THE PRIMARY STRUCTURE OF RIBOSOMAL PROTEIN S7 FROM *E. COLI* STRAINS K AND B

Joseph REINBOLT and Denis TRITSCH
Laboratoire de Biochimie de l'Institut de Biologie Moléculaire et Cellulaire du CNRS, F-67000 Strasbourg, France

and

Brigitte WITTMANN-LIEBOLD
Max-Planck-Institut für Molekulare Genetik, Abt. Wittmann, D-1000 Berlin-Dahlem, Germany

Received 3 May 1978

1. Introduction

Protein S7 from the small subunit of *Escherichia coli* ribosomes has been reported to interact specifically with 16 S RNA (reviewed [1]). The proportion of the RNA involved in the binding has been identified as part of the 3′-end [2–4]. On the other hand the part of the protein that binds to the RNA has been investigated by covalently linking S7 to the RNA. Protein S7 can be crosslinked to 16 S RNA by ultraviolet irradiation of the small subunit [5]. Ultraviolet irradiation experiments have also been made on an in vitro complex between protein S7 and 16 S RNA. The results indicated that at least four peptides were crosslinked to the RNA [6]. It can be expected that knowledge of the primary structure of protein S7 will contribute to a deeper understanding of the molecular mechanisms of protein–RNA interaction.

It has been shown that *E. coli* strains differ in their S7 proteins [7–11]. Ribosomes from strains B, C and MRE600 contain a protein S7 which differs extensively in size, charge, amino acid composition and immunological properties from the protein S7 in strain K [11]. The amino acid sequences of protein S7 isolated from *E. coli* strain B and K are reported here. Protein S7B consists of 153 amino acids and protein S7K contains 177 amino acids. This difference of 24 amino acids between S7B and S7K occurs at the C-terminal end of the protein chain.

2. Materials and methods

Protein S7 was provided by Professor H. G. Wittmann and was isolated from 30 S ribosomal subunits of *E. coli* K and B as in [12]. Its purity was determined by two-dimensional gel electrophoresis [13].

2.1. *Enzymic digestions*

Protein S7 was digested with the following endoproteases: TPCK-trypsin; *Staphylococcus aureus* protease [14]; thermolysin; and pepsin. Digestions were also performed with the exopeptidases: carboxypeptidases A and B; and carboxypeptidase C according to [15].

2.2. *Specific cleavages*

Specific cleavages were performed:

(i) At arginine residues by modification of the lysine residues with ETPA (exo-*cis*-3,6-endoxo-Δ^4-tetrahydrophthalic anhydride) [16].

(ii) At methionine residues by treatment with BrCN in 70% formic acid for 20 h at room temperature in the dark [17].

(iii) At the tryptophan residues by treatment with BNPS-skatole (2-(2-nitrophenylsulphenyl)-3-methyl-3′bromoindolenine) [18].

2.3. *Isolation of peptides*

The peptides obtained with the different endopro-

teases were separated by the fingerprinting method on cellulose thin-layer plates and extracted for analysis of their amino acid composition [19]. For preparative purposes peptides were isolated on a microcolumn (0.3 X 10 cm) of Dowex M 71 at 50°C using pyridine–formate gradients [20,21] or by gel filtration on Sephadex G50 superfine (1 X 140 cm) in 10% acetic acid, followed, when necessary, by chromatography on cellulose thin-layer sheets. In a few cases combination of all the methods was necessary.

2.4. *Amino acid analyses*

Amino acid analyses of the peptides were performed on a Durrum D-500 amino acid analyser (Palo Alto). The presence of tryptophan was investigated by amino acid analysis after hydrolysis with methane sulfonic acid containing 0.2% 3-(2-aminoethyl)indole [22] and by spraying of fingerprint plates with *p*-dimethylaminobenzaldehyde.

2.5. *Sequence determination*

Sequence determination of the peptides were performed either by the combined manual dansyl-Edman procedure [23,24] or by automatic Edman degradation in a solid-phase sequenator with attachment of the C-terminal group to amino-polystyrene resin by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide [25–27]. A few large peptides and the N-terminal part of S7 were sequenced by the automatic Edman degradation [28] in an improved Beckman sequenator [29].

3. Results and discussion

3.1. *Amino acid composition*

It is known that the molecular weight of S7K is higher than that of S7B by approx. 15% [11]. We have determined that this difference occurs at the C-terminal region [30]. Protein S7K consists of 177 amino acid residues and has mol. wt 19 732, calculated from the sequence. Its amino acid composition is the following: Asp5, Asn7, Thr5, Ser15, Glu14, Gln5, Pro7, Gly10, Ala22, Val16, Met4, Ile5, Leu16, Tyr4, Phe5, His3, Lys14, Arg18, Trp2. Protein S7B is shorter by 24 amino acids than is S7K, and consists of 153 amino acids with mol. wt 17 131, i.e., 2601 daltons less than that of S7K. Its amino acid composition is as follows: Asp5, Asn6, Thr5, Ser10, Glu14, Gln3, Pro6, Gly8, Ala19, Val16, Met4, Ile5, Leu12, Tyr3, Phe4, His2, Lys13, Arg17, Trp1. No cysteine residue could be detected, either in S7B or in S7K. The high number of basic residues as compared to acidic ones (35 versus 19 for S7K and 32 versus 20 for S7B) is consistent with the high isoelectric points of S7K and S7B [31].

3.2. *Strategy of the sequence determination*

The strategy of the sequence determination was based mainly on two types of cleavage of the protein:

(i) In order to get small peptides, protein S7 was digested with endoproteases, as shown in fig.1. The principal procedure used to determine the sequences of the peptides obtained by these digestions was the automatic Edman degradation technique in a solid-phase sequenator. Sometimes we also used the manual dansyl-Edman technique.

(ii) Larger peptides were obtained by two chemical methods for specific cleavage of the protein chain:

1. By treatment of the two protein chains with the reagent BNPS-skatole, which cleaves after tryptophan residues, it was found that treatment of protein S7K resulted in one peptide more than the corresponding treatment of S7B. We determined that this extra peptide is situated at the C-terminal end of protein S7K [30].

Fig.1. The primary structure of proteins S7K and S7B. *Abbreviations:* Tryp, tryptic peptides; LSQ, Edman degradation in an improved Beckman sequenator including automatic conversion reaction; BrCN, peptides resulting from BrCN cleavage; SP, peptides from digestion with *Staphylococcus aureus* protease; Peps, peptides from digestion with pepsin; Th, peptides from digestion with thermolysin; ETPA, peptides obtained from a tryptic digest after blocking the amino groups by ETPA; BNPS, peptides resulting from BNPS-skatole cleavage; SSQ, sequence elucidated by solid-phase sequenator. The peptides specific to S7B or S7K are mentioned as 'B' or 'K'. All the other peptides are obtained from both proteins.

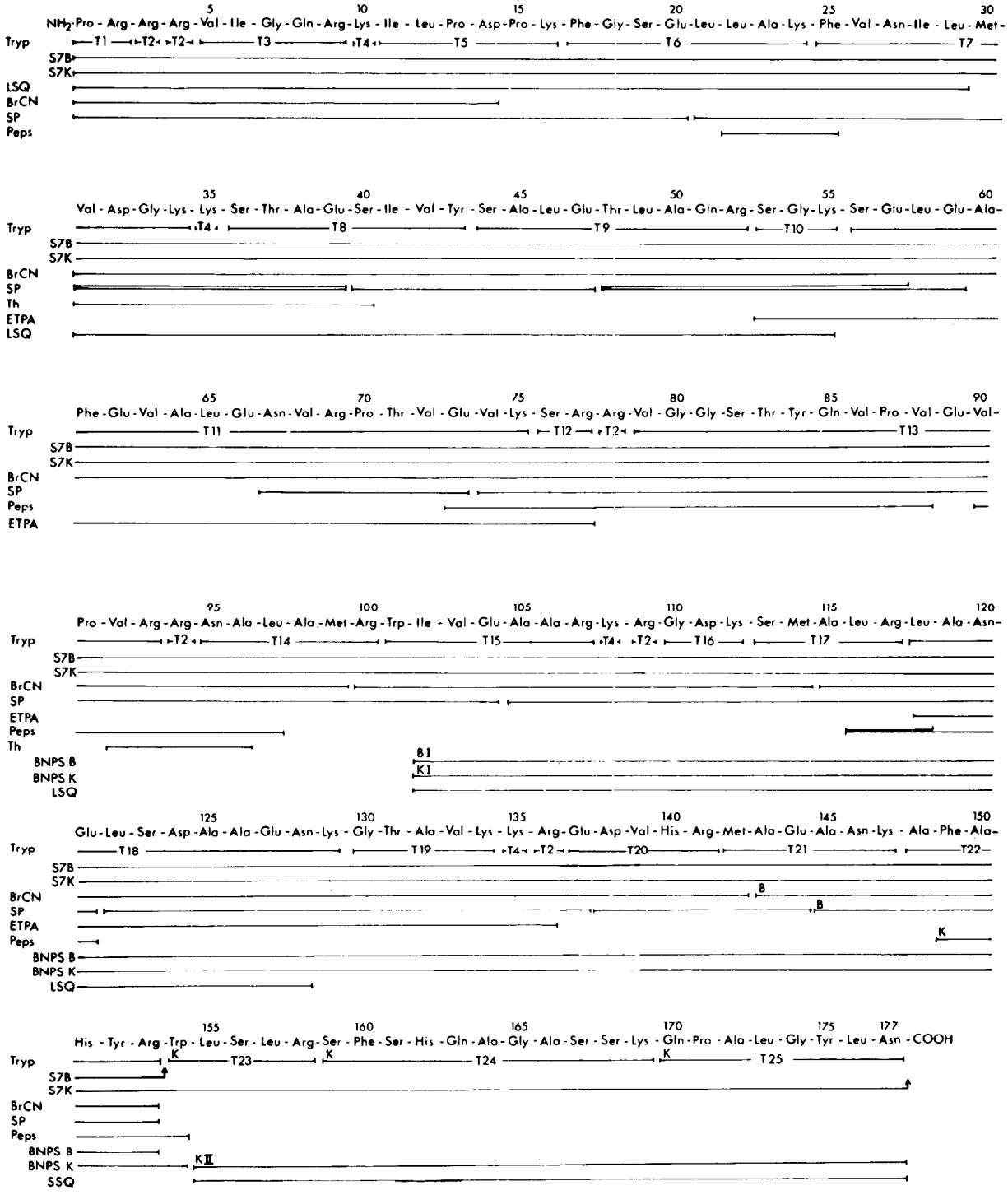
NH2-Pro - Arg - Arg - Arg - Val - Ile - Gly - Gln - Arg - Lys - Ile - Leu - Pro - Asp - Pro - Lys - Phe - Gly - Ser - Glu - Leu - Leu - Ala - Lys - Phe - Val - Asn - Ile - Leu - Met-
Tryp
S7B
S7K
LSQ
BrCN
SP
Peps
Val - Asp - Gly - Lys - Lys - Ser - Thr - Ala - Glu - Ser - Ile - Val - Tyr - Ser - Ala - Leu - Glu - Thr - Leu - Ala - Gln - Arg - Ser - Gly - Lys - Ser - Glu - Leu - Glu - Ala-
Tryp
S7B
S7K
BrCN
SP
Th
ETPA
LSQ
Phe - Glu - Val - Ala - Leu - Glu - Asn - Val - Arg - Pro - Thr - Val - Glu - Val - Lys - Ser - Arg - Arg - Val - Gly - Gly - Ser - Thr - Tyr - Gln - Val - Pro - Val - Glu - Val-
Tryp
S7B
S7K
BrCN
SP
Peps
ETPA
Pro - Val - Arg - Arg - Asn - Ala - Leu - Ala - Met - Arg - Trp - Ile - Val - Glu - Ala - Ala - Arg - Lys - Arg - Gly - Asp - Lys - Ser - Met - Ala - Leu - Arg - Leu - Ala - Asn-
Tryp
S7B
S7K
BrCN
SP
ETPA
Peps
Th
BNPS B
BNPS K
LSQ
Glu - Leu - Ser - Asp - Ala - Ala - Glu - Asn - Lys - Gly - Thr - Ala - Val - Lys - Lys - Arg - Glu - Asp - Val - His - Arg - Met - Ala - Glu - Ala - Asn - Lys - Ala - Phe - Ala-
Tryp
S7B
S7K
BrCN
SP
ETPA
Peps
BNPS B
BNPS K
LSQ
His - Tyr - Arg - Trp - Leu - Ser - Leu - Arg - Ser - Phe - Ser - His - Gln - Ala - Gly - Ala - Ser - Ser - Lys - Gln - Pro - Ala - Leu - Gly - Tyr - Leu - Asn - COOH
Tryp
S7B
S7K
BrCN
SP
Peps
BNPS B
BNPS K
SSQ

Fig.1

2. After cleavage of protein S7 with BrCN we were able to isolate a large peptide (position 31–99) by gel filtration on Sephadex G50 superfine. The sequence of this peptide was studied in an improved Beckman sequenator after reaction with Braunitzer's reagent IV [32].

3. It should be noted that the peptide bond Asp–Pro (position 14–15) appears to be extremely labile as it was split under the acidic conditions used by BrCN treatment. Full details of the sequence determination will be given elsewhere [33].

3.3. *Characteristics of the sequence*

There are many repeated sequences. Thirty seven dipeptides are present up to 5 times within the sequence of S7K, e.g., the dipeptide Ala–Leu occurs in the following positions: 45–46, 64–65, 96–97, 115–116, 172–173. Five tripeptides occur two times in the sequence: Arg–Arg–Val in positions 3–5 and 77–79; Ser–Glu–Leu in 19–21 and 56–58; Ala–Leu–Glu in 45–47 and 64–66; Val–Glu–Val in 72–74 and 88–90; Val–Pro–Val in 86–88 and 90–92. Because of the high number of basic residues (35 out of 177 for S7K and 32 out of 153 for S7B) it is not surprising to find groups of basic amino acids, e.g., in positions 2–10, 75–78, 107–112 and 134–141.

3.4. *Secondary structure of S7*

The secondary structure was predicted using computer programmes based on four different methods [34–38]. According to these calculations protein S7K contains about 38% α-helical regions, 3% β-structure and 13% turns (or loops), and protein S7B contains about 45% α-helical regions, 3% β-structure and 12% turns (or loops).

3.5. *Homologies of sequences*

The primary structure of S7K, of S7B and of other ribosomal proteins of known sequence, were compared by a computer programme. One pentapeptide of S7K and S7B: Thr–Val–Glu–Val–Lys (position 71–75) was identical with one pentapeptide of protein L19 (position 24–28). Thirteen tetrapeptides of S7K were found to be identical with tetrapeptides belonging to proteins S4, S8, S9, S11, S13, L1, L6, L7/L12, L11, L15, L16, L25 and L27. In the case of S7B ten tetrapeptides were found to be identical with tetrapeptides of the proteins S4, S8, S11, S13, L1, L6, L7/L12, L15, L16, L25 and L27. This degree of homology can be explained on a random basis [39].

Acknowledgements

We thank Mrs C. Rether and Mr A. Lehmann for excellent technical assistance.

References

[1] Zimmermann, R. A. (1974) in: Ribosomes, Monograph Series, pp. 225–269, Cold Spring Harbor Laboratory, Long Island, NY.
[2] Muto, A., Ehresmann, C., Fellner, P. and Zimmermann, R. A. (1974) J. Mol. Biol. 86, 411–432.
[3] Zimmermann, R. A., Mackie, G. A., Muto, A., Garrett, R. A., Ungewickell, E., Ehresmann, C., Stiegler, P., Ebel, J. P. and Fellner, P. (1975) Nucleic Acid Res. 2, 279–302.
[4] Rinke, J., Yuki, A. and Brimacombe, R. (1976) Eur. J. Biochem. 64, 77–89.
[5] Möller, K. and Brimacombe, R. (1975) Mol. Gen. Genet. 141, 343–355.
[6] Ehresmann, B., Reinbolt, J., Backendorf, C., Tritsch, D. and Ebel, J. P. (1976) FEBS Lett. 67, 316–319.
[7] Leboy, P. S., Cox, E. C. and Flaks, J. G. (1964) Proc. Natl. Acad. Sci. USA 52, 1367–1374.
[8] Birge, E. A., Craven, G. R., Hardy, S. J. S., Kurland, C. G. and Voynow, P. (1969) Science 164, 1285–1286.
[9] Takata, R., Dekio, S., Otaka, E. and Osawa, S. (1969) Mol. Gen. Genet. 105, 113–121.
[10] Sypherd, P. S. (1969) J. Bacteriol. 99, 379–382.
[11] Kaltschmidt, E., Stöffler, G., Dzionara, M. and Wittmann, H. G. (1970) Mol. Gen. Genet. 109, 303–308.
[12] Hindennach, I., Stöffler, G. and Wittmann, H. G. (1971) Eur. J. Biochem. 23, 7–11.
[13] Kaltschmidt, E. and Wittmann, H. G. (1970) Anal. Biochem. 36, 401–412.
[14] Houmard, J. and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[15] Tscheche, H. and Kupfer, S. (1972) Eur. J. Biochem. 26, 33–36.
[16] Riley, M. and Perham, R. N. (1970) Biochem. J. 118, 733–739.
[17] Gross, E. and Witkop, B. (1962) J. Biol. Chem. 237, 1856–1860.
[18] Fontana, A. (1972) Methods Enzymol. 25, 419–423.
[19] Bitar, K. G. and Wittmann-Liebold, B. (1976) Hoppe Seyler's Z. Physiol. Chem. 356, 1343–1352.

[20] Chen, R. and Wittmann-Liebold, B. (1975) FEBS Lett. 52, 139–140.
[21] Heiland, I., Brauer, D. and Wittmann-Liebold, B. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 1751–1770.
[22] Liu, T.-Y. and Chang, Y. H. (1971) J. Biol. Chem. 246, 2842–2848.
[23] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[24] Chen, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. 357, 873–886.
[25] Previero, A., Derancourt, J., Coletti-Previero, M. A. and Laursen, R. A. (1975) FEBS Lett. 33, 135–138.
[26] Wittmann-Liebold, B. and Lehmann, A. (1975) in: Solid Phase Methods in Protein Sequence Analysis (Laursen, R. A. ed) pp. 81–90, Pierce Chemical Co., Rockford, IL.
[27] Reinbolt, J., Tritsch, D. and Wittmann-Liebold, B. (1977) in: Solid Phase Methods in Protein Sequence Analysis (Previero, A. and Coletti-Previero, M. A. eds) INSERM Symp. 5, pp. 209–218, Elsevier/North-Holland Biomedical Press, Amsterdam, New York.
[28] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[29] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[30] Tritsch, D., Reinbolt, J. and Wittmann-Liebold, B. (1977) FEBS Lett. 77, 89–93.
[31] Kaltschmidt, E. (1971) Anal. Biochem. 43, 25.
[32] Braunitzer, G., Schrank, B., Ruhfus, A., Petersen, S. and Petersen, U. (1971) Hoppe-Seyler's Z. Physiol. Chem. 352, 1730–1732.
[33] Reinbolt, J., Tritsch, D. and Wittmann-Liebold, B. (1978) in preparation.
[34] Burgess, A., Ponnuswamy, P. and Scheraga, H. (1974) Israel J. Chem. 12, 239–286.
[35] Chou, P., Adler, A. and Fasman, G. (1975) J. Mol. Biol. 96, 29–45.
[36] Nagano, K. (1977) J. Mol. Biol. 109, 251–274.
[37] Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327–356.
[38] Dzionara, M., Robinson, S. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.
[39] Wittmann-Liebold, B. and Dzionara, M. (1976) FEBS Lett. 65, 281–283.