# PRIMARY STRUCTURE OF PROTEIN S20 FROM THE SMALL RIBOSOMAL SUBUNIT OF *ESCHERICHIA COLI*

Brigitte WITTMANN-LIEBOLD, Elke MARZINZIG and Arnold LEHMANN
*Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, Berlin-Dahlem, Germany*

Received 28 June 1976

## 1. Introduction

Protein S20 is a very basic component of the small subunit of the *E. coli* ribosome. It specifically binds to the 16S RNA of *E. coli* [1–8]. The binding site interacts with the 5′-part of the RNA and lies within the binding region for protein S4 [9]. Evidence was presented that the C-terminus of S20 is involved in the RNA binding [10]. For the elucidation of this interaction the primary structure of S20 is an essential prerequisite.

Protein S20 is located at or near the subunit interface [11]. In immuno-electronmicroscopic studies the anti-S20 attachment site was found on the forehead of the small subunit in the vicinity of proteins S19 and S11 [12].

A functional role of protein S20 in tRNA binding has been described [13]. Furthermore, S20 has shown to be altered in 'revertants' from temperature-sensitive mutants with altered alanyl-tRNA synthetases which led to an increased amount of stable RNA [14]. The less basic S20 proteins of the mutants differ in their C-termini from the wild type protein (S. Robinson and B. Wittmann-Liebold, unpublished).

In this paper we present the complete primary structure of the wild type protein S20 which is very rich in alanine and lysine. Out of 86 residues it contains 24 basic amino acid residues. Predictions of the secondary structure indicate a rather high α-helical content in this protein. Finally the primary structure of S20 is compared with the amino acid sequences of other ribosomal proteins of known primary structure and some regions identical among them were detected.

## 2. Materials and methods

Protein S20, isolated from *E. coli* strain K as described earlier [15] was provided by Dr H. G. Wittmann. The identity and purity of the protein was judged by two-dimensional polyacrylamide gel electrophoresis [16].

Enzymatic digestions employing trypsin treated with 1-chloro-4-phenyl-3-tosylamidobutan-2-one (TPCK) from Merck (Darmstadt) and thermolysin from Serva (Heidelberg) were performed at pH 8.0 at 37°C for 4 and 1 h, respectively. Digestions with α-chymotrypsin from Worthington (Freehold, N. J.) were made at pH 8.0 and 37°C for 1 h; treatment with *Staphylococcus aureus* protease [17] was performed in 0.05 M ammonium acetate buffer, pH 4.0, for 48 h at 37°C. Details for the enzymatic digestions are given elsewhere [18,19]. The total protein and the peptides were treated with carboxypeptidase A and B from Boehringer (Mannheim) in 0.1 M *N*-methylmorpholine buffer at pH 8.1. Cyanobromide cleavage of protein S20 was performed in 70% formic acid (0.02% mercaptoethanol) for 16 h at 37°C with 30-fold molar excess of reagent over methionine residues. Trypsin and thermolysin peptides were isolated by the fingerprinting technique on cellulose thin-layer plates [18,19] followed by separation on a micro column of Dowex 50 × 7 (0.3 × 10 cm) at 55°C with pyridine formate gradients. Chymotryptic peptides were separated on sulfoethylcellulose with pyridine formate gradients and purified with paper chromatography [20]. In addition, they were obtained with the fingerprinting technique. The SP-


*North-Holland Publishing Company – Amsterdam*

peptides derived from digestion with *Staphylococcus aureus* protease were separated by gel filtration on Sephadex G25 (super-fine; 190 X 0.5 cm) in 0.05 M pyridine acetate buffer, pH 7.0, and were purified by one-dimensional chromatography on cellulose thin-layer plates or fingerprinting technique. The CNBr-peptides were isolated either on Sephadex G50 (fine) in 20% acetic acid or on Sephadex G25 (fine) in 5% acetic acid followed by rechromatography on a Dowex 50 X 7 micro column. Amino acid analyses were performed on Durrum or LKB analyzers as described previously [21].

Edman degradations of the peptides from trypsin, thermolysin and *Staphylococcus aureus* protease digestion were performed by the solid phase technique [22] with attachment of the C-terminal carboxyl groups to aminopolystyrene resin [23,24].

## 3. Results and discussion

Protein S20 has the following amino acid composition: $Asp_2$, $Asn_6$, $Thr_2$, $Ser_4$, $Glu_3$, $Gln_5$, $Pro_1$, $Gly_2$, $Ala_{19}$, $Val_2$, $Met_3$, $Ile_7$, $Leu_3$, $Tyr_1$ $Phe_2$, $His_3$, $Lys_{14}$ and $Arg_7$. Appropriate methods for the detection of tryptophan and cysteine showed that these amino acids are absent. The presence of unusual amino acids cannot be completely excluded. Protein S20 contains 86 residues and has a molecular weight of 9554 daltons. Its primary structure is shown in fig.1.

After re-chromatography of small amounts of protein S20 on Sephadex G100 on a micro scale the main protein fractions gave analyses which agreed well with the amino acid composition resulting from the above given sequence. However, the value for glutamic acid was slightly higher when analyzing the total protein compared to the sum received from peptide analyses. The content of 24 basic and 5 acidic residues is in good agreement with the high isoelectric point of S20 [25].

The sequence determination was performed by combining the results [26,27] obtained from liquid phase Edman degradation of the intact protein with those from peptides isolated after digestions of protein S20 with trypsin, chymotrypsin, thermolysin and *Staphylococcus aureus* protease. Furthermore, peptides from cyanobromide treatment of total protein ensured the alignment of the peptides from the middle and the C-terminal part of the chain (fig.1). The purification of the peptides was mainly performed by combined separation on a Dowex 50 X 7 micro column with thin-layer chromatography. Paper chromatography and even thin-layer fingerprinting did not separate all of them. Full details of the procedures and results will be given elsewhere.

The N-terminal amino acids of the trypsin and thermolysin peptides were determined by dansylation. The peptides were partially sequenced by the combined Dansyl-Edman degradation [28] and those underlined with an arrow (see fig.1) were sequenced by the solid phase technique. Analyses of SP-3 (1–14); SP-4 (15–39); SP-5 (40–52); SP-2 (53–70) and SP-1 (71–86) and Edman degradations performed on SP–3, SP–5, SP–2 and SP-1 with the solid phase technique gave unambiguous results and were in accordance with the sequence information obtained from the tryptic and thermolysin peptides and with the results from total protein analyses. The acidic amino acids and their amides were identified as their phenylthiohydantoin derivatives after liquid and solid phase degradation of the corresponding peptides and their electrophoretic mobility at pH 6.5 and 4.4. These results were in full agreement with the specificity of *Staphylococcus aureus* protease cleavage at pH 4.0. However, the isolation of SP-2 (positions 53–70) with a C-terminal lysine and of SP-1 (positions 71–86) starting with the tryptic peptide T13 Ala–Ala–Arg (positions 71–73) cannot be expected by treatment with *Staphylococcus areaus* protease. Its occurrence can be explained by tryptic side reaction, as the digestion had been performed under drastic conditions to complete the cleavage of all glutamic acid peptide bonds.

In order to obtain a bridge peptide to overlap the region of SP-2 and SP-1 the shorter CNBr-peptides, isolated by gel filtration, were further purified on a Dowex 50 X 7 micro column. One CNBr-peptide gave the sum of the amino acids of the C-terminal region (positions 54–86), which was in excellent accordance with the amino acid composition of SP-2 plus SP-1 (without the N-terminal methionine residue); the other showed the amino acid composition of positions (28–86) and served as as bridge for SP-4 – SP-5 – SP-2 – SP1. In addition chymotryptic bridges were analyzed; however, as their separation was difficult, only small amounts were isolated. The sequence data presented

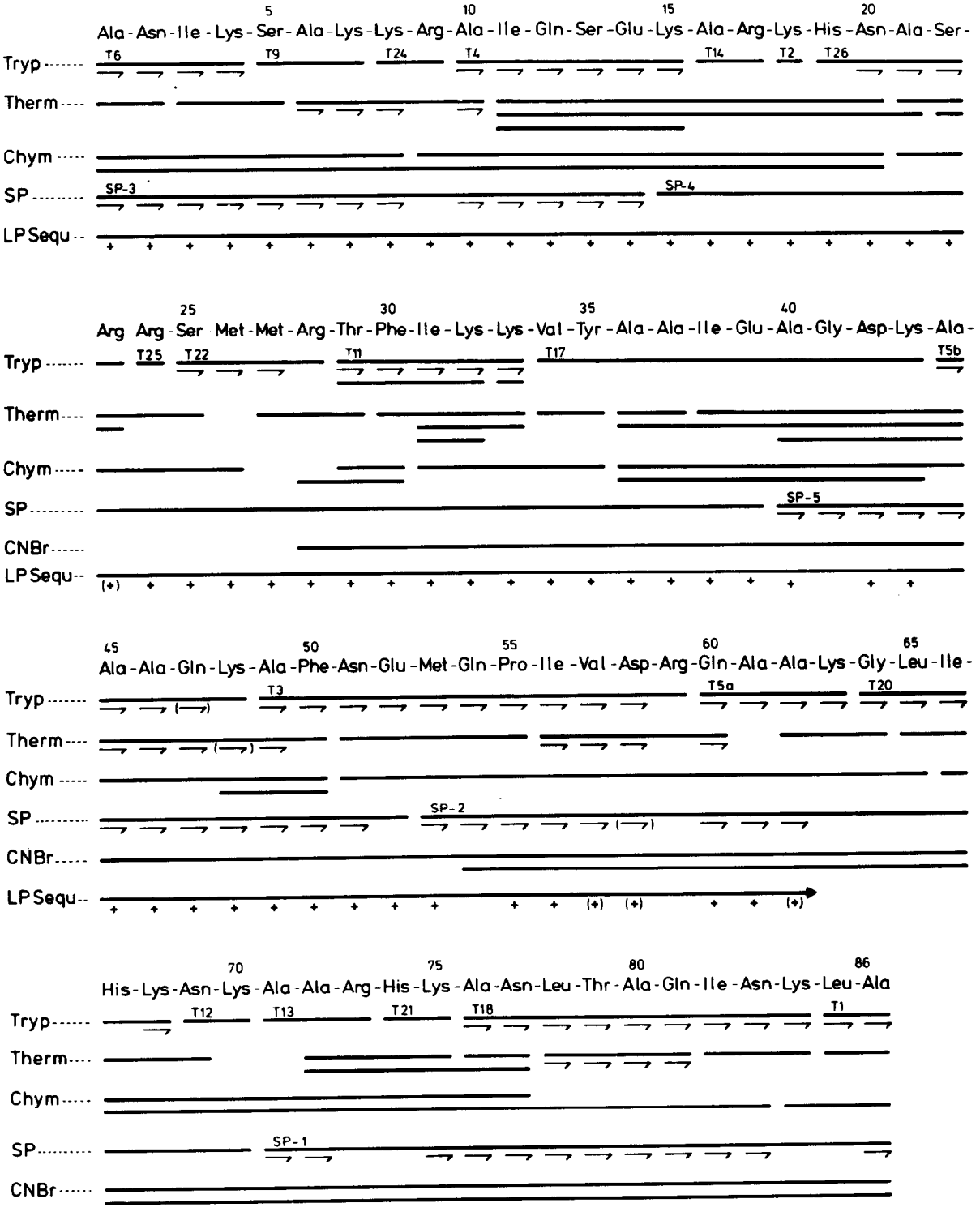
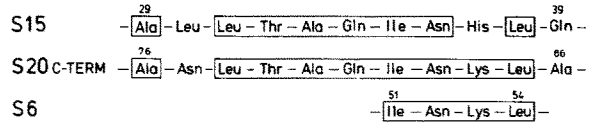

5
10
15
20
Ala - Asn - Ile - Lys - Ser - Ala - Lys - Lys - Arg - Ala - Ile - Gln - Ser - Glu - Lys - Ala - Arg - Lys - His - Asn - Ala - Ser -
Tryp
T6
T9
T24
T4
T14
T2
T26
Therm
Chym
SP
SP-3
SP-4
LP Sequ
25
30
35
40
Arg - Arg - Ser - Met - Met - Arg - Thr - Phe - Ile - Lys - Lys - Val - Tyr - Ala - Ala - Ile - Glu - Ala - Gly - Asp - Lys - Ala -
T25
T22
T11
T17
T5b
CNBr
SP-5
45
50
55
60
65
Ala - Ala - Gln - Lys - Ala - Phe - Asn - Glu - Met - Gln - Pro - Ile - Val - Asp - Arg - Gln - Ala - Ala - Lys - Gly - Leu - Ile -
T3
T5a
T20
SP-2
70
75
80
86
His - Lys - Asn - Lys - Ala - Ala - Arg - His - Lys - Ala - Asn - Leu - Thr - Ala - Gln - Ile - Asn - Lys - Leu - Ala
T12
T13
T21
T18
T1
SP-1

Fig.1




| | protein | positions | protein | positions |
|---|---|---|---|---|
| Ala-Lys-Lys-Arg | S20 | 6-9 | S16 | 11-14 |
| Ser-Glu-Lys-Ala | S20 | 13-16 | L23 | 17-20 |
| Glu-Lys-Ala-Arg | S20 | 14-17 | S5 | middle part |
| Ala-Ser-Arg-Arg | S20 | 21-24 | L25 | 16-19 |
| Glu-Ala-Gly-Asp | S20 | 39-42 | S6 | 125-128 |
| Asn-Lys-Ala-Ala | S20 | 69-72 | S8 | 20-23 |
| Ala-Arg-His-Lys | S20 | 72-75 | L20 | 11-14 |

Fig.2. Sequence homologies of ribosomal protein S20 with other proteins from *E. coli* ribosomes.

here, give well overlapping bridges in good agreement to each other, and the results are summarized in fig.1.

Comparison of the sequence of S20 with the other ribosomal proteins of known structure, cf. [29], revealed several similarities which are summarized in fig.2. As can be seen from this figure, an identical hexapeptide is present in ribosomal protein S15 [30] and in the C-terminal region of protein S20. Both proteins bind specifically to the 16S RNA, although at different regions [8,9]. It remains to be seen, whether the occurrence of the identical hexapeptide in proteins S15 and S20 has a biological meaning, since it was shown recently [3], that identical peptides of that length can occur in two proteins on a random basis.

The secondary structure of protein S20 was predicted using the method of Burgess et al. [32]. α-helix regions can be predicted for positions 5–16; 36–54; 58–62; and 70–82. β-Sheet structure was not predicted and bends were calculated for positions 17–18; 20–24; 26–29; 63–64; and 68–69. Alternatively, following the rules of Chou et al. [33,34] the following secondary structure is predicted: α-helix for positions 5/7–16; 32/33–52/54; 57/58–62/63 and 65–82. β-Sheet structures were obtained for positions 26–31 and positions 53–57 and are likely for positions 77–82. The high content of α-helical regions in protein S20 is remarkable; both predictions gave more than 50% α-helix content.

## Acknowledgements

We thank Dr G. R. Drapeau for a gift of *Staphylococcus aureus* protease and Mr Horst Graffunder and Mr Alfred Franz for excellent technical assistance. We thank Dr E. Dabbs for kindly reading the manuscript.

## References

[1] Mizushima, S. and Nomura, M. (1970) Nature (London) 226, 1214–1218.
[2] Schaup, H. W., Green, M. and Kurland, C. G. (1970) Mol. Gen. Genet. 109, 193–205.
[3] Stöffler, G., Daya, L., Rak, K. H. and Garrett, R. A. (1971) J. Mol. Biol. 62, 411–414.
[4] Schaup, H. W., Green, M. and Kurland, C. G. (1971) Mol. Gen. Genet. 112, 1–8.
[5] Garrett, R. A., Rak, K. H., Daya, L. and Stöffler, G. (1971) Mol. Gen. Genet. 114, 112–124.
[6] Zimmermann, R. A., Muro, A., Fellner, P., Ehresmann, C. and Branlant, C. (1972) Proc. Nat. Acad. Sci. USA 69, 1282–1286.
[7] Held, W. A., Ballou, B., Mizushima, S. and Nomura, M. (1974) J. B. C. 249, 3103–3111.
[8] Zimmermann, R. A., Mackie, G. A., Muto, A., Garrett, R. A., Ungewickell, E., Ehresmann, C., Ebel, J. P. and Fellner, P. (1975) Nucleic Acids Res. 2, 279–302.
[9] Ungewickell, E., Garrett, R. A., Ehresmann, C., Stingler, P. and Fellner, P. (1975) Europ. J. Biochem. 51, 165–180.
[10] Daya-Grosjean, L., Reinholt, J., Pongs, O. and Garrett, R. A. (1974) FEBS Lett. 44, 253–256.
[11] Morrison, C. A., Garrett, R. A., Zeichhardt, H. and Stöffler, G. (1973) Mol. Gen. Genet. 127, 359–368.
[12] Tischendorf, G. W., Zeichhardt, H. and Stöffler, G. (1975) Proc. Nat. Acad. Sci. USA 72, 4820–4824.
[13] Lelong, J. C., Gros, D., Gros, F., Bollen, A., Maschler, R. and Stöffler, G. (1974) Proc. Nat. Acad. Sci. USA 71, 248–252.

Fig.1. Primary structure of ribosomal protein S20. (Tryp) tryptic digestion; (Therm) Thermolysin digestion; (Chym) chymotryptic digestion; (SP) hydrolysis with *Staphylococcus aureus* protease at pH 4.0; (LP Sequ.) Edman degradation of intact protein in an improved Beckman sequenator including automatic conversion reaction; (+) means unambiguous identification of the released phenylthiohydantoin derivatives by thin-layer chromatography; (——→) means unambiguous identification with solid phase Edman degradation.

[14] Böck, A., Ruffler, D., Piepersberg, W. and Wittmann, H. G. (1974) Mol. Gen. Genet. 134, 325–332.
[15] Hindennach, I., Kaltschmidt, E. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 12–16.
[16] Kaltschmidt, E. and Wittmann, H. G. (1970) Anal. Biochem. 36, 401–412.
[17] Houmard, J. and Drapeau, G. R. (1972) Proc. Nat. Acad. Sci. USA 69, 3506–3509.
[18] Bitar, K. G. and Wittmann-Liebold, B. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1343–1352.
[19] Stadler, H. and Wittmann-Liebold, B. (1976) Europ. J. Biochem. 66, 49–56.
[20] Wittmann-Liebold, B. (1971) Hoppe-Seyler's Z. Physiolog. Chem. 352, 1705–1714.
[21] Wittmann-Liebold, B., Greuer, B. and Pannenbecker, R. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1977–1979.
[22] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[23] Previero, A., Derancourt, J., Coletti-Previero, M.-A. and Laursen, R. A. (1973) FEBS Lett. 33, 135–138.
[24] Wittmann-Liebold, B. and Lehmann, A. (1975) in: Solid phase methods in protein sequence analysis (R. A. Laursen, ed.) pp. 81–90. Pierce Chemical Company, Rockford, USA.
[25] Kaltschmidt, E. (1971) Analyt. Biochem. 43, 25–31.
[26] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[27] Wittmann, H. G. and Wittmann-Liebold, B. (1974) in: Ribosomes, pp. 115–140, Cold Spring Harbor Monograph Series.
[28] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[29] Wittmann-Liebold, B. and Dzionara, M. (1976) FEBS Lett. 61, 14–19.
[30] Morinaga, T., Funatsu, G., Funatsu, M. and Wittmann, H. G. (1976) FEBS Lett. 64, 307–309.
[31] Wittmann-Liebold, B. and Dzionara, M. (1976) FEBS Lett. 65, 281–283.
[32] Burgess, A. W., Ponnuswamy, P. K. and Scheraga, H. A. (1974) Israel J. Chem. 12, 239–286.
[33] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 211–222.
[34] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 222–245.