The complete amino acid sequence of ribosomal protein H-S11 from the archaebacterium *Halobacterium marismortui*

Evelyn Arndt, Gabriele Breithaupt and Makoto Kimura

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittman, D-1000 Berlin 33, Dahlem, Germany

Received 18 October 1985

The complete amino acid sequence of ribosomal protein H-S11 from the extremely halophilic archaebacterium *Halobacterium marismortui* is presented. This has been achieved by the sequence analysis of peptides derived by enzymatic digestions with trypsin, pepsin and *Staphylococcus aureus* protease, as well as by chemical cleavage with cyanogen bromide and *o*-iodosobenzoic acid. The protein consists of 155 amino acid residues and has a molecular mass of 17545 Da. Comparison of this sequence with other ribosomal proteins by the computer programmes RELATE and ALIGN showed that the C-terminal two-thirds of H-S11 is homologous to the eubacterial ribosomal protein S15 and that the N-terminal one-third of H-S11 is possibly related to the N-terminal region of the eubacterial ribosomal protein S8. To explain this finding, possible genetic events during evolution, e.g. fusion or splitting of genes, are discussed.

(Halobacterium marismortui) *Halophilic ribosomal protein* *Evolution* *Amino acid sequence*

1. INTRODUCTION

Archaebacteria, postulated as a third phylogenetic kingdom of organisms [1], are bacteria living in extraordinary environments such as anaerobic conditions, high salt concentration, high temperature or extremely acidic media. Halophiles are well characterized archaebacteria with regard to ribosomal structure. A striking feature of the halobacterial ribosomes is that unlike all other organisms, most of their ribosomal proteins are acidic [2]. Furthermore, the ribosomes of halobacteria share eubacterial as well as eukaryotic features. The primary structure of the 16 S rRNAs and their gene organization in *Halobacterium cutirubrum* and *H. volcanii* are more similar to the eubacterial 16 S rRNAs than to the eukaryotic counterparts [3,4]. On the other hand, the structures of 5 S rRNA and the ribosomal 'A' protein are more related to eukaryotic than to eubacterial components [5,6].

To study the evolutionary relationship between halobacteria and other organisms, a number of ribosomal proteins from *H. cutirubrum* have been characterized by N-terminal sequence analysis [7]. These studies, however, did not reveal a significant sequence homology to eubacterial and eukaryotic ribosomal proteins. In view of the distant relationship between halobacteria and other organisms, sequence similarities may not have been conserved throughout their entire lengths, implying that the full sequence is necessary to find a sequence homology which is present only in small areas. Therefore, we have begun the complete amino acid sequence analyses of ribosomal proteins from *H. marismortui*. Our recently completed sequence analysis of the ribosomal protein H-S17 from *H. marismortui* has revealed that it is homologous to ribosomal proteins E-S9 from *Escherichia coli* and B-S9 from *Bacillus stearothermophilus* [8].

Here we describe the isolation and the complete amino acid sequence of the ribosomal protein H-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

S11 from *H. marismortui*, and compare it with all ribosomal proteins from other organisms thus far sequenced. This comparison revealed interesting structural homologies between this archaebacterial ribosomal protein and 2 ribosomal proteins from eubacteria.

2. EXPERIMENTAL

H. marismortui cells grown by the method of Mevarech et al. [9] were a gift from Mr A. Shevack from our institute. The preparation of the ribosomes and the zonal separation of ribosomal subunits have been described [10].

2.1. *Purification of ribosomal proteins*

The isolation of ribosomal proteins on DEAE-cellulose was performed as described in [7]. The extracted proteins were dissolved in the starting buffer (75 mM Tris-citrate, pH 7.9, containing 30% dimethylformamide) and dialyzed for 2 h against the starting buffer. The protein solution was applied to a DEAE-cellulose column (2 × 20 cm) which had been equilibrated with the starting buffer. The proteins were eluted with a linear gradient of 0–0.3 M KCl in the starting buffer using a gradient volume of 1 l. Fractions of 3 ml were collected and the proteins were detected by analysing 100 μl samples from every fifth fraction by SDS slab gel electrophoresis [11]. Proteins were concentrated after dialysis against 10% acetic acid and purified further by gel filtration on Sephadex G-75 superfine (2 × 180 cm) in 10% acetic acid.

2.2. *Polyacrylamide gel electrophoresis*

The procedure for the two-dimensional electrophoresis was essentially the same as that described by Strom and Visentin [12], except that the electrophoresis was run using a smaller chamber (10 × 10 cm) and the urea concentration was reduced from 8 to 6 M. The sample (in 50 μl sample buffer) containing 50–100 μg total protein was overlaid. For the first dimension the electrophoresis was carried out at 1.5 mA per gel until the dye migrated to the bottom and for the second dimension at 20 mA per gel for 24 h. Slab gel electrophoresis was carried out as described in [11] with an acrylamide concentration of 15%.

2.3. *Sequence determination*

The methods used for the determination of the primary structure of ribosomal protein in our laboratory have been described in [13]. The protein (1 mg) was subjected to digestions with trypsin, pepsin and *Staphylococcus aureus* protease. The protein was also cleaved with cyanogen bromide and *o*-iodosobenzoic acid [14]. All resulting peptides were separated by a combination of gel filtration and fingerprinting on thin-layer sheets. The amino acid analyses were carried out with a Durrum 500 amino acid analyzer or by an HPLC system using *o*-phthaldialdehyde as a derivatised reagent [15]. The amino acid sequences of the peptides were performed by the DABITC/PITC double coupling method [16], manual solid-phase method [17], or with the aid of a solid-phase sequencer (LKB) [18].

2.4. *Computer analysis*

The computer analysis for a comparison of the ribosomal proteins was performed with the aid of 2 programmes RELATE and ALIGN [19]. RELATE was used to assess the statistical significance of the relationship between 2 proteins. As scoring matrix the mutation data matrix was employed and the segment length was set at 20 residues. Furthermore, ALIGN was used to align the homologous proteins for maximal homology. As scoring matrix the mutation data matrix was used, and a break penalty of 20 was employed. The proteins compared with H-S11 are listed in table 1.

2.5. *Nomenclature*

The ribosomal protein H-S11 from *H. marismortui* is designated according to its position on two-dimensional gel electrophoresis. The prefix E (for *E. coli*) and B (for *B. stearothermophilus*) are added to discriminate between the ribosomal proteins from these 2 organisms.

3. RESULTS AND DISCUSSION

3.1. *Isolation of the protein*

Fig.1 shows a two-dimensional polyacrylamide gel electrophoresis pattern of *H. marismortui* 30 S proteins. The numbers were assigned by numbering the spots according to the position on the gel in the same manner as those of other prokaryotic organisms. In this system we routinely identified

Table 1

List of ribosomal proteins whose primary structures were compared with that of H-S11

Organisms	Protein	Reference
E. coli	all ribosomal proteins	[20,21]
B. stearothermophilus	S5, L30	[22]
	S9, L6, L9, L17, L27	[23]
	L1, L14, L15, L23, L24, L29	[24]
	L2	[25]
	L32	[26]
	S2, S4, S8, S12, S15, S16, S20, L5, L18	[a]
Sac. cerevisiae	S10	[27]
	S33	[28]
	S24, L46	[29]
	L3	[30]
	L29	[31]
	L17a,L25	[32]
	YL27, YL35, YL44	[33]
Rat (liver)	L37	[34]
	L39	[35]
	P2	[36]

[a] M. Kimura, J. Kimura and E. Arndt, unpublished

22 proteins. To isolate these proteins, the protein mixture was applied to a DEAE-cellulose column (2 × 20 cm) equilibrated with the starting buffer, and then the proteins were eluted with a linear gradient of 0–0.3 M KCl. The elution pattern as analysed by SDS slab gel electrophoresis of every fifth fraction is shown in fig.2. Appropriate fractions from this column were pooled and concentrated as described in section 2. The proteins were further purified by gel filtration on a Sephadex G-75 column (2 × 180 cm) in 10% acetic acid. The identity of each protein was established by two-dimensional gel electrophoresis. The yields were about 2–3 mg per purified protein when 5000 A_{260} units of 30 S subunits were used as starting material.

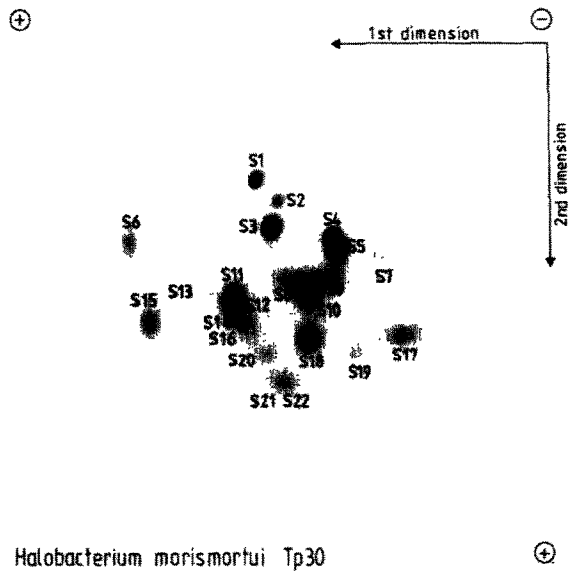

Fig.1. Two-dimensional gel electrophoresis of 30 S ribosomal proteins of *H. marismortui*. The proteins are designated according to their positions on the gel.

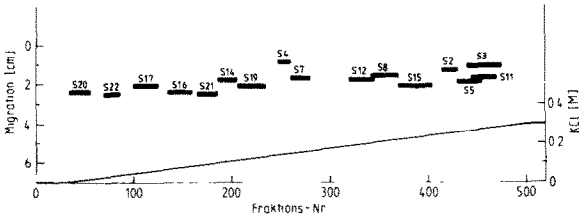

Fig.2. Ion-exchange column chromatography of *H. marismortui* 30 S ribosomal proteins on DEAE-cellulose. The proteins were eluted with a linear gradient of 0–0.3 M KCl in 75 mM Tris-citrate buffer, pH 7.9, containing 30% dimethylformamide. Bars indicate the positions of individual proteins. The distance of migration after SDS-polyacrylamide gel electrophoresis is given on the ordinate.

Protein H-S11 was eluted at 0.26 M KCl from the DEAE-cellulose column. The protein, purified further by gel filtration on Sephadex G-75, migrated as a single component in two-dimensional gel electrophoresis.

3.2. *Protein sequence determination*

The amino acid sequence of H-S11 as shown in fig.3 was determined by splitting the protein with trypsin, pepsin, *S. aureus* protease, cyanogen

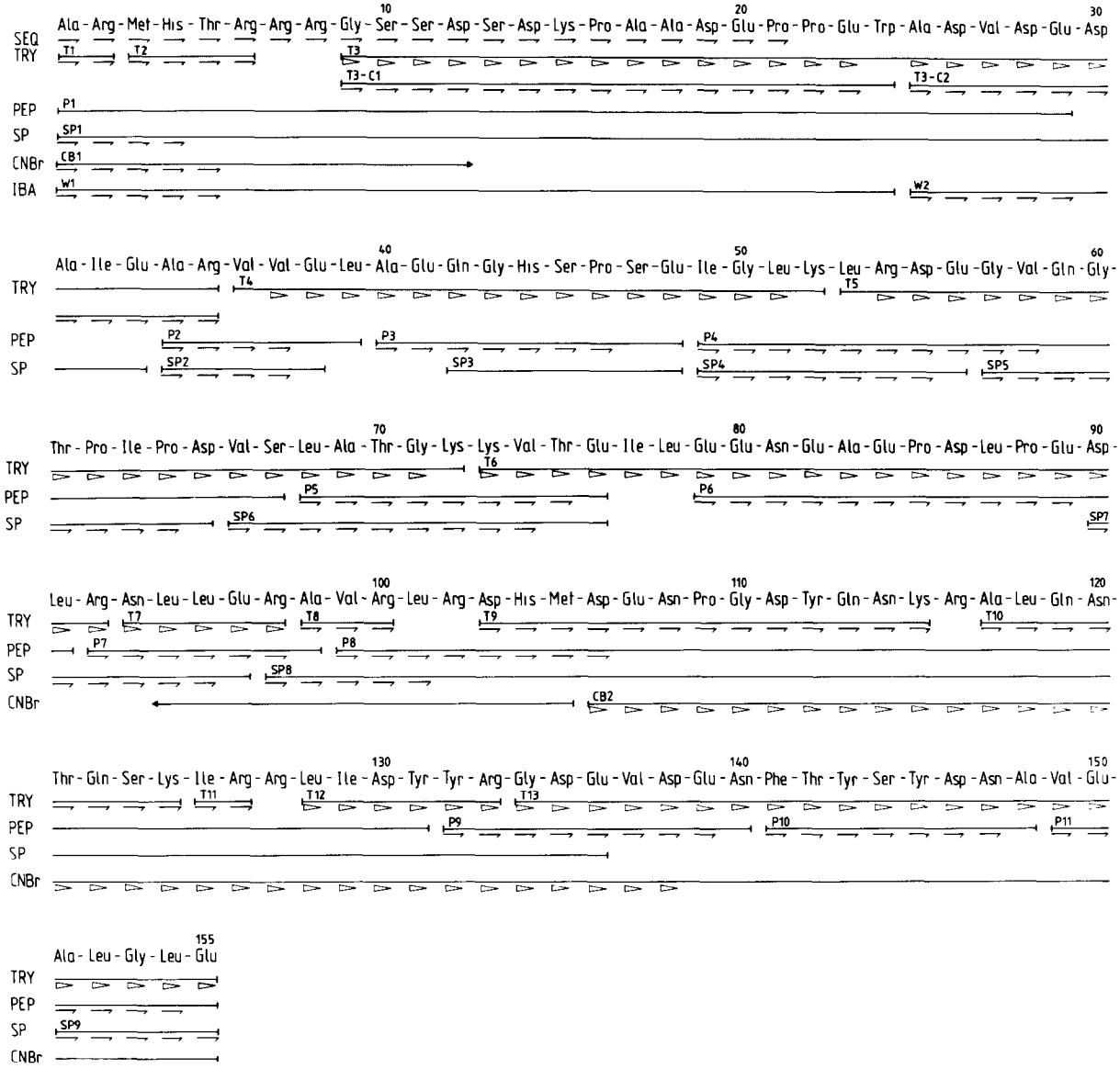

Fig.3. Amino acid sequence of protein H-S11 from *H. marismortui* sequence data on individual peptides are indicated as follows: ⟶, sequenced by the 4-*N,N*-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate double-coupling method; ▷, sequenced by the solid-phase procedure. SEQ indicates Edman degradation of intact protein. TRY, PEP, SP, CNBr and IBA indicate peptide derived from cleavage with trypsin, pepsin, *S. aureus* protease, cyanogen bromide and *o*-iodosobenzoic acid, respectively.

bromide and *o*-iodosobenzoic acid followed by sequencing the resulting peptides. The tryptic digestion produced 13 peptides which were isolated by peptide mapping. The complete sequence of all peptides except for T3 (pos. 9–35) was determined by the sequencing methods described above. The amino acid sequence of peptide T3 was obtained by isolation and sequence analysis of the chymotryptic peptides T3-C1 (pos. 9–24) and T3-C2 (pos. 25–35). The tryptophan residue at position 24 could not be identified clearly by the manual sequencing method. However, the

presence of tryptophan at this position was revealed by 2 findings: (i) the tryptophan analysis using dimethylaminobenzaldehyde on a tryptic peptide map showed the presence of tryptophan only in peptide T3; (ii) cleavage of the intact H-S11 by *o*-iodosobenzoic acid produced 2 fragments which were separated by Sephadex G-75 superfine (1 × 180 cm) gel filtration. The N-terminal sequences of the fragments thus obtained began with A-R-M and A-D-V, identical with the sequences at position 1–3 and 25–27, respectively. This shows that the tryptophan is located at position 24.

For the alignment of the tryptic peptides, suitable overlapping peptides were produced by digestion of the intact H-S11 with pepsin and *S. aureus* protease. The resulting peptides were separated by fingerprinting and sequenced. These results, together with the N-terminal sequence analysis of the intact protein, allowed us to align the tryptic peptides T1-T9 and T12-T13. However, the alignment of the peptides T9-T12 still remained unclear. This was finally established by sequencing the cyanogen bromide peptide CB2 (pos. 106–155). The cyanogen bromide treatment of H-S11 caused cleavage behind the methionine residue at position 105 and produced 2 peptides CB1 (pos. 1–105) and CB2 (pos. 106–155). These peptides were separated by gel filtration on Sephadex G-75 superfine and the C-terminal peptide was subjected to solid-phase sequencing. The amino acid sequence thus determined allowed the alignment of the remaining tryptic peptides. By combining all results the amino acid sequence of H-S11 was unambiguously established as shown in fig.3.

3.3. *Characterization of the protein*

The amino acid composition and the molecular mass of H-S11 as calculated from its primary structure are given in table 2. The composition derived from the amino acid sequence and that calculated from the amino acid analysis of the protein agree well as shown in table 2. The protein contains a high amount of acidic amino acids randomly distributed throughout the molecule. From the number of acidic residues and basic residues in H-S11 a net charge of −14 is calculated.

Another striking feature of the protein is the cluster of aromatic amino acid residues at the C-terminal region. H-S11 contains one phenylalanine and 5 tyrosine residues. Five of these aromatic residues (Phe-141, Tyr-131, Tyr-132, Tyr-143 and Tyr-145) are located at the C-terminus. Interestingly, clustering of aromatic amino acids at the C-terminus was also found [8] in the protein H-S17.

Table 2

Amino acid composition and M_r of protein H-S11 from *H. marismortui*

Amino acid	Protein hydrolyte	Sequence
Asp/Asn	24.0	17/7
Thr	5.6	6
Ser	7.9	8
Glu/Gln	25.1	20/5
Pro	8.8	9
Gly	9.9	9
Ala	13.4	13
Val	8.1	9
Met	2.2	2
Ile	6.2	6
Leu	13.4	14
Tyr	5.9	5
Phe	1.1	1
His	2.4	3
Lys	4.9	6
Arg	13.3	14
Trp[a]	+	1
Total		155
M_r		17545

[a] Presence of tryptophan was determined by staining with *p*-dimethylaminobenzaldehyde

3.4. *Sequence comparison*

As mentioned above, the N-terminal sequences of 10 ribosomal proteins from *H. cutirubrum* 30 S subunits were reported by Yaguchi et al. [7]. The comparison of the present result with those obtained for *H. cutirubrum* shows that protein H-S11 from *H. marismortui* is structurally homologous to the *H. cutirubrum* ribosomal protein H-S11. The complete amino acid sequence of H-S11 was compared with the sequences of the ribosomal proteins listed in table 1 using the computer programme RELATE. The segment comparison score thus obtained showed that H-S11 exhibits a significant homology to the eubacterial ribosomal proteins E-S15, B-S15, E-S4 and E-S8 (fig.4) but not

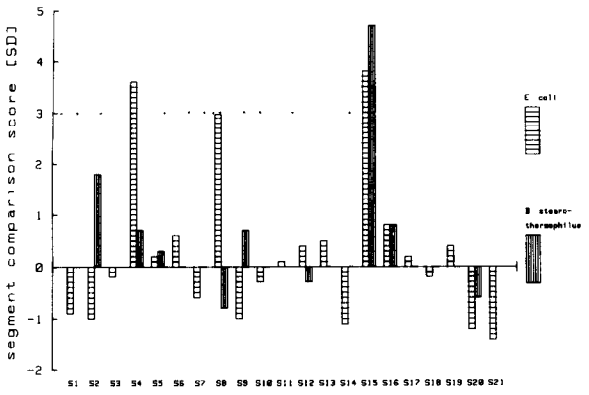

Fig.4. The segment comparison scores of H-S11 with ribosomal proteins from *E. coli* and *B. stearothermophilus* obtained by the program RELATE. As scoring matrix the mutation data matrix was used and the segment length was set at 20 residues. The segment comparison score 3.0 SD unit which suggests a possible relatedness of the sequences is indicated by the dotted line.

to the eukaryotic proteins. Accordingly, comparisons with these 4 proteins were further examined using the computer programme ALIGN. First H-S11 was aligned with E-S15 and B-S15 for maximum homology (fig.5); this yielded alignment scores of 8.53 and 8.38 for E-S15 and B-S15, respectively. These scores are statistically significant and indicate that the C-terminal region (pos. 65–155) of H-S11 is related to the eubacterial protein S15.

As shown in fig.4 protein H-S11 also seems to be related to the *E. coli* proteins E-S4 and E-S8. Therefore, these proteins as well as B-S4 and B-S8 were compared with the programme ALIGN. The comparison showed that the N-terminal 60 residues of H-S11 can be aligned with E-S4 and E-S8 as illustrated in fig.6. Proteins E-S4 and E-S8 contain short segments which seem to be homologous to the N-terminal part of H-S11, although the alignment scores (2.23 for E-S4 and 2.32 for E-S8) indicate less similarity than that between H-S11 and the eubacterial protein S15. Although it is not excluded that the relatedness of H-S11 with E-S4 or E-S8 obtained by the programme RELATE might be fortuitous it is possible that particularly S8 is related to the N-terminal part of H-S11 for several relevant observations. First, the proteins E-S8 and E-S15 bind to neighbouring sequences that lie in the middle of the *E. coli* 16 S rRNA [37]. Second, E-S8 and E-S15 can bind specifically to the 16 S rRNA from *H. cutirubrum*, which is related to *H. marismortui*, and the protein binding sites on the *H. cutirubrum* 16 S rRNA are homologous to those on the cognate 16 S rRNA [38]. Third, the protein E-S8 is composed of 2 structural domains [39] one

```
                    10                  20                  30                  40                  50
HS11  A R M H T R R R G S S D S D K P A A D E P P E W A D V D E D A I E A R V V E L A E Q G H S P S E I G
BS15
ES15

                    60                  70                  80                  90                 100
HS11  L K L R D E G V Q G T P I P D V S L A T G K K V T E I L E E N E A E P D L P E - D L R N L L E R A V
BS15                      A L T Q E R K R E I I E Q F K V H E N D T G S P E V Q I A I L T E Q I N
ES15                      S L S T E A T A K I V S E F G R D A N D T G S T E V Q V A L L T A Q I N

                   110                 120                 130                 140                 150
HS11  R L R D H M D E N P G D Y Q N K R A L Q N T Q S K I R R L I D Y Y R G D E V D E N F T Y S Y D N A V
BS15  N L N E H L R V R K K D H H S R R G L L K M V G K R R R L L A Y L R N K D V A - - - - - R Y R E I V
ES15  H L Q G H F - A E K K D H H S R R G L L R M V S Q R R K L L D Y L K R K D V A - - - - - R Y T Q L I

HS11  E A L G L E
BS15  E K L G L R R
ES15  E R L G L R R
```

Fig.5. Comparison of the amino acid sequences of H-S11 from *H. marismortui* with those of S15 from *B. stearothermophilus* (B-S15) and *E. coli* (E-S15). A maximum homology was obtained by the programme ALIGN with a break penalty of 20 residues. Identical residues are enclosed in boxes.

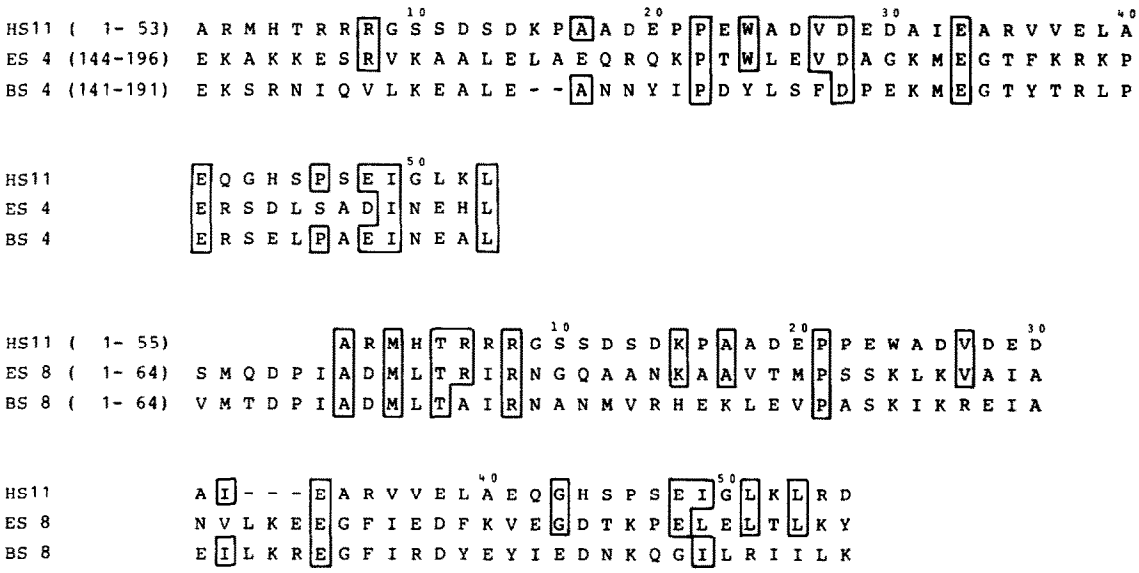

Fig.6. Alignments of the N-terminal part of H-S11 from *H. marismortui* with the homologous sequences of E-S4 (144–196), B-S4 (141–191), E-S8 (1–64) and B-S8 (1–64). The programme ALIGN was used for a maximum homology with a condition identical to that described in fig.5. Identical residues are enclosed in boxes.

of which is nearly identical to the region which shows some homology to the N-terminal part of H-S11.

Two hypotheses concerning the genesis of H-S11 are considered. On the one hand, protein H-S11 may be related to the eubacterial S15 only, leaving a large N-terminal region which shows no homology to any other ribosomal proteins (fig.7a). On the other hand, the gene for H-S11 could have evolved by a fusion of the genes for the eubacterial protein S15 and for the N-terminal part of protein S8 (fig.7b). Alternatively, it is also possible that the ancestral gene for protein H-S11 has been split into 2 parts during evolution: one evolved into the gene for the eubacterial protein S15 and the other into that for the N-terminal region of S8. This splitting should have occurred in the evolutionary line leading to eubacteria after it had diverged from that leading to halobacteria.

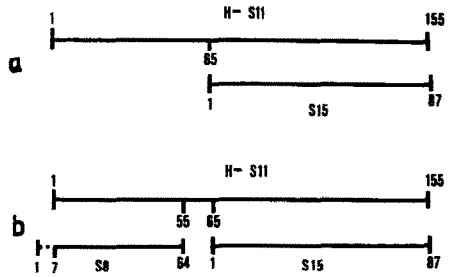

Fig.7. Two suggested models for the evolutional relationship between the halobacterial protein S11 (H-S11) and eubacterial proteins S15 and S8.

ACKNOWLEDGEMENTS

We would like to thank Dr H.G. Wittmann for his interest and encouragement, and Drs C. Pon and B. Wittmann-Liebold for valuable discussion and critical reading of the manuscript.

REFERENCES

[1] Woese, C.R. and Fox, G.E. (1977) Proc. Natl. Acad. Sci. USA 74, 5088–5090.
[2] Visentin, L.P., Chow, C., Matheson, A.T., Yaguchi, M. and Rollin, F. (1972) Biochem. J. 130, 103–110.
[3] Hui, I. and Dennis, P.P. (1985) J. Biol. Chem. 260, 899–906.
[4] Gupta, R., Lanter, J.M. and Woese, C.R. (1983) Science 221, 656–659.

[5] Matheson, A.T. and Yaguchi, M. (1982) Zentralbl. Bakteriol. Hyg., I. Abt. Orig. C3, 192–199.
[6] Hori, H. and Osawa, S. (1980) in: Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes (Osawa, S. et al. eds) pp.539–551, University of Tokyo Press, Tokyo.
[7] Yaguchi, M., Visentin, L.P., Zuker, M., Matheson, A.T., Roy, C. and Strom, A.R. (1982) Zentralbl. Bakteriol. Hyg., I. Abt. Orig. C3, 200–208.
[8] Kimura, M. and Langner, G. (1984) FEBS Lett. 175, 213–218.
[9] Mevarech, M., Leicht, W. and Werber, M.M. (1976) Biochemistry 15, 2383–2387.
[10] Shevack, A., Gewitz, H.S., Hennemann, B., Yonath, A. and Wittmann, H.G. (1985) FEBS Lett. 184, 68–71.
[11] Laemmli, U.K. (1970) Nature 227, 680–685.
[12] Strom, A.R. and Visentin, L.P. (1973) FEBS Lett. 37, 274–280.
[13] Wittmann-Liebold, B. and Kimura, M. (1984) in: Methods in Molecular Biology, vol.1, Proteins (Walker, J.M. ed.) pp.221–242, The Humana Press, New York.
[14] Fontana, A., Dalzoppo, D., Grandi, C. and Zambonin, M. (1983) Methods Enzymol. 91, 311–318.
[15] Ashman, K. and Bosserhof, A. (1985) in: Modern Methods in Analytical Protein Chemistry (Tschesche, H. ed.) De Gruyter, Heidelberg, in press.
[16] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[17] Chang, J.Y. (1979) Biochim. Biophys. Acta 578, 188–195.
[18] Wittmann-Liebold, B. and Lehmann, A. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, C. ed.) pp.49–72, Elsevier, Amsterdam, New York.
[19] Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, vol.5, suppl.3, National Biomedical Research Foundation, Washington, DC.
[20] Wittmann, H.G. (1982) Annu. Rev. Biochem. 51, 155–183.
[21] Wittmann-Liebold, B. (1984) Adv. Protein Chem. 36, 56–78.
[22] Kimura, M. (1984) J. Biol. Chem. 259, 1051–1055.
[23] Kimura, M. and Chow, C.K. (1984) Eur. J. Biochem. 139, 225–234.
[24] Kimura, M., Kimura, J. and Ashman, K. (1985) Eur. J. Biochem. 150, 491–497.
[25] Kimura, M., Kimura, J. and Watanabe, K. (1985) Eur. J. Biochem., in press.
[26] Tanaka, I., Kimura, M., Kimura, J. and Dijk, J. (1984) FEBS Lett. 166, 343–346.
[27] Leer, R.J., Van Raamsdonk-Duin, M.M.C., Molenaar, C.M.T., Cohen, L.H., Mager, W.H. and Planta, R.J. (1982) Nucleic Acids Res. 10, 5869–5878.
[28] Leer, R.J., Van Raamsdonk-Duin, M.M.C., Cornelissen, M.T.E., Cohen, L.H., Mager, W.H. and Planta, R.J. (1983) Nucleic Acids Res. 11, 7759–7768.
[29] Leer, R.J., Van Raamsdonk-Duin, M.M.C., Kraakmann, P., Mager, W.H. and Planta, R.J. (1985) Nucleic Acids Res. 13, 701–709.
[30] Schultz, L.D. and Friesen, J.D. (1983) J. Bacteriol. 155, 8–14.
[31] Käufer, N.F., Fried, H.M., Schwindinger, W.F., Jasin, M. and Warner, J.R. (1983) Nucleic Acids Res. 11, 3123–3135.
[32] Leer, R.J., Van Raamsdonk-Duin, M.M.C., Hagenboom, M.J.M., Mager, W.H. and Planta, R.J. (1984) Nucleic Acids Res. 12, 6685–6700.
[33] Itoh, T., Higo, K., Otaka, E. and Osawa, S. (1980) in: Genetics and Evolution of RNA Polymerase, tRNA and Ribosomes (Osawa, S. et al. eds) pp.609–624, University of Tokyo Press, Tokyo.
[34] Lin, A., McNally, J. and Wool, I.G. (1983) J. Biol. Chem. 258, 10664–10671.
[35] Lin, A., McNally, J. and Wool, I.G. (1984) J. Biol. Chem. 259, 487–490.
[36] Lin. A., Wittmann-Liebold, B. and Wool, I.G. (1982) J. Biol. Chem. 257, 9189–9197.
[37] Zimmermann, R.A. (1979) in: Ribosomes (Chambliss, G. et al. eds) pp.135–169, University Park Press, Baltimore.
[38] Thurlow, D.L. and Zimmermann, R.A. (1982) in: Archaebacteria (Kandler, O. ed.) pp.347, Fischer, Stuttgart.
[39] Paterakis, K. and Littlechild, J. (1982) FEBS Lett. 149, 328–333.