THE PRIMARY STRUCTURE OF PROTEIN S5 FROM THE SMALL SUBUNIT OF THE *ESCHERICHIA COLI* RIBOSOME

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

Received 14 August 1978

1. Introduction

Protein S5 plays an important role in both assembly and function of the small ribosomal subunit. 30 S subunits reconstituted in the absence of protein S5 sediment at 28 S and are only 30–50% active in the in vitro poly (U)-directed polyphenylalanine synthesis system. The ability of these depleted particles to bind fMet-tRNA in the presence of AUG and IF-2 is drastically reduced [1]. These data are in good agreement with antibody-blocking experiments: anti-S5 is a strong inhibitor of the translation of natural and synthetic mRNA and exhibits these effects mainly by inhibition of the initiation step [2]. Furthermore, protein S5 is, together with proteins S2 and S9, implied in the association of ribosomal subunits and stimulates the EF-G-dependent GTPase activity [3]. On the other hand, specific antibodies to protein S5 inhibit neither subunit reassociation [4] nor EF-G-dependent GTP-hydrolysis [5].

Hydrodynamic studies on isolated protein S5 revealed an elongated shape of this protein [6,7]. A more compact structure for S5 was deduced from neutron scattering experiments [8]. Three antibody binding sites have been detected on the 30 S subunit by immune electron microscopy [9]: one of these sites is located in the lower region of the head, a second in the neck region and a third at the upper part of one of the lobes of the small subunit.

Protein S5, besides protein S7, differs among various *Escherichia coli* strains [10], and altered S5 proteins have been detected in a great variety of mutants:

(i) In spectinomycin resistant mutants [11];
(ii) In revertants from streptomycin dependence to independence [12–14];
(iii) In cold-sensitive [15] and temperature-sensitive [16] mutants;
(iv) In mutants suppressing a defect alanyl-tRNA synthetase [17];
(v) In neamine resistant mutants [18].

In this paper the complete primary structure of protein S5 is presented, and predictions for the secondary structure of this protein are made based on four different computer programmes. Furthermore, the sequence of protein S5 is compared for homology with 40 other *E. coli* proteins of known primary structure. Finally, the determination of the sequence of S5 allows the exact localization of the amino acid exchanges in the S5 protein chains which are altered in the different types of mutants mentioned above. The results obtained so far with the mutants are summarized at the end of this paper.

2. Materials and methods

Protein S5 was isolated from *E. coli* K strain A19 as in [19] and provided by Dr H. G. Wittmann. The identity and purity of the protein were checked by two-dimensional polyacrylamide gel electrophoresis [20].

Enzymatic digestions of the protein with trypsin, α-chymotrypsin and thermolysin were performed by preincubation of the water-soluble protein with these enzymes at neutral pH and 37°C for 0.5–1 h before adding dropwise 0.2 M *N*-methyl-morpholine acetate buffer (pH 8.1) or dilute ammonia to reach a final

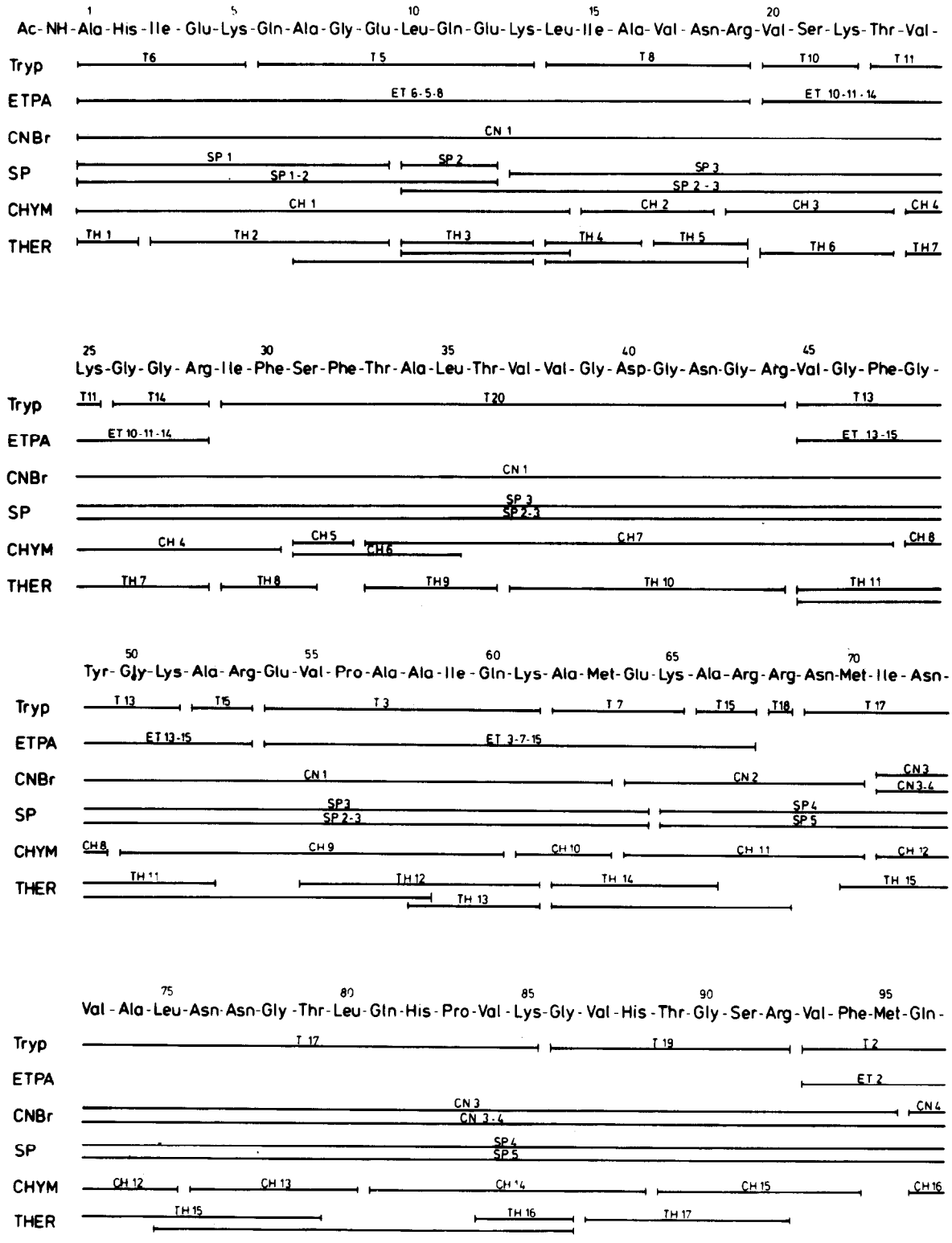
1 5 10 15 20
Ac-NH-Ala-His-Ile-Glu-Lys-Gln-Ala-Gly-Glu-Leu-Gln-Glu-Lys-Leu-Ile-Ala-Val-Asn-Arg-Val-Ser-Lys-Thr-Val-
Tryp
ETPA
CNBr
SP
CHYM
THER
T6
T5
T8
T10
T11
ET 6-5-8
ET 10-11-14
CN 1
SP 1
SP 2
SP 3
SP 1-2
SP 2-3
CH 1
CH 2
CH 3
CH 4
TH 1
TH 2
TH 3
TH 4
TH 5
TH 6
TH 7
25 30 35 40 45
Lys-Gly-Gly-Arg-Ile-Phe-Ser-Phe-Thr-Ala-Leu-Thr-Val-Val-Gly-Asp-Gly-Asn-Gly-Arg-Val-Gly-Phe-Gly-
Tryp
ETPA
CNBr
SP
CHYM
THER
T11
T14
T20
T13
ET 10-11-14
ET 13-15
CN 1
SP 3
SP 2-3
CH 4
CH 5
CH 6
CH 7
CH 8
TH 7
TH 8
TH 9
TH 10
TH 11
50 55 60 65 70
Tyr-Gly-Lys-Ala-Arg-Glu-Val-Pro-Ala-Ala-Ile-Gln-Lys-Ala-Met-Glu-Lys-Ala-Arg-Arg-Asn-Met-Ile-Asn-
Tryp
ETPA
CNBr
SP
CHYM
THER
T13
T15
T3
T7
T15
T18
T17
ET 13-15
ET 3-7-15
CN 1
CN 2
CN 3
CN 3-4
SP 3
SP 2-3
SP 4
SP 5
CH 8
CH 9
CH 10
CH 11
CH 12
TH 11
TH 12
TH 13
TH 14
TH 15
75 80 85 90 95
Val-Ala-Leu-Asn-Asn-Gly-Thr-Leu-Gln-His-Pro-Val-Lys-Gly-Val-His-Thr-Gly-Ser-Arg-Val-Phe-Met-Gln-
Tryp
ETPA
CNBr
SP
CHYM
THER
T17
T19
T2
ET 2
CN 3
CN 3-4
CN 4
SP 4
SP 5
CH 12
CH 13
CH 14
CH 15
CH 16
TH 15
TH 16
TH 17

Fig.1, continued

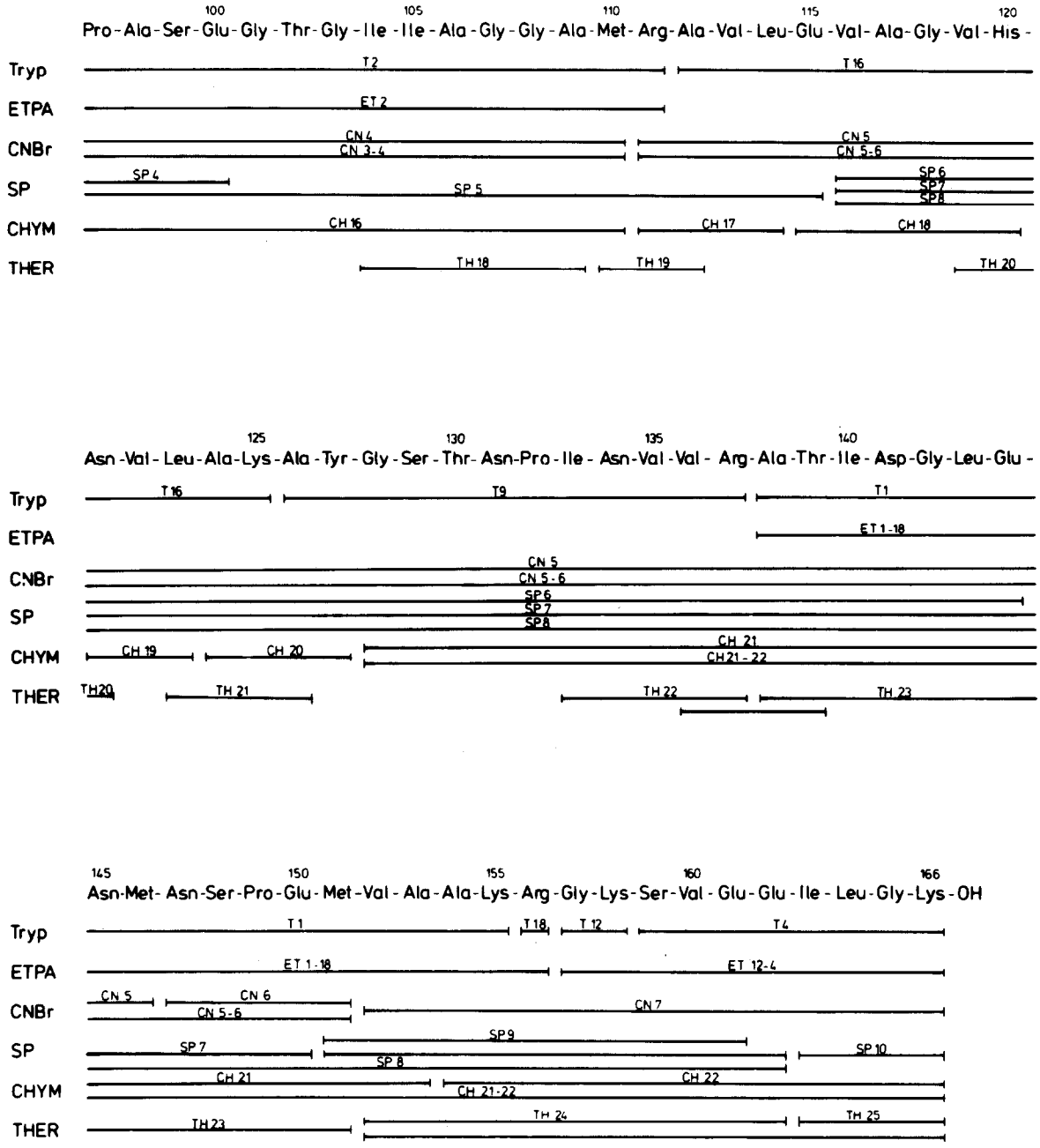

Fig.1. The primary structure of protein S5K from *Escherichia coli. Abbreviations:* Tryp, trypsin digestion (peptides T); ETPA, trypsin digestion after blocking the amino groups by ETPA (peptides ET); CNBr, cleavage with cyanogen bromide (peptides CN); SP, digestion with *Staphylococcus aureus* protease; CHYM, digestion with α-chymotrypsin (peptides CH); THER, digestion with thermolysin (peptides TH).

pH of 7.8–8.0. Digestions with trypsin were then done for 4 h, with chymotrypsin for 1–3.5 h at 37°C and with thermolysin for 2 h at 37°C or 55°C. Furthermore, tryptic digestions were made after blocking of the ε-amino groups of the lysines with exo-*cis*-3,6-endoxo-Δ^4-tetrahydrophthalic anhydride (ETPA) [21,22]. Cleavage with *Staphylococcus aureus* protease was done in 50 mM ammonium acetate buffer at pH 4.0 for 24–48 h at 37°C [23]. Digestions with carboxypeptidase A and B were performed at pH 7.8 and 37°C. The protein was digested between 10 min and 4 h and peptides for 2–4 h. Cleavage with cyanogen bromide was made in 70% formic acid at 37°C for 24–48 h with a 100-fold molar excess of the reagent over methionine.

Separation of the peptides was achieved by chromatography on sulfoethyl-cellulose columns [24], on Dowex M71 micro columns [25] and by the fingerprint technique using cellulose thin-layer sheets [22]. The tryptic peptide T20 was isolated as a sediment from the lyophilized tryptic digestion which was dissolved in water. Larger peptides, mainly deriving from *Staph. aureus* protease digestion and cyanogen bromide treatment, were fractionated by gel filtration on Sephadex G-50, G-75 and G-100 in 10% acetic acid or 20% pyridine. The smaller peptides eluted from these Sephadex column runs were further purified by one-dimensional thin-layer chromatography or by the thin-layer fingerprint technique. Amino acid analyses were done on a Durrum D-500 or LKB 3201 analyzer as detailed in [22].

Sequencing of the various peptides was performed by the following methods:

(i) Manual Edman degradation combined with dansylation of the free N-terminal amino acid residue after each degradation cycle [26,27]. In cases of Glu/Gln and Asp/Asn the released thiazolinones were manually converted to the PTH-amino acids which were identified by thin-layer chromatography or electrophoresis.

(ii) Solid-phase sequencing [28] with attachment of the C-terminal carboxyl groups of the peptides to aminopolystyrene resin by means of a water-soluble carbodiimide [29]. The thiazolinones were converted to the PTH-amino acids with 20% TFA and were identified by thin-layer chromatography on silica gel sheets [30].

(iii) Insoluble and large peptides were sequenced by automated Edman degradation with the liquid-phase technique [31] using an improved Beckman sequencer together with automated conversion device (reviewed in [32]). The PTH-amino acids were identified by thin-layer chromatography as in [30].

3. Results and discussion

3.1. *Sequence determination*

Protein S5 was subjected to digestion with various proteases and to chemical cleavage employing cyanogen bromide. The resulting peptides which are shown in fig.1 were isolated by a combination of column chromatography, sulfoethyl-cellulose, Dowex M71 micro method, or gel filtration on Sephadex with thin-layer chromatography or fingerprint technique.

All tryptic peptides, except T6 and T20, were sequenced by the manual dansyl-Edman technique and by the solid-phase method with attachment of the C-terminal carboxyl group of the peptides to amino-polystyrene resin. Tryptic peptide T20, which was insoluble in dilute buffer solutions, was sequenced in the improved Beckman sequenator.

The peptides deriving from the other cleavages were mainly sequenced by the dansyl-Edman technique. Peptide SP4 (position 65–100) was degraded in the liquid-phase sequencer (up to position 83). The sequence of SP9 (position 151–161/162) and the N-terminal sequence of peptide SP6 (position 116–144) up to position 127 were obtained by means of the solid-phase method.

Peptide T6, located at the N-terminus of wild-type protein S5, could not be degraded by the Edman method because it has a blocked N-terminus. However, several mutants with altered S5 protein were found to have unblocked N-terminus [33]. The N-terminal peptides T6 of these mutants were sequenced by the solid-phase technique and gave the sequence shown in fig.1.

Knowing that the amino acid composition of wild-type T6 was the same as that of mutant T6, we further studied wild-type T6 with blocked N-terminus by digestion of the peptide with thermolysin which released two fragments: T6-THa (Ala_1, His_1) and T6-THb (Ile–Glu–Lys); the sequence of the latter

was obtained by dansyl-Edman degradation. T6-THa was treated with carboxypeptidase B which resulted in free histidine. The carboxypeptidase digest was chromatographed on a thin-layer cellulose sheet, and diverse zones, not staining with ninhydrin reagent, were eluted for amino acid analysis. A zone containing only alanine, which co-chromatographed with authentic *N*-mono-acetyl-alanine was examined in a Varian CH-7 mass spectrometer provided with an electron impact ionic source. The alanine containing zone gave the molecular ion $[M + 1]^+$ at $m/e = 132$ and the $[M + 2]^+$ ion at $m/e = 133$. In addition, the thermolytic peptide T6-THa was directly analyzed in a Varian 311 mass spectrometer employing the field desorption technique. The molecular mass ion of *N*-mono-acetyl-alanyl-histidine [M + 1] at $m/e = 269$ was obtained.

3.2. *Alignment of the isolated peptides*

The alignment of the tryptic peptides T6–T5–T8–T10–T11–T14–T20 in the N-terminal region of the protein chain was made by combining the sequence results on *Staph. aureus* protease, chymotrypsin and thermolysin peptides. It was confirmed by liquid-phase sequencing of the unblocked S5 protein derived from mutant JE386 [33] up to position 34. In addition, the adjacent positions of peptide T8–T10 were known from protein-chemical studies on the protein S5 from mutant $spc^{R\text{-}9}$ [34] in which Arg_{19} of T8 is replaced by leucine resulting in the tryptic 'bridge-peptide' T8–T10.

The alignment of the tryptic peptides T20–T13–T15–T3–T7 was obtained from peptides CH7, CH9 and TH14. The sequence of CN2 and the N-terminal sequence of SP4 provided the alignment of T7–T15–T18–T17. Chymotryptic peptide CH14 served as a bridge for T17–T19 and peptide CH15 for T19–T2.

The order of T2–T16 was known from mutants of S5 in which Arg_{111} is replaced by leucine in the mutant N-421 [35] and in d-1023 [36]. The same result came from analysis of mutant *nea*-319 where this arginine residue is replaced by serine and from *nea*-314 where it is exchanged by glycine [37].

The order of peptides T16–T9–T1–T18–T12–T4 derived from the sequence of CH20, the composition and the liquid-phase sequencing of peptide SP6 as well as solid-phase sequencing of peptide SP9 and CN7. This alignment is in good agreement with results from treatment of the intact protein S5 with carboxypeptidase A and B, which releases lysine, glycine, leucine and isoleucine. Furthermore, the mutant *sup* 0-1 has a shortened S5 protein chain with a smaller peptide T4 than in the wild-type [35].

3.3. *The primary structure of protein S5*

Combination of all results, which will be detailed elsewhere, led to the primary structure of protein S5 of *E. coli* as shown in fig.1. Protein S5 has 166 amino acid residues and the composition derived from the sequence is: Asp_2 Asn_{11} Thr_8 Ser_7 Glu_{11} Gln_5 Pro_5 Gly_{21} Ala_{20} Val_{19} Met_6 Ile_{10} Leu_9 Tyr_2 Phe_4 His_4 Lys_{12} Arg_{10}. Tryptophan and cysteine have not been found by the appropriate methods for their detection. The blocked N-terminus of S5 was identified as *N*-mono-acetyl-alanine. The molecular weight calculated on the basis of the sequence given in fig.1 is 17 515.

The protein has a tendency to precipitate at pH 7.5–8.0. Therefore complete cleavages with trypsin, thermolysin and chymotrypsin were only obtained by preincubating the protein with these enzymes in water prior to the adjustment of the optimal pH for the digestion.

Cleavage with *Staph. aureus* protease and cyanogen bromide mainly gave large fragments which were heterogeneous at their N- or C-terminal part, e.g., SP3/SP2–3 and SP6/7/8 (see fig.1). Therefore the yield of individual fragments was rather low and made only N-terminal sequence studies possible.

3.4. *Secondary structure of protein S5*

In fig.2 the secondary structure of protein S5 is presented as predicted by four different methods detailed in [38]. In the line 'PRE' of fig.2, regions for helix, extended structure and β-turns are indicated provided that at least three out of four methods are in agreement. Based on this prediction, protein S5 contains at least 34% helix, 18% β-turns and 8% extended structure.

3.5. *Comparison of protein S5 with other ribosomal proteins*

Comparison of S5 with other *E. coli* ribosomal proteins of known sequences revealed that a pentapeptide is identical in S5 and S14 and another penta-

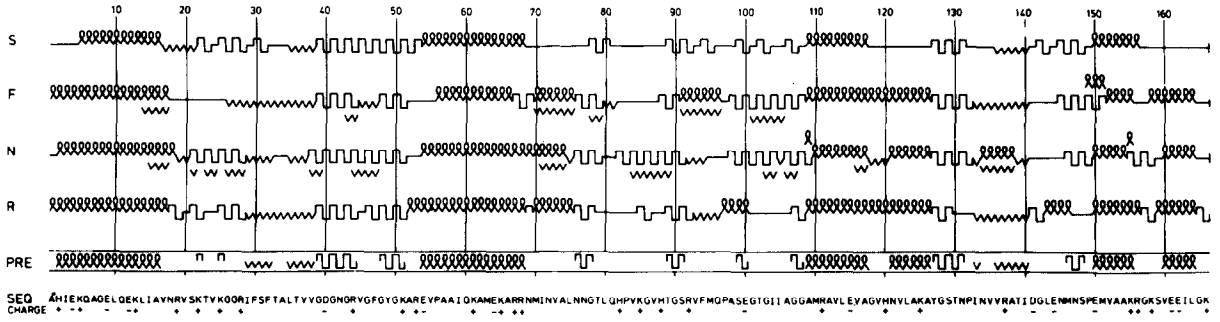

Fig.2. Secondary structure of ribosomal protein S5 predicted according to four different methods (detailed in [38]). The symbols represent residues in helical (ℓ), β-turn or loops (┐└), extended (w) and coil (——) conformational states, respectively. The diagram denoted by 'PRE' summarises the secondary structure obtained when three out of four predictions are in agreement. Å means *N*-mono-acetyl-alanine.

peptide in S5 and L29. Each of eight tetrapeptides of S5 were also found in proteins S2, S3, S4, S16, S20, L4, L14, L23 and three tetrapeptides are shared with protein L24. Furthermore, the following stretch of sequences occurs both in S5 and S9: . . . Glu–Lys . . . Glu–Lys–Leu . . . Thr–Val–Lys–Gly–Gly . . . Lys–Ala–Arg–Arg . . .

3.6. *Comparison of proteins S5 from two* E. coli *strains*

From the analysis of the tryptic peptides of proteins S5 isolated from the *E. coli* strains K and B it became obvious that one glutamic acid residue in S5K is replaced by alanine in the tryptic peptide T1 of S5B [24]. Solid-phase degradation of the peptide T1 of S5B enabled us now to locate alanine at position 150 replacing glutamic acid of S5K. The N-terminal regions of proteins S5 from *E. coli* (fig.1) and *Bacillus stearothermophilus* ([39], M. Yaguchi, personal communication) are compared in fig.3. Identical amino acids occur in ~50% of the 30 positions sequenced so far in both proteins.

3.7. *Amino acid replacements in mutationally altered S5 proteins*

Ten mutants with altered S5 proteins have been studied up to now by protein-chemical methods. The results are summarized in fig.4. Similar to mutants altered in protein S12 [40], the amino acid exchanges in mutationally altered S5 proteins are clustered in a

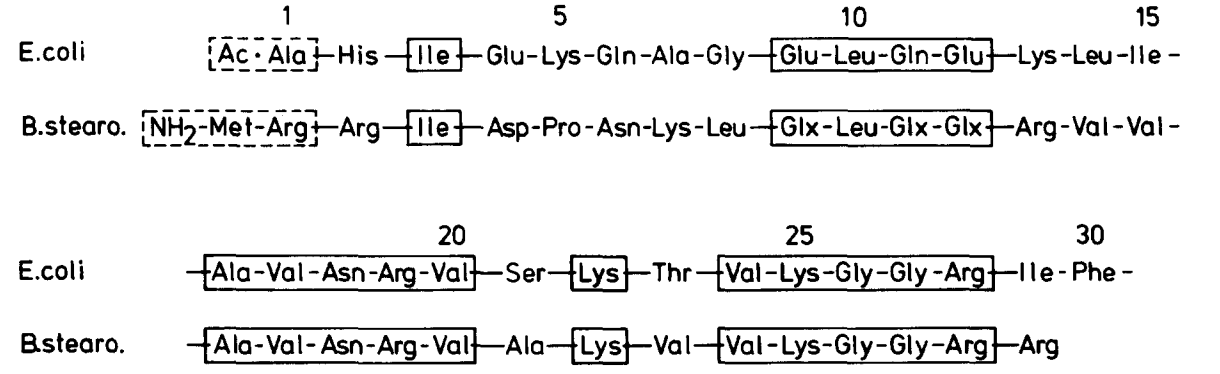

Fig.3. Comparison of N-terminal sequences of protein S5 from *E. coli* and *Bacillus stearothermophilus*.

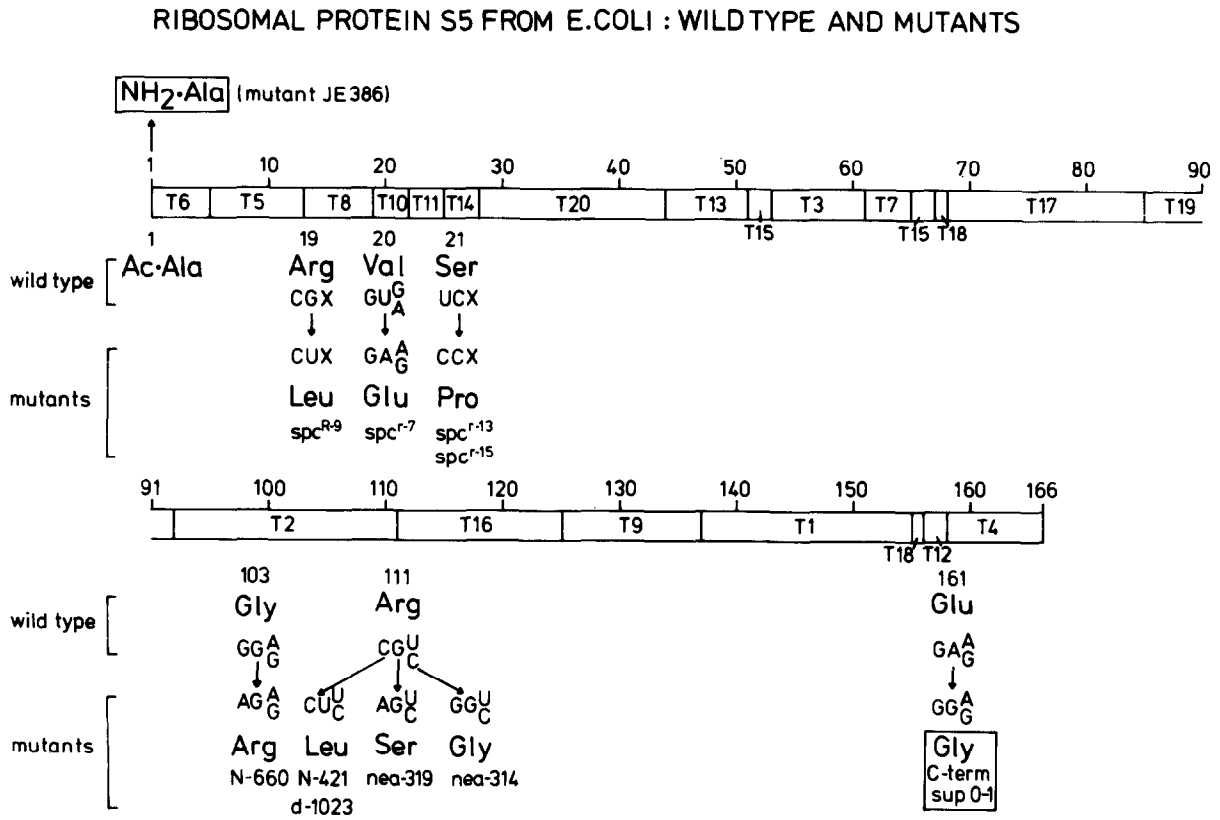

Fig.4. Amino acid replacements in mutationally altered S5 proteins. The alignment of the tryptic peptides is presented together with the positions where the amino acid exchanges have been localised.

few regions of the protein chain; in positions 19–21 for four spectinomycin-resistant mutants [34,41,42]: in positions 103 and 111 for three revertants from streptomycin-dependence to independence [35,36] and in position 111 for two neamycin-resistant mutants [37]. Protein S5 of mutant *sup* 0-1 in which the temperature-sensitive phenotype of an alanyl-tRNA synthetase mutation is partially suppressed, is shortened by five amino acids at the C-terminus and terminates with –Ser–Val–Gly [35]. Finally, mutants have recently been isolated whose N-terminus is not acetylated [33].

Acknowledgements

We thank Mr A. Lehmann for excellent technical assistance and Dr K. Isono for his interest in this work and for providing protein samples of the temperature-sensitive S5 mutants. We are indepted to Dr Höhne, Institut für Organische Chemie at the Technische Universität, Berlin, for the analysis in the Varian 311 mass spectrometer.

References

[1] Nomura, M. and Held, W. A. (1974) in: Ribosomes (Nomura, M. et al. eds) pp. 193–223, Cold Spring Harbor Press, Long Island, NY.
[2] Lelong, J. C., Gros, D., Gros, F., Bollen, A., Maschler, R. and Stöffler, G. (1974) Proc. Natl. Acad. Sci. USA 71, 248–252.
[3] Sander, G., Marsh, R. C. and Parmeggiani, A. (1973) FEBS Lett. 33, 132–134.
[4] Morrison, C. A., Garrett, R. A., Zeichhardt, H. and Stöffler, G. (1973) Mol. Gen. Genet. 127, 359–368.

[5] Highland, J. H., Ochsner, E., Gordon, J., Bodley, R., Hasenbank, R. and Stöffler, G. (1974) Proc. Natl. Acad. Sci. USA 71, 627–630.
[6] Rohde, M. F., O'Brien, S., Cooper, S. and Aune, K. C. (1975) Biochemisty 14, 1079–1087.
[7] Georgalis, Y. and Giri, L. (1978) FEBS Lett. 95, 99–102.
[8] Engelman, D. M., Moore, P. B. and Schoenborn, B. P. (1975) Proc. Natl. Acad. Sci. USA 72, 3888–3892.
[9] Tischendorf, G. W., Zeichhardt, H. and Stöffler, G. (1975) Proc. Natl. Acad. Sci. USA 72, 4820–4824.
[10] Kaltschmidt, E., Stöffler, G., Dzionara, M. and Wittmann, H. G. (1970) Mol. Gen. Genet. 109, 303–305.
[11] Bollen, A., Davies, J., Ozaki, M. and Mizushima, S. (1969) Science 165, 85–86.
[12] Stöffler, G., Deusser, E., Wittmann, H. G. and Apirion, D. (1971) Mol. Gen. Genet. 111, 334–341.
[13] Hasenbank, R., Guthrie, C., Stöffler, G., Wittmann, H. G., Rosen, L. and Apirion, D. (1973) Mol. Gen. Genet. 127, 1–18.
[14] Dabbs, E. R. and Wittmann, H. G. (1976) Mol. Gen. Genet. 149, 303–309.
[15] Nashimoto, H., Held, W., Kaltschmidt, E. and Nomura, M. (1971) J. Mol. Biol. 62, 121–138.
[16] Isono, K., Cumberlidge, A. G. and Isono, S. (1977) Mol. Gen. Genet. 152, 239–243.
[17] Wittmann, H. G., Stöffler, G., Piepersberg, W., Buckel, P., Ruffler, D. and Böck, A. (1974) Mol. Gen. Genet. 134, 225–236 (1974).
[18] DeWilde, M., Cábezon, T., Herzog, A. and Bollen, A. (1975) Mol. Gen. Genet. 142, 19–33.
[19] Hindennach, J., Stöffler, G. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 7–11.
[20] Kaltschmidt, E. and Wittmann, H. G. (1970) Anal. Biochem. 36, 401–412.
[21] Riley, M. and Perham, R. N. (1970) Biochem. J. 118, 733–739.
[22] Hitz, H., Schäfer, D. and Wittmann-Liebold, B. (1977) Eur. J. Biochem. 75, 497–512.
[23] Houmard, J. and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[24] Wittmann-Liebold, B. and Wittmann, H. G. (1971) Biochim. Biophys. Acta 251, 44–53.
[25] Heiland, J., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.
[26] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[27] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.
[28] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[29] 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.
[30] Wittmann-Liebold, B., Geissler, A. W. and Marzinzig, E. (1975) J. Supramol. Struct. 3, 426–447.
[31] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[32] Wittmann-Liebold, B., Brauer, D. and Dognin, J. (1977) in: Solid Phase Methods in Protein Sequence Analysis (Previero, A. and Coletti-Previero, M. A. eds) pp. 219–232, Elsevier/North-Holland Biomedical Press, Amsterdam, New York.
[33] Cumberlidge, A. G. and Isono, K. (1978) Cell, in press.
[34] DeWilde, M. and Wittmann-Liebold, B. (1978) Mol. Gen. Genet. 127, 273–276.
[35] Piepersberg, W., Böck, A., Yaguchi, M. and Wittmann, H. G. (1975) Mol. Gen Genet. 143, 43–52.
[36] Itoh, T. and Wittmann, H. G. (1973) Mol. Gen. Genet. 127, 19–32.
[37] Yaguchi, M., Wittmann, H. G., Cábezon, T., DeWilde, M., Villarroel, R., Herzog, A. and Bollen, A. (1975) Mol. Gen. Genet. 142, 35–43.
[38] Dzionara, M., Robinson, S. M. L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.
[39] Yaguchi, M., Matheson, A. T. and Visentin, L. P. (1974) FEBS Lett. 46, 296–300.
[40] Funatsu, G., Yaguchi, M. and Wittmann-Liebold, B. (1977) FEBS Lett. 73, 12–17.
[41] Funatsu, G., Schiltz, E. and Wittmann, H. G. (1971) Mol. Gen. Genet. 114, 106–111.
[42] Funatsu, G., Nierhaus, K. and Wittmann-Liebold, B. (1972) J. Mol. Biol. 64, 201–209.