptides; TL, thermolysin peptides; CH, chymotryptic peptides; SP, *Staphylococcus aureus* protease peptides; P, peptic peptides of unoxidized protein; P′, peptic peptides of oxidized protein.

jacent to peptide T2 from the overlapping sequences provided by peptides CH1 and P1′.




CH4

Peptide CH4 was completely sequenced by the dansyl-Edman method without hydrolyzing the ATZ derivatives or converting them into Pth-amino acids. CH4 provides the alignment of tryptic peptides T6–T10. The fragments CH4a–d obtained from residual tryptic activity of the chymotrypsin confirmed the sequence. Peptides CH4a (positions 44–60) and CH4c (positions 50–59) were characterized by amino acid analysis while peptides CH4b (positions 44–49) and CH4d (positions 61–62) were sequenced by the dansyl-Edman procedure.

*Peptides Resulting from Digestion by Staphylococcus aureus Protease.* After digestion of protein L31 with *S. aureus* protease three peptides were isolated in significant amount (Figure 4e; supplementary material). From their amino acid composition (Table III; supplementary material) their positions in the protein chain were clearly identified. Peptides SP1 (residues 1–4) and SP2 (residues 5–11) were identical with peptides TL1 and TL2, respectively. Peptide SP3 (positions 51–62) contains peptides TL10 and TL11 (see Figure 5). The absence of an additional Asp or Glu residue in the amino acid composition of SP3 confirms the assignment of TL10 and TL11 to the C terminus of the protein.

*Peptic Peptides.* Peptic peptides (Figure 4d; supplementary material) were utilized to confirm overlaps previously established including these of TL3 and TL8 or CH1 and CH3, respectively, which were sequenced simultaneously as they were presumably linked by disulfide bridges. The amino acid compositions of the peptic peptides were in good agreement with the results obtained from other peptides (Table III; supplementary material). P1 (positions 1–22) and P1c (positions 12–22) are linked to peptide P3 (positions 35–50), presumably by disulfide bridges. Minor cleavages resulted also in peptides P1a (positions 1–11) and P1b (positions 1–9). Peptide P2 confirmed the alignment of tryptic peptides T3–T5. Peptide P4 is identical with SP3 and covers the C-terminal region of the protein. Digestion of the oxidized protein resulted only in cleavage after leucine in position 34. The two halves were isolated by peptide mapping and sequenced by the dansyl-Edman method without conversion or hydrolysis of the ATZ derivatives from the organic phases. P1′ was sequenced up to position 15. In order to save material steps 7–10 were not dansylated as this region did not provide any further overlaps. Thus, the alignment of peptides T1–T3 was once more confirmed. Finally peptide P2′ was sequenced through position 11, confirming once more the alignment of tryptic peptides T5 and T6.

*The Complete Sequence of Protein L31.* The primary structure of L31 is summarized in Figure 5. The entire sequence was established and confirmed using about 700 μg of protein: 665 μg for enzymatic digestion, and 35 μg for two-dimensional gel electrophoresis, amino acid composition, and N-terminal analysis of the intact protein.

## Discussion

Protein L31 has a single chain of 62 amino acids (Figure 5) giving a calculated molecular weight of 6967, in good agreement with the sodium dodecyl sulfate gel molecular weight of 7100 (Alakhov et al., 1975). The amino acid sequence is in close agreement with the measured amino acid composition (Table II).

Determination of the total sequence of protein L31 with a limited amount of material (700 μg ≅ 100 nmol) was facilitated mainly by the incorporation of two refinements of the methodology: (1) isolation of peptides from enzymatic digests exclusively from thin-layer peptide maps and (2) location of peptide spots by staining with fluorescamine, rather than ninhydrin. Peptides were well resolved and obtained in good yield from the thin-layer plates. Use of fluorescamine avoids destruction of N-terminal amino acids and lysine residues, affording higher yields of intact peptides. The relative merits of the ninhydrin and fluorescamine methods for visualizing peptide spots have been discussed (Brosius, 1976; Schiltz et al., 1977). Thus it was possible to sequence through every position of the protein chain from two to four times with peptides derived from different enzymatic digests.

In a search for homology with known sequences from other ribosomal proteins, only two tetrapeptide sequences were found to be identical, and are probably not statistically significant. This result is in accord with previous findings that little or no

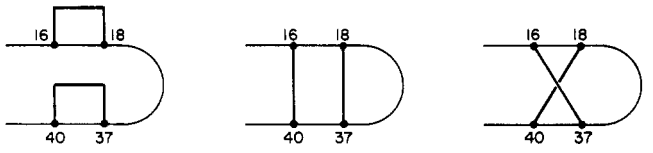



FIGURE 6: Schematic representation of the possible configurations of the sulfur bridges in L31. Evidence presented rules out arrangement (a), but does not distinguish between the remaining possibilities (b and c).

sequence homology exists between different ribosomal proteins (Wittmann-Liebold and Dzionara, 1976).

Secondary structure calculations, as described by Burgess et al. (1974), Chou and Fasman (1974a,b), Chou et al. (1975), Nagano (1977), and Robson and Suzuki (1976) have been applied to the sequence of protein L31. There is general agreement between the various methods that no $\alpha$ helix is present in the L31 structure. The methods of Burgess et al. (1974) and Chou et al. (1975) suggest that some $\beta$ structure may exist in positions 15–22 and 33–37. Detailed results of these calculations are described by Dzionara et al. (1978).

The presence of a small molecular weight protein from the 50S subunit with three half-cystine-containing tryptic peptides was previously reported by Acharya and Moore (1973) and Moore (1975), who were unable to correlate their findings with the properties of any of the known ribosomal proteins. Comparison of their findings with my results strongly suggests that their component "Y" is in fact protein L31.

There is considerable evidence that the four half-cystine residues in protein L31, as isolated, are linked by disulfide bridges. In every case where peptide maps were carried out on a digest of protein that had not been subjected to performic acid oxidation, all of the half-cystine-containing peptides were found in a single spot. Thus, in tryptic digests of nonoxidized protein, peptides T3, T5, and T6 were all found in the same two spots, which differed presumably due to the different oxidation state of the methionine residue in T3; in chymotryptic digests, both peptides CH1 and CH3 were found in each of three spots, differing presumably by the different oxidation states of the two methionines in peptide CH1; and finally, peptic peptides P1 and P3, and P1c and P3, respectively, were found to comigrate. Yet, in digests of performic acid oxidized protein, all of the cysteic acid containing peptides migrated independently of each other.

There are three possible arrangements of two pairs of disulfide bridges between four half-cystine residues (Figure 6). Arrangement a can be ruled out, since it would predict that peptide T3 would migrate independently of T5 and T6, CH1 would be unlinked to CH3, and P1 would be unlinked to P3. However, the evidence does not distinguish between arrangements b and c. This is mainly due to the very close positioning of half-cystine residues 16 and 18, making it difficult to cleave specifically between them with available methods. The biological significance of these disulfide bridges is not clear. It is not possible to estimate from these studies whether the oxidation state or disulfide arrangement of protein L31 as isolated is the same as in the intact ribosome. However, Acharya and Moore (1973) have reported that the number of half-cystines is equal to the number of titratable thiols in the denatured 50S subunit, and conclude that there are no disulfide bridges in the intact 50S ribosomal subunit. Thus, it is probable that these disulfide bonds arose during the isolation of L31. It is interesting, however, that disulfide bridge formation is complete and apparently specific, suggesting that this protein has the capability of forming specific disulfide bridges.

Acknowledgment

I would like to thank Mrs. Ursula Arfsten for excellent technical assistance, Dr. Garth Cumberlidge for mini two-dimensional polyacrylamide gel electrophoresis and limericks, Dr. K. Katze for discussions, Dr. B. Wittmann-Liebold for providing computer data including secondary structure predictions and homologies, and Drs. H. G. Wittmann, Roger Garrett, and Harry Noller for critically reading the manuscript.

Supplementary Material Available

Table III giving the amino acid compositions of all peptides, Figure 1 showing the two-dimensional polyacrylamide gel electrophoresis of the pH 3.8 protein fraction, Figure 3 showing the two-dimensional polyacrylamide gel electrophoresis of purified L31, and figures of several peptide maps (Figures 4d-f) as indicated in the text (6 pages). Ordering information is given on any current masthead page.

References

Acharya, A. S., and Moore, P. B. (1973), *J. Mol. Biol. 76,* 207–221.

Alakhov, Yu. B., Mednikova, T. A., Momuz, L. P., Markova, L. F., Dovgas, N. V., Kashparov, I. V., and Ovchinnikov, Yu. A. (1975), *Bioorg. Khim. 1,* 581–587.

Brauer, D. (1977), Dissertation, Technische Universität, Berlin.

Brimacombe, R., Nierhaus, K. H., Garrett, R. A., and Wittmann, H. G. (1976), *Prog. Nucleic Acid Res. Mol. Biol. 18,* 1–44.

Brosius, J. (1976), Dissertation, Johann Wolfgang Goethe-Universität, Frankfurt am Main.

Brosius, J., and Arfsten, U. (1978), *Biochemistry* (following paper in this issue).

Brosius, J., and Chen, R. (1976), *FEBS Lett. 68,* 105–109.

Brosius, J., Schiltz, E., and Chen, R. (1975), *FEBS Lett. 56,* 359–361.

Bruton, C. J., and Hartley, B. S. (1970), *J. Mol. Biol. 52,* 165–178.

Burgess, A. W., Ponnuswamy, P. K., and Scheraga, A. h. (1974), *Isr. J. Chem. 12,* 239–286.

Chen, R., Brosius, J., Wittmann-Liebold, B., and Schäfer, W. (1977), *J. Mol. Biol. 111,* 173–181.

Chou, P. Y., and Fasman, G. D. (1974), *Biochemistry 13,* 211–222.

Chou, P. Y., and Fasman, G. D. (1974), *Biochemistry 13,* 222–245.

Chou, P. Y., Adler, A. J., and Fasman, G. D. (1975), *J. Mol. Biol. 96,* 29–45.

Dzionara, M., Robinson, S. M. L., and Wittmann-Liebold, B. (1978), *J. Supramol. Struct.,* submitted for publication.

Edman, P., and Henschen, A. (1975), in Protein Sequence Determination, Needleman, S. B., Ed., New York, N.Y., Springer-Verlag, pp 232–279.

Gray, W. R., and Hartley, B. S. (1963), *Biochem. J. 89,* 379–380.

Hartley, B. S. (1970), *Biochem. J. 119,* 805–816.

Hindennach, I., Stöffler, G., and Wittmann, H. G. (1971), *Eur. J. Biochem. 23,* 7–11.

Hirs, C. H. W. (1956), *J. Biol. Chem. 219,* 611–621.

Houmard, J., and Drapeau, G. R. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 3506–3509.

Inagami, T., and Murakami, K. (1972), *Anal. Biochem. 47,* 501–504.

Isono, K. (1974), *Mol. Gen. Genet. 133,* 77-86.
Kaltschmidt, E., and Wittmann, H. G. (1970), *Proc. Natl. Acad. Sci. U.S.A. 67,* 1276-1282.
Moore, P. B. (1975), *J. Mol. Biol. 98,* 439-444.
Nagano, K. (1977), *J. Mol. Biol. 109,* 251-274.
Robson, B., and Suzuki, E. (1976), *J. Mol. Biol. 107,* 327-356.
Roseau, G., and Pantel, P. (1969), *J. Chromatogr. 44,* 392-395.
Sanger, F., and Thompson, E. O. P. (1953), *Biochem. J. 53,* 366-374.
Schiltz, E., Schnackerz, K. D., and Gracy, W. R. (1977), *Anal. Biochem. 79,* 33-42.
Schwabe, E. (1972), *Hoppe-Seyler's Z. Physiol. Chem. 353,* 1899-1906.
Smithies, O., Gibson, D. M., Fanning, E. M., Goodfliesh, R. M., Gillman, J. G., and Ballantyne, G. L. (1971), *Biochemistry 10,* 4912-4921.
Tschesche, H., and Kupfer, S. (1972), *Eur. J. Biochem. 26,* 33-36.
Vandekerckhove, J., and Van Montagu, M. (1974), *Eur. J. Biochem. 44,* 279-288.
Weber, H. J. (1972), *Mol. Gen. Genet. 119,* 233-248.
Weiner, A. M., Platt, T., and Weber, K. (1972), *J. Biol. Chem. 247,* 3242-3251.
Wittmann-Liebold, B., and Dzionara, M. (1976), *FEBS Lett. 65,* 281-283.
Yaguchi, M., Wittmann, H. G., Cabézon, T., DeWilde, M., Villarroel, R., Herzog, A., and Bollen, A. (1975), *Mol. Gen. Genet. 142,* 35-43.

# Primary Structure of Protein L19 from the Large Subunit of *Escherichia coli* Ribosomes†

Jürgen Brosius*,‡ and Ursula Arfsten

ABSTRACT: Protein L19, a component of the *Escherichia coli* 50S ribosomal subunit implicated in 30S-50S subunit interaction, was sequenced by the dansyl-Edman method. L19 consists of a single polypeptide chain of 114 amino acids giving a calculated molecular weight of 13 002. Peptides obtained from various enzymatic cleavages were isolated on thin-layer peptide maps or by gel filtration. Automated Edman degradation using a liquid phase sequenator was carried out on the whole protein as well as on a large 58-residue fragment arising from digestion with *Staphylococcus aureus* protease. Every position in protein L19 was confirmed at least twice. Results of secondary structure estimation and homologies with other *E. coli* ribosomal protein sequences are presented.

Protein L19 is a component of the structure of the 50S ribosomal subunit, and has been implicated in the 30S-50S subunit association (Morrison et al., 1973; Tischendorf et al., 1974). It is accessible to antibody in the free 50S subunit (Morrison et al., 1977) but not in the undissociated 70S particle. Thus, it is one of the 50S proteins that is shielded by the 30S subunit against antibodies (Zeichhardt, 1976). Variable results have been reported for the binding of protein L19 to ribosomal RNA (Garrett et al., 1974). When chloramphenicol bound to 50S subunits is irradiated with ultraviolet (UV) light, the antibiotic is found to become covalently attached to protein L19 (Sonenberg et al., 1976). Protein L19 has been shown to have an antigenic determinant near that of L16 (Tischendorf et al., 1975), a protein which has also been implicated in the binding of chloramphenicol (Nierhaus and Nierhaus, 1973; Pongs et al., 1973). These findings suggest that L19 is located at the 30S-50S interface, and that it may play a role in the structure and function of the ribosomal A site. Knowledge of its primary structure may thus lead to further insights into the mechanism of protein synthesis at the molecular level.

## Materials and Methods

If not stated otherwise chemicals were analytical grade purchased from Merck.

*Isolation of Protein L19.* Protein L19 was isolated according to Hindennach et al. (1971) and provided by Dr. H. G. Wittmann. Its identity and purity were checked by two-dimensional polyacrylamide gel electrophoresis (Kaltschmidt and Wittmann, 1970).

*Amino Acid Analysis.* Unoxidized and oxidized protein (10 μg each) were subjected to amino acid analysis as described in the preceding paper (Brosius, 1978).

Tryptophan-containing peptides were located by spraying tryptic peptide maps with 1% dimethylbenzaldehyde (Riedel de Haen, Hanover) in 2 N HCl (Spies and Chambers, 1949) or by fluorescence under long-wavelength UV light, as described in the preceding paper (Brosius, 1978). Furthermore, Trp can be detected as its Pth[1] derivative by silica gel thin-layer chromatography (Brosius, 1978).

The N-terminal amino acid was determined by dansylation of 2 μg of the intact protein.

*Enzymatic Cleavage.* Methods used for enzymatic cleavage of protein L19 are summarized in Table I. Blocking of the lysine residues with citraconic acid anhydride prior to tryptic

† From the Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, West Germany. *Received September 8, 1977.*

‡ Present address: Thimann Laboratories, University of California, Santa Cruz, Calif. 95064.

[1] The same abbreviations and symbols are used as in the preceding paper (Brosius, 1978). In addition, the following were used: Tos-Lys$CH_2Cl$, *N*-α-*p*-tosyl-L-lysylchloromethane; BT, peptides resulting from digestion with trypsin after blocking the lysine residues with citraconic acid anhydride; CH, peptides resulting from digestion with Tos-Lys$CH_2Cl$-chymotrypsin; CH′, peptides resulting from digestion with α-chymotrypsin; SP, peptides resulting from digestion with *Staphylococcus aureus* protease at pH 4.0; SP′, peptides resulting from digestion with *Staphylococcus aureus* protease at pH 7.8; —, solid phase Edman degradation.

0006-2960/78/0417-0508$01.00/0 