PRIMARY STRUCTURE OF PROTEIN L23 FROM THE *ESCHERICHIA COLI* RIBOSOME

Brigitte WITTMANN-LIEBOLD and Barbara GREUER
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, 1000 Berlin 33 (Dahlem), Germany

Received 2 October 1979

1. Introduction

Protein L23 is, according to electron microscopical studies, located within the 'seat region' of the 50 S subunit of the *Escherichia coli* ribosome [1]. Antibodies against this protein strongly inhibit the association of 30 S and 50 S subunits indicating that L23 is at, or near, the subunit interface (reviewed in [2]). The elongation factor EF-Tu can be crosslinked to protein L23 [3].

The binding site of L23 on the 23 S RNA has been shown by nuclease protection studies on single protein–RNA complexes to be in the vicinity of nucleotide 1600 [4]. This result is in agreement with the findings that L23 rebinds to the 18 S RNA fragment which spans from nucleotide position 1300 to the 3′-end of the 23 S RNA [5] and that in reconstitution experiments with 23 S RNA fragments protein L23 binds to a fragment spanning from nucleotide position 1300–2000 [6].

Several *E. coli* mutants with altered ribosomal protein L23 have been isolated by two different isolation procedures [7,8]. The phenotype of a kasugamycin-dependent mutant is the product of a mutation in gene *ksgA* leading to resistance to kasugamycin and in the gene for protein L23 leading to dependence on this drug [9]. Since kasugamycin acts on the 30 S subunit it is interesting that an altered 50 S protein, namely L23, influences the kasugamycin phenotype of the ribosome. This is consistent with the conclusion mentioned above that L23 is an 'interphase protein'.

In this paper the complete primary structure of protein L23 is given, and its secondary structure as predicted from the amino acid sequence is presented. In addition the results on the comparison of the L23 sequence with the primary structures of 53 other ribosomal proteins which have already been sequenced are discussed.

2. Materials and methods

Protein L23 was isolated from 50 S subunits of *E. coli* strain K12 as described [10]. The protein samples were provided by Drs H. G. Wittmann and J. Dijk. The purity of the protein was checked by two-dimensional gel electrophoresis [11]. Furthermore, a fragment of protein L23 isolated as a byproduct during purification of the salt extracted protein [12] was provided by Dr J. Dijk. This fragment spans the sequence from amino acid position 10 to the C-terminus.

For the sequencing studies peptides were isolated applying the following enzymatic hydrolyses:

(i) Trypsin digestion at pH 8.1 (in 0.1 M *N*-methylmorpholine acetate buffer, at 37°C for 4–6 or 12 h);
(ii) Thermolysin digestion at pH 8.1 (same buffer, at 37 and 50°C for 2–4 h);
(iii) Chymotrypsin digestion at pH 8.1 (same buffer, at 37°C for 1 h);
(iv) Pepsin digestion at pH 2 (in 16% acetic acid at 37°C for 4 h);
(v) Digestion with *Staphylococcus aureus* protease at pH 4.0 (in 0.1 M ammonium acetate buffer at 37°C for 48 h).

In addition, partial acid hydrolysis of protein L23 for cleaving peptide bonds adjacent to aspartic acid residues was performed (in 2% acetic acid for 15 h at 110°C).

For specific cleavage at arginine residues the lysines were blocked with ETPA (exo-*cis*-3,6-endoxo-Δ^4-tetrahydrophthalic acid) [13] with 10 mg reagent/2 mg protein in 0.25 M borate buffer at pH 9 for 3 h at room temperature and for 3 h at 4°C. After dialysis in dilute ammonia (pH 9) the tryptic digestion was performed and followed by deblocking of the lysines in 10% acetic acid overnight.

The resulting peptides were isolated as follows:

(i) By thin-layer fingerprinting (tryptic, thermolytic, chymotryptic, *Staphylococcus aureus* and ETPA peptides);

(ii) By column chromatography on Dowex 50 (2 × 90 mm, 50°C) and one-dimensional preparative thin-layer purification (tryptic peptides);

(iii) By gel filtration on Sephadex G-50 superfine (180 × 1 cm columns) in 10% acetic acid and purification by fingerprints or by one-dimensional chromatography on thin-layer sheets (tryptic, ETPA, *Staphylococcus aureus* and peptic peptides) as summarized in [14].

Undigested or insoluble 'core' peptides, e.g., resulting from incomplete hydrolysis during trypsin digestion, of the regions positions 45–64 or 34–87, were dissolved in small volumes of 8 M urea, diluted to 2 M with morpholine buffer (pH 8.1) and were further digested with trypsin or thermolysin. Pepsin digestion of the 'core' peptides was without urea in 16% acetic acid.

The Edman degradations were performed manually with the combined dansyl-Edman technique [15] or by the DABITC/PITC (4-*NN*-dimethyl aminoazobenzene-4′-isothiocyanate/phenylisothiocyanate) double coupling method [16]. Solid-phase sequencing of peptides was made by attaching the carboxyl groups of the C-terminal end of the peptides with a water-soluble carbodiimide on amino-polystyrene resin [17] at pH 5 and 30°C, for 1–2 h. The released PTH-amino acids were identified by thin-layer techniques.

Amino acid analyses were performed in a Durrum D-500 analyser, using 1–2 nmol peptides for the hydrolysis (in 50 μl of double-distilled 5.7 N HCl, for 20 h at 110°C; sensitivity scale of 0.5 *A*). Protein L23 was hydrolysed in 100 μg amounts (in 50 μl HCl at 110°C for 20, 48 and 72 h) with aliquots of 20% of the sample at 2.0 *A* sensitivity, for each analysis. The analyser columns and buffers were self-prepared.

3. Results and discussion

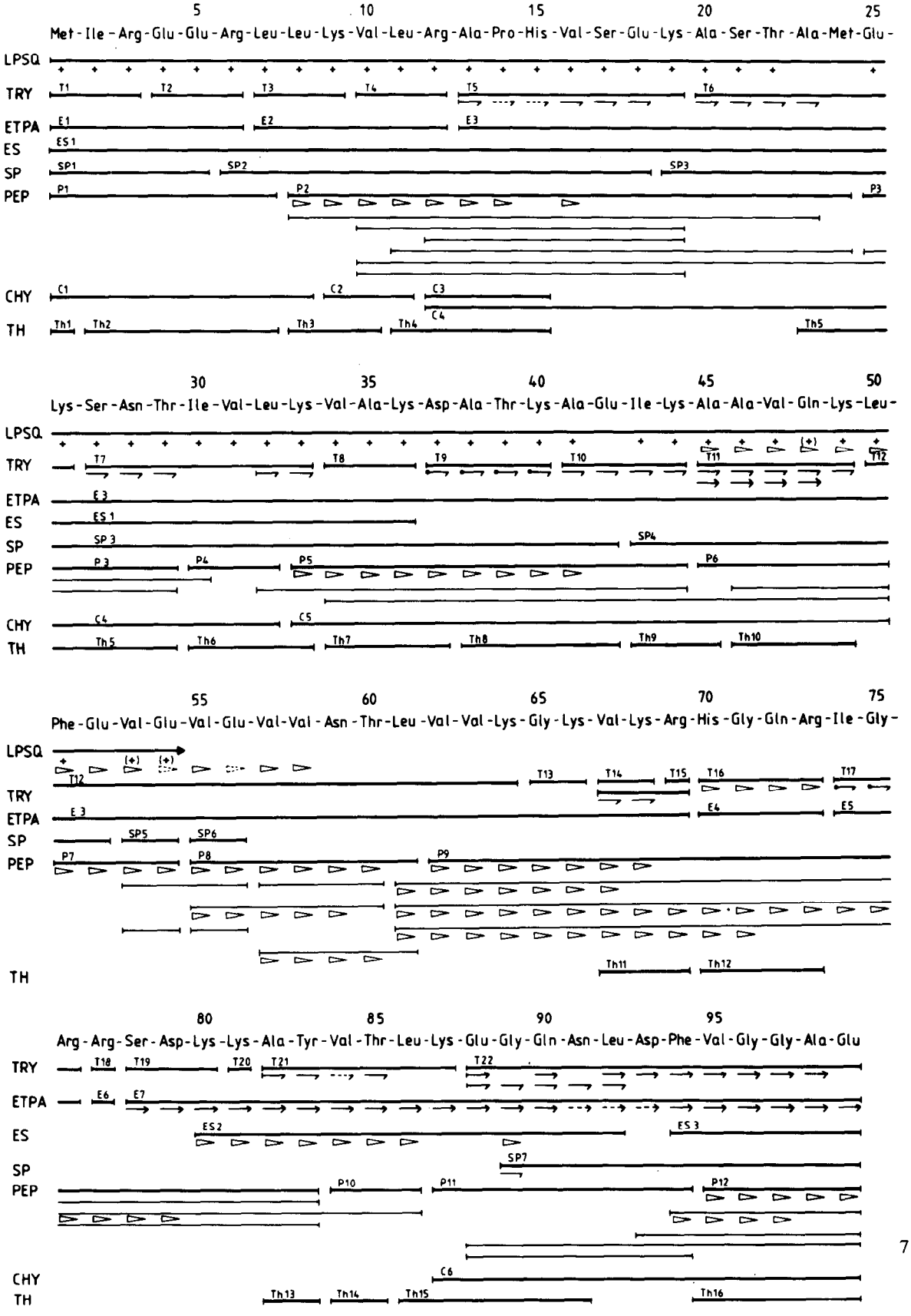
3.1. *Sequence determination*

The N-terminal sequence of protein L23 up to position 51 was determined [18] by liquid-phase Edman degradation performed in a considerably modified Beckman sequencer as reviewed [19]. Thin-layer chromatography and hydrolysis of the released PTH-amino acids followed by their analysis were used for the identification as in [18]. The sequence within the N-terminal region of the chain was confirmed by isolating peptides after digestions with trypsin, pepsin, *Staphylococcus aureus* protease, chymotrypsin, thermolysin as well as by tryptic peptides after blocking of the lysines by ETPA. The alignment of the tryptic peptides T1 to T12 derived from the sequencer result as shown in fig.1. All other peptides obtained could be placed into this sequence and were partially sequenced by the combined dansyl-Edman technique by solid-phase sequencing or by the manual DABITC/PITC double-coupling method (as detailed in fig.1).

Determination of the sequence of the middle part of the protein was more difficult for several reasons.

(1) The tryptic cleavage between T11 and T12 was incomplete, and the 'core' peptide from this area, not soluble in water or dilute acids, turned out to be heterogeneous at the N-terminus, starting mainly with alanine (pos. 45) and to a

Fig.1. Primary structure of protein L23 of *Escherichia coli* ribosomes. LPSQ: Degradation of intact protein L23 by means of a modified Beckman sequencer, + means unambiguously identified residue by thin-layer technique and additionally by amino acid analysis after hydrolysis. Digestion with trypsin, TRY; with trypsin after reaction with ETPA, ETPA; with *Staphylococcus aureus* protease, SP; with pepsin, PEP; with chymotrypsin, CHY; and with thermolysin, TH; ES means peptides from partial hydrolysis in dilute acetic acid. Symbols are as follows: ——, combined dansyl-Edman degradation, →, solid-phase sequencing in a solid-phase sequencer; ▷, manual DABITC/PITC method.


5 10 15 20 25
Met-Ile-Arg-Glu-Glu-Arg-Leu-Leu-Lys-Val-Leu-Arg-Ala-Pro-His-Val-Ser-Glu-Lys-Ala-Ser-Thr-Ala-Met-Glu-
LPSQ
TRY
ETPA
ES
SP
PEP
CHY
TH
T1 T2 T3 T4 T5 T6
E1 E2 E3
ES1
SP1 SP2 SP3
P1 P2 P3
C1 C2 C3 C4
Th1 Th2 Th3 Th4 Th5
30 35 40 45 50
Lys-Ser-Asn-Thr-Ile-Val-Leu-Lys-Val-Ala-Lys-Asp-Ala-Thr-Lys-Ala-Glu-Ile-Lys-Ala-Ala-Val-Gln-Lys-Leu-
T7 T8 T9 T10 T11 T12
E3
ES1
SP3 SP4
P3 P4 P5 P6
C4 C5
Th5 Th6 Th7 Th8 Th9 Th10
55 60 65 70 75
Phe-Glu-Val-Glu-Val-Glu-Val-Val-Asn-Thr-Leu-Val-Val-Lys-Gly-Lys-Val-Lys-Arg-His-Gly-Gln-Arg-Ile-Gly-
T12 T13 T14 T15 T16 T17
E3 E4 E5
SP5 SP6
P7 P8 P9
Th11 Th12
80 85 90 95
Arg-Arg-Ser-Asp-Lys-Lys-Ala-Tyr-Val-Thr-Leu-Lys-Glu-Gly-Gln-Asn-Leu-Asp-Phe-Val-Gly-Gly-Ala-Glu
T18 T19 T20 T21 T22
E6 E7
ES2 ES3
SP7
P10 P11 P12
C6
Th13 Th14 Th15 Th16

lower extent with leucine (pos. 50). Therefore, only a few nanomoles of peptide T12 and the 'bridging' peptide T11–T12 could be purified.

(2) The sequencing of these peptides was difficult because of the repetitive sequence Glu–Val–Glu–Val–Glu–Val–Val (pos. 52–58) and a further Val–Val sequence (pos. 62–63). The best result was obtained on peptide T11–T12 with the manually performed DABITC/PITC double coupling method [16] which led to the sequence determination up to position 58.

(3) The presence of ten lysine peptides (T5–T14) without any arginine residue gave rise to a long ETPA peptide, E3 (pos. 13–69) which contained the difficult sequence area mentioned above.

(4) The thermolysine peptides of the middle part of the protein chain consisted of many small pieces with similar sequences which could not be purified or aligned unambiguously.

(5) The alignment of peptides T13 and T14 and that of the single lysine and arginine residues (pos. 81, 69 and 77) complicated a final proof for this part of the chain.

These difficulties were overcome by the detailed study of the rather complex peptide mixture obtained after pepsin hydrolysis of the tryptic 'core' applying the DABITC/PITC technique on a few nanomoles of purified peptides. In this way it was possible to determine the entire sequence of this region up to pos. 79 (see fig.1). Details will be described elsewhere.

The primary structure of the C-terminal part of the protein chain was determined by sequencing the ETPA peptide, E7 (pos. 78–99) with the solid-phase sequencer and by sequencing the peptide ES2 (pos. 80–92) obtained from partial acid hydrolysis with the DABITC/PITC method. Additional isolated fragments and sequenced peptides of this region are shown in fig.1.

3.2. *Characteristics of the sequence*

The sequence determination of protein L23 was derived from several sets of peptides. The tryptic and peptic peptides independently span the entire sequence of L23 as shown in fig.1. The amino acid composition of protein L23 is Asp_3 Asn_3 Thr_5 Ser_4 Glu_{10} Gln_3 Pro_1 Gly_6 Ala_{10} Val_{14} Met_2 Ile_4 Leu_8 Tyr_1 Phe_2 His_2 Lys_{14} Arg_7. Cysteine and tryptophan were not detected. The molecular weight based on the sequence amounts to 11 013. The amino acid analyses of the entire protein are in good agreement with the sum of the amino acid residues derived from the sequence given in fig.1.

Protein L23 has a rather high amount of hydrophobic amino acids, namely 26% valine, leucine and isoleucine. The valine residues are clustered in one part of the protein between amino acids 31–67 and especially in region 53–63 which contains 6 valines out of 11 residues. Because of the repetitive sequence and the incomplete hydrolysis in this region with most of the applied enzymes the sequence analysis of this part of the chain was rather difficult as mentioned above.

The charged amino acids in protein L23 (which contains 36% acidic and basic residues) are distributed over the whole protein chain with a positively charged cluster at positions 64–81 and an accumulation of glutamic acid residues at positions 52–56. Furthermore, charged amino acids are clustered in the N-terminal area up to pos. 19. The single proline residue occurs in the N-terminal region at pos. 14, the two phenylalanine residues are in the valine region at pos. 51 and at pos. 94 and the single tryrosine in the C-terminal area at pos. 83.

3.3. *Secondary structure predictions of protein L23*

Four different prediction methods were employed for calculating the secondary structure of protein L23 as presented in fig.2 and described in [20]. In case of the predictions according to Chou and Fasman recent parameters [21,22] were used. According to these predictions two pronounced regions of secondary structure, predicted in accordance with all methods (see fig.2), can be given, namely an extended helix area for positions 33–54 in the middle part of the chain, and furthermore one to two strong turn or loop areas at positions 64–81. Moreover, several areas of extended structure or pleated sheet are probably located before and after the central helix part (around pos. 30, 60) and at the C-terminal end of the protein chain. The amount of secondary structure calculated for protein L23 based on 3 out of 4 agreeing predictions is as follows: 35% helix, 14% turns or loops and at least 8% of extended structure (see line 'PRE' in fig.2).

Recently, CD-measurements were performed on protein L23 prepared by the salt extraction method [23]. The CD-spectrum measured in 300 mM potassium fluoride and 5 mM potassium phosphate at

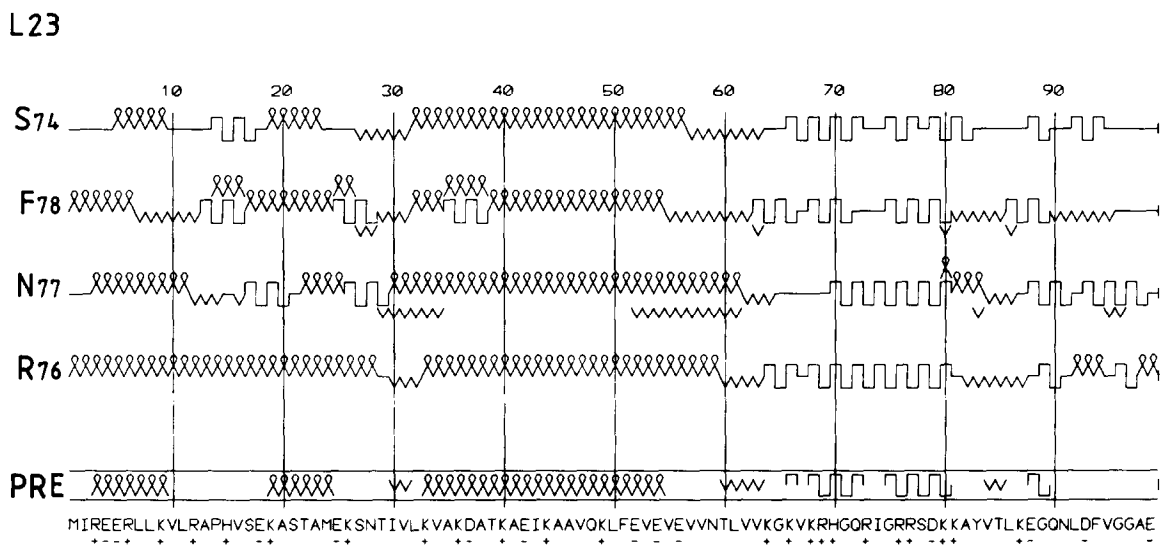


Fig.2. Predictions of the secondary structure of protein L23 according to 4 different methods, for the details see text and [20].

pH 7.0 shows relatively low levels of ordered structure (J. Dijk, personal communication).

3.4. *Comparison with sequences of other ribosomal proteins*

The sequence of protein L23 was compared with 48 sequenced *E. coli* ribosomal proteins (reviewed in [24]) and with 5 complete structures of proteins from other organisms (yeast, *Bacillus subtilis* and *Artemia salina*; reviewed in [25]). The results are given in table 1 in which identical sequence regions are listed.

Table 1
Identical peptides in protein L23 and other ribosomal proteins from *Escherichia coli*

L23 positions	Sequences	Protein	Positions
17–20	Ser–Glu–Lys–Ala	S20	13–16
23–26	Ala–Met–Glu–Lys	S5	62–65
24–27	Met–Glu–Lys–Ser	S17	16–19
28–31	Asn–Thr–Ile–Val	S11	28–31
30–33	Ile–Val–Leu–Lys	L6	130–133
32–35	Leu–Lys–Val–Ala	S8	31–34
33–36	Lys–Val–Ala–Lys	S12	14–17
37–40	Asp–Ala–Thr–Lys	L1	15–18
44–47	Lys–Ala–Ala–Val	S8	21–24
44–47	Lys–Ala–Ala–Val	L22	42–45
52–55/54–57	Glu–Val–Glu–Val	L7/L12	116–119
65–68	Gly–Lys–Val–Lys	L24	22–25
65–68	Gly–Lys–Val–Lys	L27	64–67
65–68	Gly–Lys–Val–Lys	L14	50–53
67–70	Val–Lys–Arg–His	S21	52–55
86–89	Leu–Lys–Glu–Gly	L7/L12	94–97
86–89	Leu–Lys–Glu–Gly	NS1	36–39
96–99	Gly–Gly–Ala–Glu	S3	157–160

Acknowledgements

We wish to thank Dr J. Dijk for a gift of protein samples of protein L23 and its fragment.

References

[1] Tischendorf, G. W. and Stöffler, G. (1975) Proc. Natl. Acad. Sci. USA 72, 4820–4824.
[2] Stöffler, G. and Wittmann, H. G. (1977) in: Molecular Mechanisms of Protein Biosynthesis (Weissbach, H. and Pestka, S. eds) pp. 117–202, Academic Press, New York.
[3] Fabian, U. (1976) FEBS Lett. 71, 256–260.
[4] Branlant, C., Krol, A., Sriwidada, J., Ebel, J. P., Sloof, P. and Garrett, R. (1975) FEBS Lett. 52, 195–201.
[5] Spierer, P., Zimmermann, R. A. and Mackie, G. A. (1975) Eur. J. Biochem. 52, 459–468.
[6] Chen-Schmeisser, U. and Garrett, R. (1976) Eur. J. Biochem. 69, 401–410.
[7] Isono, K., Krauss, J. and Hirota, Y. (1976) Mol. Gen. Gen. 149, 297–302.
[8] Dabbs, E. R. and Wittmann, H. G. (1976) Mol. Gen. Genet. 149, 303–309.
[9] Dabbs, E. R. (1979) Mol. Gen. Genet. in press.
[10] Hindennach, J., Kaltschmidt, E. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 12–16.
[11] Kaltschmidt, E. and Wittmann, H. G. (1970) Analyt. Biochem. 36, 401–412.
[12] Dijk, J. and Littlechild, J. A. (1979) Methods Enzymol. 59, 481–502.
[13] Hitz, H., Schäfer, D. and Wittmann-Liebold, B. (1977) Eur. J. Biochem. 75, 497–512.
[14] Wittmann-Liebold, B., Brauer, D. and Dognin, M. (1977) in: Solid Phase Methods in Protein Sequence Analysis, Inserm Symp. no. 5, (Previero, A. and Coletti-Previero, M.-A. eds) pp. 219–232, Elsevier/North-Holland, Amsterdam, New York.
[15] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.
[16] Chang, Y. J., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[17] Wittmann-Liebold, B. and Lehmann, A. (1975) in: Solid Phase Methods in Protein Sequence Analysis, (Laursen, R. A. ed) pp. 83–90, Pierce Chemical Company, Rockford.
[18] Wittmann-Liebold, B., Geissler, A. W. and Marzinzig, E. (1975) J. Supramol. Struct. 3, 426–447.
[19] Wittmann-Liebold, B. (1979) in: 12th Miles Int. Symp. (Beers, F. ed) in press.
[20] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.
[21] Chou, P. Y. and Fasman, G. D. (1977) J. Mol. Biol. 115, 135–175.
[22] Chou, P. Y. and Fasman, G. D. (1978) Ann. Rev. Biochem. 47, 251–276.
[23] Dijk, J. and Ackermann, J. (1979) J. Biol. Chem. submitted.
[24] Wittmann, H. G., Littlechild, J. and Wittmann-Liebold, B. (1979) in: Ribosomes (Davies, J. et al. eds) University Park Press, Baltimore, in press.
[25] Matheson, A. T. and Möller, W. (1979) in: Ribosomes (Davies, J. et al. eds) University Park Press, Baltimore, in press.