The complete amino acid sequence of the ribosomal protein HS3 from *Halobacterium marismortui*, an archaebacterium

Makoto Kimura and Gabriele Langner

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33 (Dahlem), Germany

Received 23 July 1984

The complete amino acid sequence of the ribosomal protein HS3 from *Halobacterium marismortui* has been determined, which is, to our knowledge, the first complete protein sequence from an archaebacterial ribosome. Comparison of the primary structure of HS3 with all sequenced ribosomal proteins from other organisms revealed that HS3 is significantly homologous with protein S9 from the gram-negative *Escherichia coli* and the gram-positive *Bacillus stearothermophilus*, having 27 and 37% identical residues, respectively. Based on these and other results it is possible that some ribosomal components of archaebacteria are related to eubacteria, especially gram-positive bacteria, and others to eukaryotes.

Amino acid sequence Halobacterium marismortui *Ribosomal protein* *Sequence comparison*

1. INTRODUCTION

In addition to the two kingdoms of organisms, the eukaryotes and the eubacteria, authors in [1] have deduced the existence of the third kingdom, the archaebacteria, from comparative studies on partial sequences of the 16 S ribosomal RNAs. Since that time, a phylogenetical relationship among the 3 kingdoms has been examined with a comparative study on cellular components. From the complete sequence and secondary structure of the *Halobacterium volcanii* 16 S ribosomal RNA, it has been reported that this RNA is closer to the eubacterial 16 S ribosomal RNA than to the eukaryotic 16 S-like ribosomal RNA, but that it also shows specific sequence similarity to its eukaryotic counterpart. From this, it was concluded that the 3 kingdoms are almost equivalent to each other [2].

In contrast, it has been suggested from comparisons of the acidic ribosomal 'A' proteins from several species that the archaebacteria are more closely related to the eukaryotes than to the eubacteria, and a phylogenetic tree was made showing divergence of the archaebacteria from the eukaryotes subsequent to the divergence of the eubacteria and the eukaryotes [3]. Furthermore, a comparison of the primary and secondary structures of ribosomal 5 S RNAs also supported the conclusion derived from the comparative study on acidic ribosomal 'A' proteins [4].

To obtain a more detailed view of the evolutionary relationships between archaebacteria and the other kingdoms, we have extended our own comparative studies to ribosomal proteins from the archaebacterium *H. marismortui*. Hitherto only partial sequences of the 'A' protein [5] and N-terminal sequences of 10 ribosomal proteins from archaebacteria [6] have been reported. Here we present the complete amino acid sequence of the 30 S subunit protein HS3 of *H. marismortui*, and it is shown that it is homologous with the eubacterial ribosomal proteins S9 from *E. coli* and *B. stearothermophilus*.

2. MATERIALS AND METHODS

30 S ribosomal proteins from *H. marismortui* were isolated as in [7], and purification of protein HS3 was achieved by DEAE–cellulose chromatography followed by gel filtration on Sephadex G-75 as in [6]. Details of the isolation and purification

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

of the 30 S ribosomal proteins from *H. marismortui* will be published elsewhere. The purity and identity of the isolated protein were established by two-dimensional gel electrophoresis [7]. Enzymes and other materials for sequencing were used as in [8]. Sequence determination of protein HS3 was carried out as in [9,10]. Computer analysis for a comparison of all ribosomal proteins available at present was performed by the aid of two programs, RELATE and ALIGN [11,12]. RELATE was used for the comparison of HS3 with other ribosomal proteins searching with a fragment length of 25. When segment comparison scores over 3.0 SD units were obtained for the HS3 protein with one of the other proteins these pairs were then subjected to analysis with the program ALIGN. In this case the mutant data matrix with a break penalty of 25 was employed. In the computer analysis, the relatedness of any two sequences was estimated as in [13]. The proteins compared with HS3 were the ribosomal proteins from *E. coli* [14,15], *B. stearothermophilus* [16,17], rat liver [18–21], *Saccharomyces cerevisiae* [22] and *Schizosaccharomyces pombe* [23].

3. RESULTS AND DISCUSSION

3.1. *Determination of the amino acid sequence of protein HS3*

Protein HS3 was eluted from the DEAE–cellulose column at the beginning of the gradient, at a KCl concentration of 0.02 M. After further purification by gel filtration on Sephadex G-75, the protein migrated as a single spot in two-dimensional gel electrophoresis as shown in fig.1.

The amino acid sequence of HS3 is presented in fig.2. This sequence was deduced as follows. Protein HS3 was first digested with trypsin, and the peptides were isolated by thin-layer fingerprinting. Amino acid sequences of the peptides were determined by the DABITC/PITC method [24] or by the manual solid-phase technique as in [9]. The alignment of the tryptic peptides was mainly deduced from the amino acid sequences of chymotryptic and peptic peptides. Chymotryptic and peptic digestion of HS3 gave 11 (C1–C11) and 13 (P1–P13) peptides, respectively, which were separated by fingerprinting, and the resulting peptides were sequenced. The amino acid sequence information from these peptides allowed the alignment of almost all tryptic peptides, but several overlaps were based on a single residue only. Hence, HS3 was also digested with *Staphylococcus aureus* protease, and the resulting peptides were isolated by gel filtration on Sephadex G-50 and then sequenced. From these results, the amino acid sequence of HS3 could be unambiguously determined as shown in fig.2.

Protein HS3 contains 130 amino acid residues, and from its amino acid composition a molecular

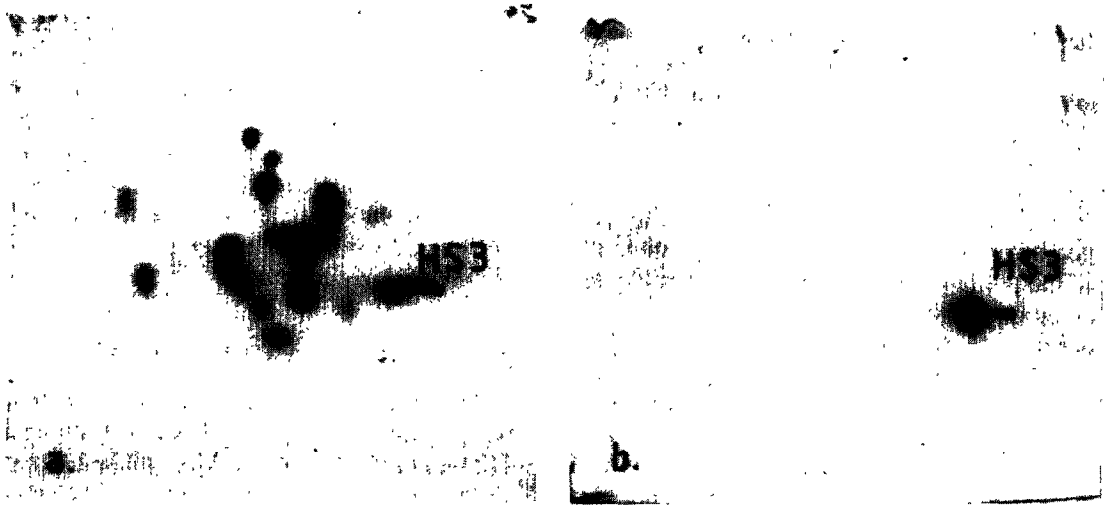

Fig.1. Two-dimensional gel electropherograms of 30 S ribosomal proteins from *H. marismortui* (a) and the purified protein HS3 (b). Protein HS3 is shown by an arrow.

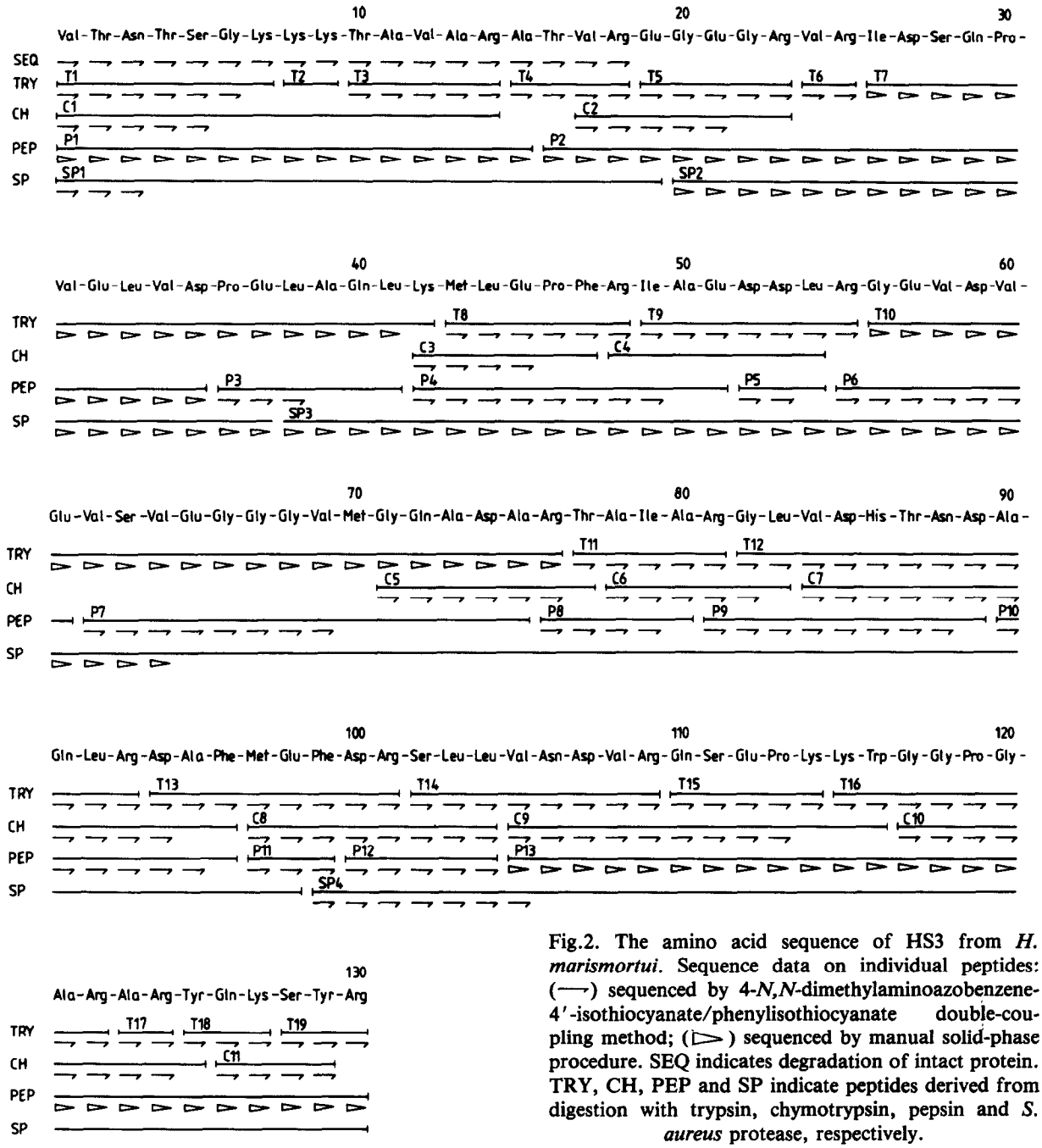

Fig.2. The amino acid sequence of HS3 from *H. marismortui*. Sequence data on individual peptides: (——) sequenced by 4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate double-coupling method; (▷) sequenced by manual solid-phase procedure. SEQ indicates degradation of intact protein. TRY, CH, PEP and SP indicate peptides derived from digestion with trypsin, chymotrypsin, pepsin and *S. aureus* protease, respectively.

mass of 14353 is calculated. The amino acid composition of HS3, as derived from the sequence, agrees well with the amino acid analysis of the protein as given in table 1. From the number of acidic residues (22 residues) and basic residues (22 including one His residue) in HS3, a net charge of −0.5 at pH 6.5 was calculated assuming that the His residue carries a charge of +0.5 [25].

Table 1

Amino acid composition and M_r of protein HS3

Amino acid	Sequence	Protein hydrolysate
Asp	11 }	14.3
Asn	3 }	
Thr	6	6.0
Ser	6	5.5
Glu	11 }	20.1
Gln	6 }	
Pro	5	4.1
Gly	12	11.8
Ala	13	13.3
Val	14	14.2
Met	3	2.7
Ile	3	3.0
Leu	9	8.5
Tyr	2	1.3
Phe	3	2.5
His	1	0.8
Lys	7	6.8
Arg	14	14.2
Trp	1	ND
Total	130	
M_r	14353	

ND, not determined

Examining the distribution of charged residues throughout the sequence, the following characterization can be deduced: 15 of 22 basic residues are located in the N-terminal (positions 1–25) and C-terminal (positions 100–130) regions. On the other hand, acidic residues exist densely in the central part of the molecule. The location of aromatic residues also seems to be non-random: All but Phe-47 are clustered within the C-terminal third of the chain, namely Phe-96, Phe-99, Trp-116, Tyr-125 and Tyr-129.

3.2. *Comparison of ribosomal proteins from* Halobacterium *with those of eubacteria and eukaryotes*

It was concluded from a comparison of the N-terminal sequences of 10 ribosomal proteins from the *H. cutirubrum* 30 S subunit that these proteins do not show significant homology to eubacterial ribosomal proteins, but do have some homology with eukaryotic ribosomal proteins [6].

After elucidation of the complete amino acid sequence of protein HS3 we compared its structure with the sequences of all ribosomal proteins thus far determined, using the computer program RELATE. The segment comparison obtained showed that only two eubacterial ribosomal proteins, namely, the S9 proteins from *E. coli* [26] and *B. stearothermophilus* ribosomes [17], exhibited a significant homology (at the level of 3.9 and 5.8 SD units, respectively) with the HS3 protein from *Halobacterium*.

Accordingly, all 3 proteins (HS3, ES9 and BS9) were aligned for maximum homology by the program ALIGN. This comparison was made as follows. Firstly, protein HS3 was aligned with BS9 (fig.3), and the alignment score thus obtained was 10.3 (in SD units). This is strong statistical evidence that HS3 is related to BS9, since it has

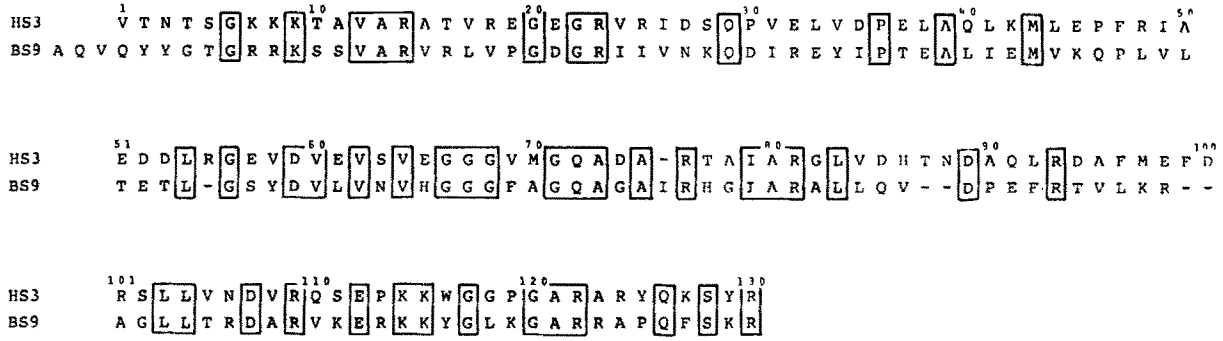

Fig.3. Comparison of the primary structure of HS3 from *H. marismortui* with that of BS9 from *B. stearothermophilus*. Identical residues are enclosed in boxes.

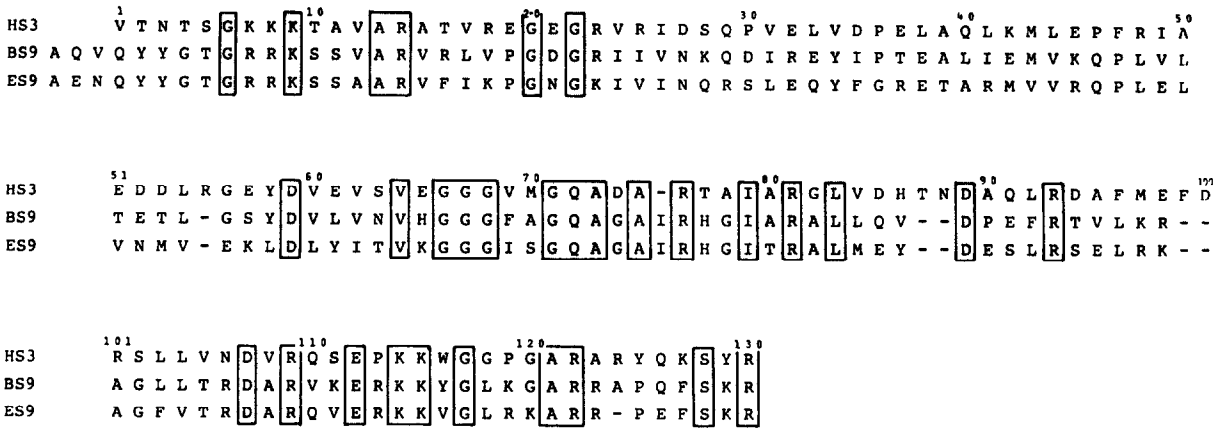

Fig.4. Comparison of the primary structures of 3 homologous proteins from *H. marismortui, B. stearothermophilus* and *E. coli*. Identical residues are enclosed in boxes.

been pointed out [13] that two proteins which give a score of more than 10 SD units are strongly related to each other.

Next ES9 was aligned with the two sequences BS9 and HS3 (fig.4). This comparison shows that HS3 can easily be aligned to the two S9 proteins with a few insertions and one deletion, and that HS3 has 37 and 27% identical residues with BS9 and ES9, respectively. Thus, protein HS3 from *Halobacterium* is more related to the corresponding protein in the gram-positive *B. stearothermophilus* than to that in the gram-negative *E. coli*.

There are 31 positions in the sequence which are completely conserved in all 3 proteins, i.e., HS3, ES9 and BS9. In particular the central region of the proteins, at positions 59–83, and the C-terminus are very homologous. This finding suggests that these regions may play an important role in the function and/or structure of the proteins.

There is no indication from our computer search program that protein HS3 is homologous to any of the eukaryotic ribosomal proteins whose primary structures are so far known. However, it is possible that this negative finding is due to insufficient sequence data for proteins from eukaryotic ribosomes. It is interesting in this context that another ribosomal protein from *Halobacterium*, the so-called 'A' protein, is significantly homologous to a ribosomal protein from yeast [5] and that the 5 S RNAs from archaebacteria and eukaryotes are related [4].

On the other hand, protein HS3 (this paper) and the 16 S RNA [2] from *Halobacterium* ribosomes show significant homology to their counterparts in eubacteria. Therefore, it is possible that some ribosomal components of archaebacteria are more closely related to eubacteria and others more closely to eukaryotes. A definite answer to this interesting hypothesis must await more sequence data from archaebacterial and eukaryotic ribosomal components.

ACKNOWLEDGEMENTS

We thank Dr H.G. Wittmann for his interest and encouragement, and Drs B. Wittmann-Liebold and R. Brimacombe for critically reading the manuscript.

REFERENCES

[1] Woese, C.R. and Fox, G.E. (1977) Proc. Natl. Acad. Sci. USA 74, 5088–5090.
[2] Gupta, R., Lanter, J.M. and Woese, C.R. (1983) Science 221, 656–659.
[3] Matheson, A.T. and Yaguchi, M. (1982) Zbl. Bakt., I. Abt. Orig. C3, 192–199.
[4] 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.
[5] Yaguchi, M., Matheson, A.T., Visentin, L.P. and Zuker, M. (1980) in: Genetics and Evolution of

RNA Polymerase, tRNA and Ribosomes (Osawa, S. et al. eds) pp.585–599, University of Tokyo Press, Tokyo.
[6] Yaguchi, M., Visentin, L.P., Zuker, M., Matheson, A.T., Roy, C. and Strom, A.R. (1982) Zbl. Bakt., I. Abt., Orig. C3, 200–208.
[7] Strom, A.R. and Visentin, L.P. (1973) FEBS Lett. 37, 274–280.
[8] Kimura, M., Foulaki, K., Subramanian, A.R. and Wittmann-Liebold, B. (1982) Eur. J. Biochem. 123, 37–53.
[9] 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.
[10] Wittmann-Liebold, B. (1981) in: Chemical Synthesis and Sequencing of Peptides and Proteins (Liu, T.Y. et al. eds) pp.75–110, Elsevier, Amsterdam, New York.
[11] George, D.G., Orcutt, B.C., Dayhoff, M.O. and Barker, W.C. (1982) National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC.
[12] Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, vol.5, suppl.3, National Biomedical Research Foundation, Washington, DC.
[13] Barker, W.C., Ketcham, L.K. and Dayhoff, M.O. (1978) J. Mol. Evol. 10, 265–281.
[14] Wittmann, H.G. (1982) Annu. Rev. Biochem. 51, 155–183.
[15] Giri, L., Hill, W.E., Wittmann, H.G. and Wittmann-Liebold, B. (1984) Adv. Protein Chem. 36, 2–46.
[16] Kimura, M. (1984) J. Biol. Chem. 259, 1051–1055.
[17] Kimura, M. and Chow, C.K. (1984) Eur. J. Biochem. 139, 225–234.
[18] Wittmann-Liebold, B., Geissler, A.W., Lin, A. and Wool, I.G. (1979) J. Supramol. Struct. 12, 425–433.
[19] Lin, A., Wittmann-Liebold, B., Mcnally, J. and Wool, I.G. (1982) J. Biol. Chem. 257, 9189–9197.
[20] Lin, A., Mcnally, J. and Wool, I.G. (1983) J. Biol. Chem. 258, 10664–10671.
[21] Lin, A., Mcnally, J. and Wool, I.G. (1984) J. Biol. Chem. 259, 487–490.
[22] Otaka, E., Higo, K. and Osawa, S. (1982) Biochemistry 21, 4545–4550.
[23] Otaka, E., Higo, K. and Itoh, T. (1983) Mol. Gen. Genet. 191, 519–524.
[24] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[25] Offord, R.E. (1966) Nature 211, 591–593.
[26] Chen, R. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1415–1430.