THE PRIMARY STRUCTURE OF THE 5 S RNA BINDING PROTEIN L5 OF *ESCHERICHIA COLI* RIBOSOMES

Robert CHEN* and Gisela EHRKE

Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, Germany

Received 11 August 1976

1. Introduction

Protein L5 is the largest of three proteins (L5, L18 and L25) that form a stable complex with 5 S RNA within the 50 S subunit of *E. coli* ribosomes [1–4]. In addition these proteins mediate the attachment of 5 S RNA to the 23 S RNA and thus integrate the 5S RNA into the 50 S particle [2]. The 5 S RNA–protein complex has been found to be active in the hydrolysis of GTP and ATP [5] and to be involved in the binding of aminoacyl-tRNA to the ribosome (see ref. [6] for a recent review). Crosslinking of L7/L12 to L5 [7] revealed that the 5 S RNA–protein complex is in close proximity to the binding site of initiation, elongation and termination factors to the ribosome. Furthermore, protein L5 is involved in the EF-G dependent binding of GDP [8].

The primary structures of 5 S RNA [9] and of proteins L18 [10] and L25 [11,12] have been determined recently. The elucidation of the amino acid sequence of protein L5 reported in this paper will further facilitate the investigation of the function and the three-dimensional structure of the 5 S RNA–protein complex.

2. Materials and methods

Protein L5 was isolated from a CM-cellulose column as described [13] and provided by Dr H. G. Wittmann.

*To whom correspondence should be addressed.

2.1. *Enzyme digestions*

The protein was digested with trypsin, chymotrypsin and thermolysin under conventional conditions. The hydrolysis of L5 with *Staphylococcus aureus* protease was in 5 nmol NH_4HCO_3, pH 7.8, for 16 h at 37°C. A partial digestion of L5 was performed on a protein sample which had been reacted at its lysine residues with citraconic anhydride using double distilled water adjusted to pH 7.5 as solvent. The protein was then desalted on a Sephadex G-25 column in 0.2% NH_4HCO_3 before it was submitted to the tryptic digestion.

2.2. *Isolation of peptides*

Ten nmol of both the hydrolysate of the *Staphylococcus aureus* protease digestion and the hydrolysate of the partial tryptic digestion were fingerprinted in order to obtain a general view of the number, size and charge of the peptides. Most of the digested protein, however, was prefractionated on a Sephadex G-50 SF column (150 × 1 cm) before it was fingerprinted. Fractions of 500 μl were collected. The peptides were detected by micro N-terminus dansylation or by the ninhydrin reaction after aliquots of all fractions had been spotted onto cellulose thin-layer plates.

Preparative fingerprints were developed on purified cellulose thin-layer plates (Polygram cell 400 or cell 300, Macherey and Nagel, Düren) [14]. The positions of the peptides were identified with ninhydrin (0.1% solution) or in some cases with fluorescamine (Roche, Basel). Peptides used for N-terminus and sequence determinations were extracted from the cellulose twice with 200 μl of 50% acetic acid. For amino acid analyses the extractions were carried out in 6 N HCl

North-Holland Publishing Company – Amsterdam

and 0.02% β-mercaptoethanol. The complete hydrolyses of the peptides were performed in the same buffer for 20 h at 110°C.

2.3. *Amino acid analyses*

All amino acid analyses were run on a Durrum D500 amino acid analyzer (Palo Alto, California) at a sensitivity range of 1–2 nmol. Cysteine was recovered as cysteic acid after oxidation with performic acid [15]. Tryptophan was first identified by its fluorescence on the peptide map when observed under ultraviolet light at 366 nm. In addition it was identified by reacting it with Ehrlich's reagent [16] which is specific for tryptophan.

2.4. *Sequence determination*

The sequence determinations of all peptides were exclusively done by an improved micro dansyl-Edman procedure [17–19]. Before the degradation was started the peptides were transferred into 0.4 × 5.0 cm glass tubes and dried twice in a desiccator under vacuum in order to remove traces of acetic acid as described previously [18]. The dansyl amino acids were identified according to their R_f-values on micro polyamide plates (F 1700, Schleicher and Schüll, Dassel).

2.5. *Assignment of amides*

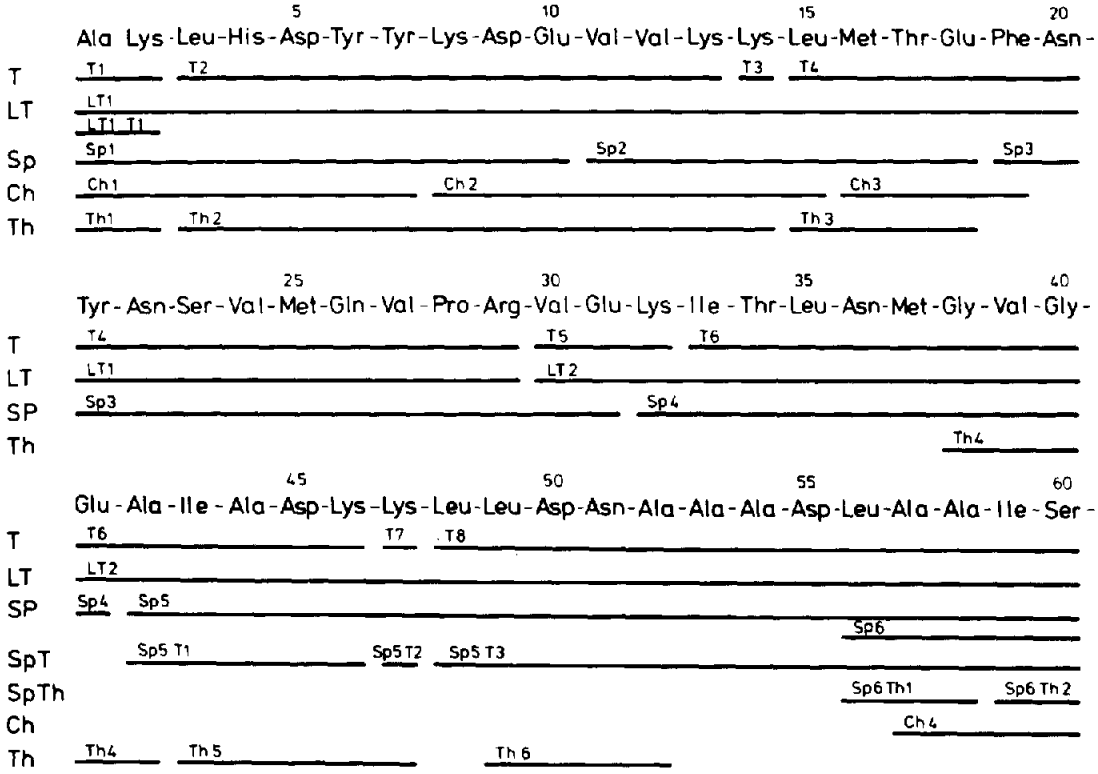
All *n*-butyl acetate extracts from the micro dansyl-Edman degradation in which the N-terminus had been determined as Glx or Asx were collected and converted to the PTH amino acids and identified as such on silica gel thin layer plates by their colour reaction with ninhydrin [18].

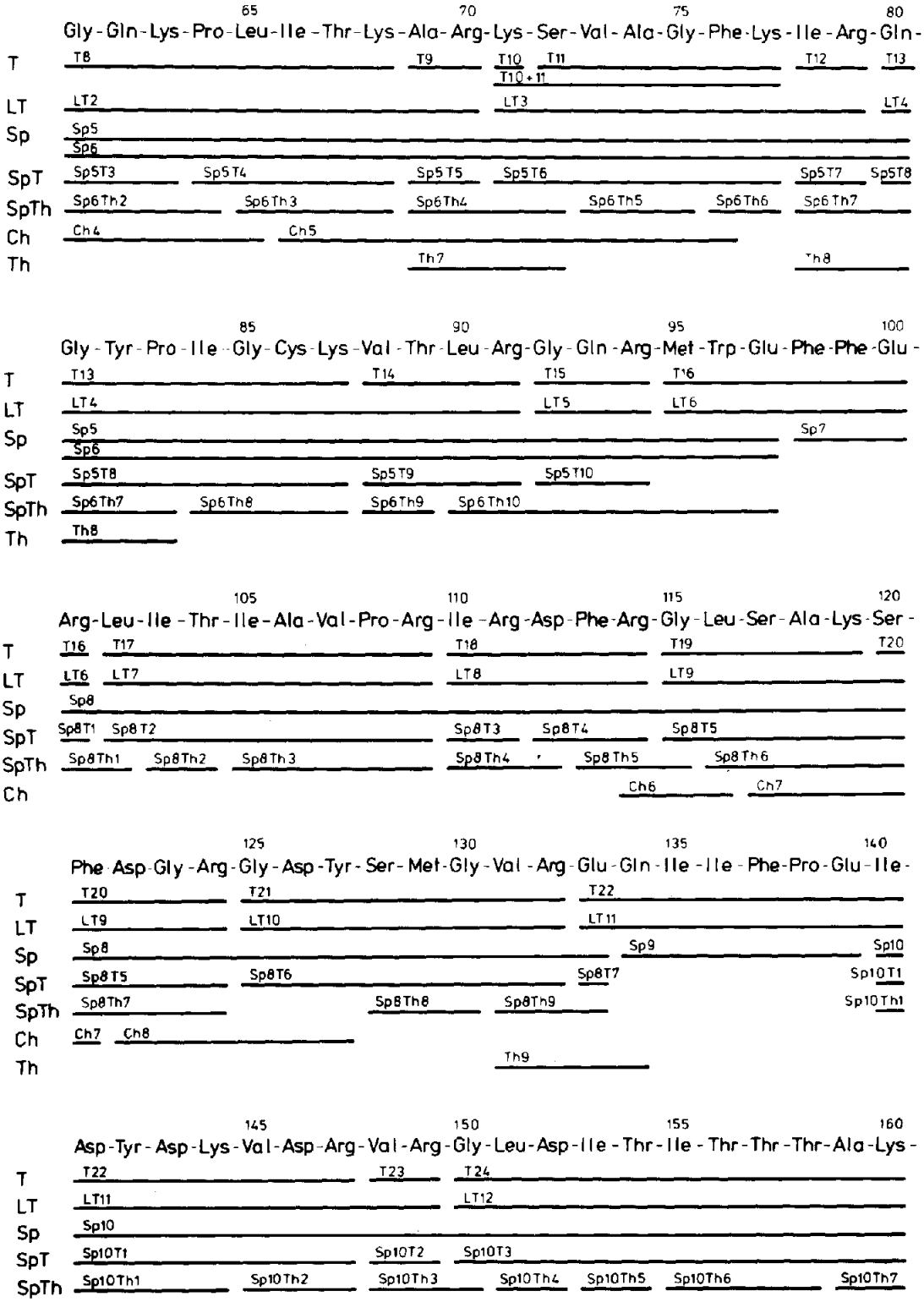
2.6. *Identification of the C-terminus*

For the determination of the C-terminal amino acid sequence of L5 the protein was digested with carboxypeptidase A as well as with carboxypeptidase A+B. The released amino acids were analysed on a Durrum analyzer.

3. Results and discussion

The complete amino acid sequence of protein L5 is presented in fig.1. The protein has a molecular weight




65 70 75 80
Gly-Gln-Lys-Pro-Leu-Ile-Thr-Lys-Ala-Arg-Lys-Ser-Val-Ala-Gly-Phe-Lys-Ile-Arg-Gln-
T T8 T9 T10 T11 T12 T13
T10+11
LT LT2 LT3 LT4
Sp Sp5
Sp6
SpT Sp5T3 Sp5T4 Sp5T5 Sp5T6 Sp5T7 Sp5T8
SpTh Sp6Th2 Sp6Th3 Sp6Th4 Sp6Th5 Sp6Th6 Sp6Th7
Ch Ch4 Ch5
Th Th7 Th8
85 90 95 100
Gly-Tyr-Pro-Ile-Gly-Cys-Lys-Val-Thr-Leu-Arg-Gly-Gln-Arg-Met-Trp-Glu-Phe-Phe-Glu-
T T13 T14 T15 T16
LT LT4 LT5 LT6
Sp Sp5 Sp7
Sp6
SpT Sp5T8 Sp5T9 Sp5T10
SpTh Sp6Th7 Sp6Th8 Sp6Th9 Sp6Th10
Th Th8
105 110 115 120
Arg-Leu-Ile-Thr-Ile-Ala-Val-Pro-Arg-Ile-Arg-Asp-Phe-Arg-Gly-Leu-Ser-Ala-Lys-Ser-
T T16 T17 T18 T19 T20
LT LT6 LT7 LT8 LT9
Sp Sp8
SpT Sp8T1 Sp8T2 Sp8T3 Sp8T4 Sp8T5
SpTh Sp8Th1 Sp8Th2 Sp8Th3 Sp8Th4 Sp8Th5 Sp8Th6
Ch Ch6 Ch7
125 130 135 140
Phe-Asp-Gly-Arg-Gly-Asp-Tyr-Ser-Met-Gly-Val-Arg-Glu-Gln-Ile-Ile-Phe-Pro-Glu-Ile-
T T20 T21 T22
LT LT9 LT10 LT11
Sp Sp8 Sp9 Sp10
SpT Sp8T5 Sp8T6 Sp8T7 Sp10T1
SpTh Sp8Th7 Sp8Th8 Sp8Th9 Sp10Th1
Ch Ch7 Ch8
Th Th9
145 150 155 160
Asp-Tyr-Asp-Lys-Val-Asp-Arg-Val-Arg-Gly-Leu-Asp-Ile-Thr-Ile-Thr-Thr-Thr-Ala-Lys-
T T22 T23 T24
LT LT11 LT12
Sp Sp10
SpT Sp10T1 Sp10T2 Sp10T3
SpTh Sp10Th1 Sp10Th2 Sp10Th3 Sp10Th4 Sp10Th5 Sp10Th6 Sp10Th7

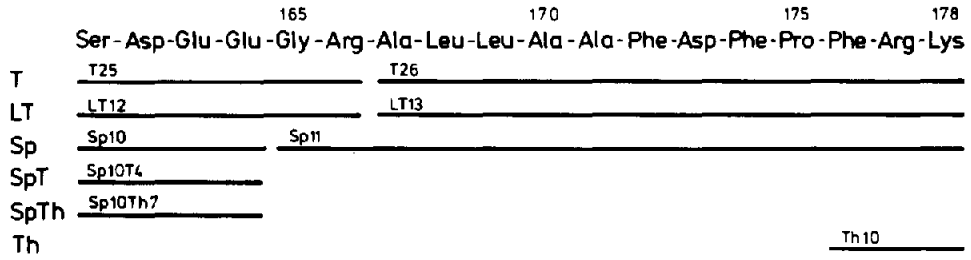


Fig.1. Amino acid sequence of ribosomal protein L5. T: tryptic peptides. LT: peptides obtained from a partial tryptic digestion of the protein that had been blocked at its lysine residues. Sp: peptides obtained from the *Staphylococcus aureus* protease digestion. SpT: tryptic peptides from a second digestion of Sp peptides. SpTh: thermolytic peptides from a second digestion of Sp peptides. Ch: chymotryptic peptides. Th: thermolytic peptides.

of 20172 daltons and consists of 178 amino acids residues. The amino acid composition derived from the sequence is as follows: Asp_{14}, Asn_4, Thr_9, Ser_7, Glu_{10}, Gln_5, Pro_6, Gly_{13}, Ala_{16}, Cys_1, Val_{12}, Met_5, Ile_{14}, Leu_{13}, Tyr_6, Phe_{10}, His_1, Lys_{16}, Arg_{15}, Trp_1. It is in excellent agreement with the amino acid analysis of the hyrolysed protein.

One tryptophan was identified on the tryptic peptide map when observed under ultraviolet light or sprayed with Ehrlich's reagent. The peptide containing tryptophan was characterized as tryptic peptide T16. Sequencing the peptide released tryptophan in step two, i.e. in position 96 of the protein. The amino acid analysis of the oxidized protein provided one cysteine. This result was substantiated by the sequence and is in agreement with the report of Moore [20] who analyzed the cysteine content of ^{35}S-labelled ribosomal proteins. From the tryptic fingerprint 19 peptides (T1, T2, T5, T8, T9, T11, T12, T13, T14, T15, T17, T18, T19, T20, T21, T22, T23, T25 and T26) were isolated and sequenced completely. Peptides T3, T7 and T10 consist of lysine residues only. The missing peptides T4, T6, T16, and T24 have not been isolated directly from the tryptic digest. However, their sequences were obtained from other peptides.

From the partial tryptic digest of L5 the lysine-blocked peptides LT1 and LT2 were isolated in pure form from the Sephadex G-50 column. The remaining LT-peptides were further purified by the two-dimensional separation technique on cellulose thin layer plates [21–23]. Peptide LT1 was sequenced up to residue 24. This gave the alignment of peptides T1-T2-T3-T4. Peptide T4 was isolated from Peptide LT1 by a second digestion with trypsin. Peptide LT2 (positions 30–70) could be completely sequenced by by our refined micro dansyl-Edman technique [18–19]. In this way the amino acid sequence of peptide T6 and the order of peptides T5-T6-T7-T8-T9 were established. The alignment of T10-T11-T12 resulted from the sequence of LT3. The sequence of LT4 gave the order: T13-T14. The amino acid sequence of T16 was determined by the sequence analysis of peptide LT6 since both peptides are identical. The study of peptide LT9 gave evidence for the alignment of T19-T20 and that of peptide LT11 resulted in the order: T22-T23. From the complete sequence of LT12 the sequence of tryptic peptide T24 and the order of peptides T24-T25 were obtained. Peptide LT13 was identical with peptide T26. Both peptides contained the C-terminus of protein L5.

Digestion of protein L5 with *Staphylococcus aureus* protease released 11 peptides. From these peptides seven (Sp1, Sp2, Sp3, Sp4, Sp7, Sp9 and Sp11) were sequenced. The results from Sp2 and Sp3 gave the sequence of tryptic peptide T4. In addition, the overlap of peptides LT1-LT2 was provided by Sp3. The alignments of LT2-LT3, LT3-LT4 and LT4-LT5-LT6 were determined from several thermolytic peptides of the peptide Sp6. The order of peptides LT6-LT7-LT8-LT9-LT10 was provided by peptide Sp8 which has been degraded manually up to its 26 residue. One peptide (Sp8Th9) obtained from a second digestion of Sp8 with thermolysin revealed that peptide LT10 was positioned before LT11. The sequence of the first 20 residues of peptide Sp10 showed that peptide LT11 is located before LT12. The sequence of peptide Sp11 ensured the alignment of LT12-LT13.

The primary structure of protein L5 was thus

Table 1
Prediction of α-helical and β-sheet conformations in L5 by the methods of Chou and Fasman [28,29] and Burgess et al. [27].
The numbers indicate the residues involved in the secondary structure.

method	α-helix					
Chou and Fasman	8–20	41–57		93–103	163–174	
Burgess et al.	10–16	41–59	66–71	95–103	166–170	
	β-sheet					
Chou and Fasman	23–27	33–37	65–70	76–92	98–107	150–159
Burgess et al.		34–37		86–90		155–158

determined completely. It is the largest protein of the 50 S subunit of *E. coli* ribosomes that has been sequenced so far. Our results agree very well with the 40 N-terminal amino acids that have been sequenced with an improved Beckman sequenator [24].

Recently we reported the occurrence of methylated α-amino groups in some *E. coli* ribosomal protein (25). Although there are reports about a low degree of methylation in L5 [26,27], we have not identified any unusual amino acid during our studies.

The probable secondary structure of protein L5 has been calculated by the computerized programs of Burgess et al. [28] and Chou and Fasman [29,30]. The results from both computations agree very well in the prediction of α-helices as well as β-sheet conformations (table 1). This coincidence suggests that there may indeed exist α-helices and β-sheets in the respective parts of the protein.

In a systematic search for similarities among the amino acid sequences of ribosomal proteins of *E. coli* one pentapeptide (Val-Thr-Leu-Arg-Gly) was found to occur in protein L5 and in proteins S9,S11 and L28. A small number of tetrapeptides of L5 are also present in duplicate in several other ribosomal proteins.

Acknowledgements

We would like to thank Dr H. G. Wittmann for his interest and support and Dr B. Wittmann-Liebold for provision of computer data including homologies and secondary structure.

References

[1] Horne, J. R. and Erdmann, V. A. (1972) Molec. gen. Genet. 119, 337–344.
[2] Gray, P. N., Garrett, R. A., Stöffler, G. and Monier, R. (1972) Eur. J. Biochem. 28, 412–421.
[3] Yu, R. S. T. and Wittmann, H. G. (1973) Biochim. Biophys. Acta 324, 375–385.
[4] Gray, P. N., Bellemare, G., Monier, R., Garrett, R. A. and Stöffler, G. (1973) J. Mol. Biol. 77, 133–152.
[5] Gaunt-Klöpfer, M. and Erdmann, V. A. (1975) Biochim. Biophys. Acta 390, 226–230.
[6] Erdmann, V. A. (1976) Progr. Nucleic Acid Res. Mol. Biol., in press.
[7] Kenny, J. W., Sommer, A. and Traut, R. R. (1975) J. Biol. Chem. 250, 9434–9436.
[8] Maassen, J. A. and Möller, W. (1974) Proc. Natl. Acad. Sci. USA 71, 1277–1280.
[9] Brownlee, G. G., Sanger, F. and Barrell, B. G. (1968) J. Mol. Biol. 34, 379–412.
[10] Brosius, J., Schiltz, E. and Chen, R. (1975) FEBS Lett. 56, 359–361.
[11] Dovgas, N. V., Markova, L. F., Mednikova, T. A., Vinokurow, L. M., Alakhov, Y. B. and Ovchinnikov, Y. A (1975) FEBS Lett. 53, 351–354.
[12] Bitar, K. G. and Wittmann-Liebold, B. (1975) Z. Physiol. Chem. 356, 1343–1352.
[13] Hindennach, I., Kaltschmidt, E. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 12–16.
[14] Chen, R., Mende, L. and Arfsten, U. (1975) FEBS Lett. 59, 96–98.
[15] Hirs, C. H. W. (1956) J. Biol. Chem. 219, 611–621.
[16] Smith, M. J. H., Taylor, K. W. (1953) Biochem. J. 55, 000–000.
[17] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[18] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.
[19] Chen, R., Arfsten, U., Brosius, J., Ehrke, G. and Mende, L (1976) 10th Internat. Congr. Biochem., Abstracts.
[20] Moore, P. B. (1975) J. Mol. Biol. 98, 439–444.

[21] Chen, R. and Wittmann-Liebold, B. (1975) FEBS Lett. 52, 139–140.
[22] Chen, R. and Ehrke, G. (1976) FEBS Lett. 215–219.
[23] Brosius, J. and Chen, R. (1976) FEBS Lett., in press.
[24] Wittmann-Liebold, B., Geissler, A. W. and Marzinzig, E. (1975) J. Supramolecular Structure 3, 426–447.
[25] Chen, R., Brosius, J., Wittmann-Liebold, B. and Schäfer, W. (1976) J. Mol. Biol., manuscript submitted.
[26] Alix, J.-H. and Hayes, D. (1974) J. Mol. Biol. 86, 139–159.
[27] Chang, C. N. and Chang, F. N. (1975) Biochemistry 14, 468–477.
[28] Burgess, A. W., Ponnuswamy, P. K. and Scheraga, H. A. (1974) Israel J. Chem. 12, 239.
[29] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 211–222.
[30] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 222–245.