Amino acid sequences of ribosomal proteins S11 from *Bacillus stearothermophilus* and S19 from *Halobacterium marismortui*

Comparison of the ribosomal protein S11 family

Makoto Kimura, Junko Kimura and Tomomitsu Hatakeyama

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

The complete amino acid sequences of ribosomal proteins S11 from the Gram-positive eubacterium *Bacillus stearothermophilus* and of S19 from the archaebacterium *Halobacterium marismortui* have been determined. A search for homologous sequences of these proteins revealed that they belong to the ribosomal protein S11 family. Homologous proteins have previously been sequenced from *Escherichia coli* as well as from chloroplast, yeast and mammalian ribosomes. A pairwise comparison of the amino acid sequences showed that *Bacillus* protein S11 shares 68% identical residues with S11 from *Escherichia coli* and a slightly lower homology (52%) with the homologous chloroplast protein. The halophilic protein S19 is more related to the eukaryotic (45–49%) than to the eubacterial counterparts (35%).

Amino acid sequence; Ribosomal protein; Archaebacteria; Evolution; (*B. stearothermophilus, H. marismortui*)

1. INTRODUCTION

Since ribosomes are ubiquitous in all organisms, their components are ideal for studies on molecular evolution. Numerous 16 S-like rRNA genes from various organisms have been cloned and sequenced. This sequence information indicated that the evolutionary relationships between eubacteria, archaebacteria and eukaryotes are equidistant [1]. Similar studies on the evolutionary relationships between these three groups were based on the sequences of the 23 S-like rRNAs [2] and the 5 S rRNAs [3]. The results of the 23 S-like rRNAs are essentially the same as those from the 16 S-like rRNAs, while the data from 5 S rRNAs are somewhat different in that the archaebacteria may have diverged directly from eukaryotes. With regard to the ribosomal proteins, a number of A-proteins, i.e. the ribosomal protein L12 family,

from a wide range of organisms have been determined. The results are consistent with those from the 5 S rRNAs, indicating that the archaebacterial A-proteins are more related to the eukaryotic than to the eubacterial counterparts [4].

Our studies in this field have been concerned with the determination of the amino acid sequences of ribosomal proteins from the Gram-positive eubacterium *Bacillus stearothermophilus* [5] and the archaebacterium *Halobacterium marismortui* [6]. These sequences were compared with those from the Gram-negative eubacterium *Escherichia coli* and from eukaryotic organisms. Our sequence studies have also been undertaken in order to support the X-ray structural analysis of crystals of the intact large ribosomal subunits from *B. stearothermophilus* and *H. marismortui* [7,8]. Furthermore, the comparison of the primary structures of ribosomal proteins from different organisms will help to identify the functionally and/or structurally important regions within the individual ribosomal proteins, since these regions have been conserved during evolution.

Correspondence address: T. Hatakeyama, Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-1000 Berlin 33 (Dahlem), Germany

Here, we report the determination of the amino acid sequences of the ribosomal protein S11 from *B. stearothermophilus* (BS11) and of its homologue from *H. marismortui* (HS19), and we compare their sequences with those of protein S11 from *E. coli* and other organisms. The *E. coli* ribosomal protein S11 plays an essential role for the selection of the correct tRNA in protein biosynthesis [9], and is located on the large lobe of the small subunit [10]. Therefore, it is interesting to compare the S11 homologous proteins from various organisms and thereby identify the conserved regions within the protein chain.

2. MATERIALS AND METHODS

2.1. *Materials*

Protein S11 from 30 S subunits of *B. stearothermophilus* (strain NCA 1503) was isolated and purified as in [11]. Ribosomal protein S19 from *H. marismortui* was obtained as described [12]. Enzymes and other materials for sequencing were used as in [13].

2.2. *Sequence determination*

Enzymatic digestions and sequence determination were performed as in [13]. An acid cleavage at aspartyl bonds with dilute HCl and an Asn-Gly cleavage with hydroxylamine were performed according to [14,15], respectively.

2.3. *Computer analysis*

A search of the NBRF data bank for homologous ribosomal proteins and the optimal alignment of related sequences were carried out on a Vax/VMS 8600 computer using the RELATE and ALIGN programs [16].

3. RESULTS AND DISCUSSION

3.1. *Amino acid sequence of S11 from* B. stearothermophilus

The complete amino acid sequence of S11 was mainly derived by the analysis of peptides obtained by cleavages with lysylendopeptidase, *Staphylococcus aureus* protease and chymotrypsin. Lysylendopeptidase digestion of S11 was performed at pH 8.1 in 0.2 M *N*-methylmorpholine acetate buffer at 37°C for 6 h. The enzymatic digestion produced an insoluble fraction of the peptide mixture, and therefore it was first separated into the soluble and pellet fractions. The precipitate was further washed with 0.2 M *N*-methylmorpholine acetate buffer to remove any

trapped soluble peptides. It was then dissolved in 0.1% trifluoroacetic acid and further purified by reverse-phase high-performance liquid chromatography (RP-HPLC). The peptides that were soluble in the buffer were lyophilized and separated by RP-HPLC. The pellet fraction comprised primarily one major peptide (peptide K3), while the soluble fraction yielded seven peptides.

The amino acid sequences of all peptides, except peptide K3, were completely determined by the DABITC/PITC double-coupling method. The C-terminal residues of peptide K3 were obtained from the sequences of peptides K3-T2 and K3-A2 derived from tryptic and acid cleavage of peptide K3. Alignment of the lysylendopeptidase peptides was accomplished by sequencing the chymotryptic and *S. aureus* protease peptides derived from protein S11. Chymotryptic and *S. aureus* protease digestion of S11 yielded nine and seven peptides, respectively which were separated by RP-HPLC. Sequencing of these peptides established the alignment of lysylendopeptidase peptides and completed the amino acid sequence of S11, as shown in fig.1.

3.2. *Amino acid sequence of S19 from* H. marismortui

The sequence of protein S19 (fig.1) was established by procedures similar to those described for protein S11. It was obtained by the isolation and characterization of lysylendopeptidase peptides which were aligned by overlapping peptides isolated from *S. aureus* and chymotryptic digests, as well as by cyanogen bromide cleavage. The combination of the sequence information obtained from these peptides could establish almost all amino acid residues of S19, except at the C-terminus of the protein, where an Asn-Gly sequence at 119–120 lowered the sequencing yield drastically.

This problem was turned into advantage by cleaving the protein at the Asn-Gly with hydroxylamine. The protein was at first cleaved with 2 M NH_2OH at pH 9.0 at 45°C for 5 h in the presence of 6 M guanidine HCl and then desalted on Sephadex G-10. The mixture was separated by RP-HPLC. This procedure gave a pure peptide, NG2, which provided the C-terminal sequence of protein S19 (fig.1).
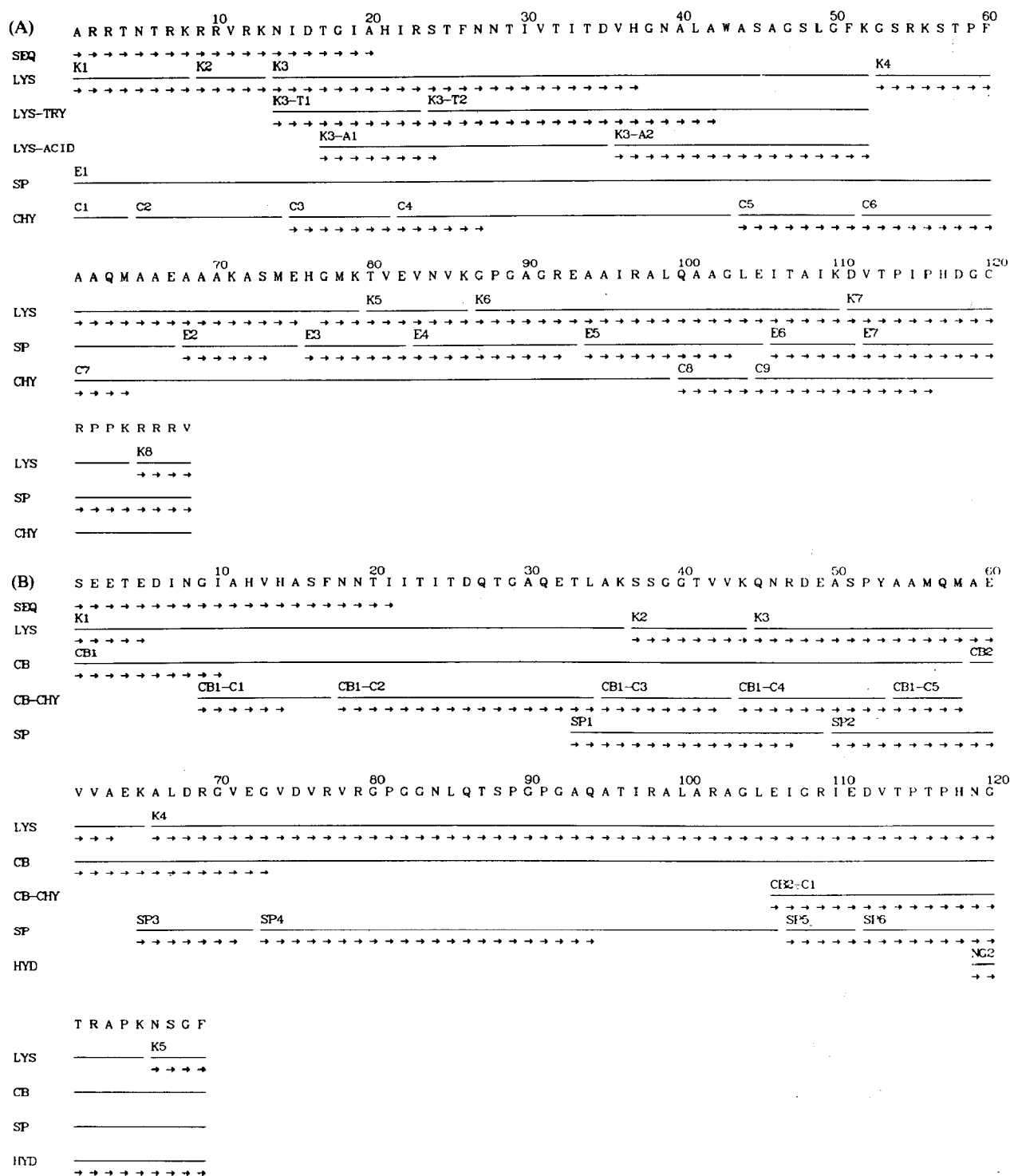
(A)

```
              10          20          30          40          50          60
A R R T N T R K R R V R K N I D T G I A H I R S T F N N T I V T I T D V H G N A L A W A S A G S L G F K G S R K S T P F
```

SEQ → → → → → → → → → → → → → → → → → → →

LYS
K1   K2   K3   K4
→ →
LYS–TRY
              K3–T1   K3–T2
              →
LYS–ACID
              K3–A1   K3–A2
              → → → → → → → →   → → → → → → → → → → → → → → → → → →
SP
E1

CHY
C1   C2   C3   C4   C5   C6
              → → → → → → → → → → → →   → → → → → → → → → → → → → → →

```
          70          80          90          100         110         120
A A Q M A A E A A A K A S M E H G M K T V E V N V K G P G A G R E A A I R A L Q A A G L E I T A I K D V T P I P H D G C
```

LYS
                              K5   K6   K7
→ →
SP
          E2   E3   E4   E5   E6   E7
          → → → → → →   → → → → → → → → → → → → →   → → → → → →   → → → → → → → → → → → →
CHY
C7   C8   C9
→ → → →   → → → → → → → → → → → → → → → → →

```
R P P K R R R V
```

LYS
        K8
        → → → →
SP
        → → → → → → →
CHY

(B)

```
              10          20          30          40          50          60
S E E T E D I N G I A H V H A S F N N T I I T I T D Q T G A Q E T L A K S S G G T V V K Q N R D E A S P Y A A M Q M A E
```

SEQ →

LYS
K1   K2   K3
→ → → → →   →
CB
CB1   CB2
→ → → → → → → → → →
CB–CHY
              CB1–C1   CB1–C2   CB1–C3   CB1–C4   CB1–C5
              → → → → → →   →
SP
                                                     SP1   SP2
                                                     → → → → → → → → → → → → →   → → → → → → → → → →

```
          70          80          90          100         110         120
V V A E K A L D R G V E G V D V R V R G P G G N L Q T S P G P G A Q A T I R A L A R A G L E I G R I E D V T P T P H N G
```

LYS
        K4
→ → →   →
CB
→ → → → → → → → → → → →
CB–CHY
                                                                    CB2–C1
                                                                    → → → → → → → → → → → → → →
SP
        SP3   SP4   SP5   SP6
        → → → → → → →   →   → → → → → → → → → → → →
HYD
                                                                                             NG2
                                                                                             → →

```
T R A P K N S G F
```

LYS
        K5
        → → → →
CB

SP

HYD
        → → → → → → → →

Fig.1. Amino acid sequences of protein S11 from *B. stearothermophilus* (A) and of S19 from *H. marismortui* (B). SEQ indicates degradation of intact proteins. LYS, SP, CHY, CB and HYD designate peptides derived by cleavages of proteins with lysylendopeptidase, *S. aureus* protease, chymotrypsin, cyanogen bromide and hydroxylamine, respectively. LYS-TRY, LYS-ACID and CB-CHY indicate peptides obtained by tryptic digestion, acid cleavage and chymotryptic digestion of the peptide. Arrows (→) show the cycles of the DABITC/PITC double-doupling method.

3.3. Sequence comparison

The search in the NBRF data bank for ribosomal protein sequences homologous to the *B. stearothermophilus* S11 (BS11) and the *H. marismortui* protein S19 (HS19) using the program RELATE showed that both proteins are related to the *E. coli* ribosomal protein S11 (ES11) [17]. The three proteins ES11, BS11 and HS19 as well as the *E. coli* S11 homologous ribosomal proteins from tobacco chloroplasts [18], yeast [19] and mammals [20,21] can be aligned by the program ALIGN as shown in fig.2. In this comparison, the chloroplast S11 and the mammalian S14 have extra amino acid residues compared with the other four proteins, and there is a 10-amino-acid extension for the chloroplast S11 and a 12-amino-acid insertion for

the mammalian S14. Furthermore, the halophilic protein HS19 and the two eukaryotic proteins have an extra Glu residue at position 49 of HS19 and an insertion of eight amino acids (from positions 79 to 86 in HS19). In spite of these alterations, there are 26 positions which are completely conserved in all six proteins.

On the basis of the distribution of conserved and substituted amino acid residues, the protein chain of the S11 family can be divided into four regions as shown in fig.2. In regions A and C, there are insertions which are especially long in the A region. Therefore, the two regions seem to have evolved under a weak functional constraint. In contrast, the B and D regions are highly conserved among all six proteins. Almost all conserved residues are

Fig.2. Alignment of amino acid sequences of six ribosomal proteins belonging to the S11 family. The alignment was first made by the computer program ALIGN, and then refined by hand to maximize the homology for all six proteins. The amino acid sequences of the *E. coli* S11, tobacco chloroplast S11, yeast RP59 and mammalian S14 are taken from [17–19,21], respectively. Amino acid residues are numbered according to the halophilic protein S19. The protein chains are divided into four regions, A–D, based on the distribution of conserved and substituted amino acid residues.

Table 1

Extent of sequence identity within the S11 family

	H. marismortui	E. coli	B. stearo- thermophilus	Tobacco	Yeast
H. marismortui S19	–				
E. coli S11	35	–			
B. stearothermophilus S11	35	68	–		
Tobacco CS11	30	49	52	–	
Yeast RP59	49	37	35	34	–
Mammalian S14	45	42	36	34	79

Values expressed as %

located in these regions, and furthermore, many of the amino acid exchanges in these regions represent conservative substitutions. This indicates that these regions are under strong constraints in the protein structure and/or function. Therefore, the B and D regions or one of these regions may play an important role in the function and/or structure of the protein S11. Interestingly, the C-terminus of the halophilic protein has a unique sequence (APKNSGF) instead of the conserved basic residues in the other five proteins. In addition, protein HS19, like eubacterial proteins, lacks at the C-terminus two Arg residues which are involved in the emetine inhibition for eukaryotic ribosomes [20]. It can be expected, therefore, that the C-terminal region of the bacterial proteins is not essential for ribosome function, and might rather be involved in the stability of the protein structure.

In table 1, a matrix showing the extent of sequence identity among the S11 homologous proteins from various organisms is presented. The halophilic protein HS19 has more homology to proteins from eukaryotes than to those from eubacteria, and even less to the chloroplast protein from tobacco. BS11 shows the best homology to ES11 sharing 68% identical residues, and a slightly lower homology to the chloroplast protein. Based on this comparison, the evolutionary tree for the ribosomal protein S11 family (fig.3) has two main branches: one for eubacterial and chloroplast S11, and another for halophilic and eukaryotic proteins. This relationship is consistent with previous results derived from the comparisons of other ribosomal protein families [4,22,23]. Similar results to those presented in fig.3 have recently
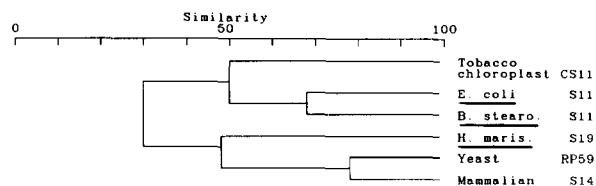

Fig.3. Evolutionary relationship of the E. coli S11 homologous ribosomal proteins based on the comparative data presented in table 1.

been obtained from sequence analyses of ribosomal proteins from the archaebacterium Methanococcus vannielii. These proteins also exhibit a significant similarity with their eukaryotic counterparts (Köpke, A., Wittmann-Liebold, B. and Böck, A., unpublished).

It is, however, generally considered, based on the sequence comparison of the large ribosomal RNAs, that three main lineages, eubacteria, archaebacteria and eukaryotes, descended from a common ancestor at the same time. If this holds true, the incompatibility of the results derived from ribosomal RNAs and proteins suggests that the rate of sequence changes for archaebacterial and eukaryotic ribosomal proteins has decreased after their divergence from eubacteria. In other words, the ribosomal proteins from eubacteria may have evolved faster than those of archaebacterial and eukaryotic ribosomal proteins.

REFERENCES

[1] Woese, C.R. and Olsen, G.J. (1986) System. Appl. Microbiol. 7, 161–177.

[2] Leffers, H., Kjems, J., Ostergaard, L., Larsen, N. and Garrett, R.A. (1987) J. Mol. Biol. 195, 43–61.

[3] Hori, H. and Osawa, S. (1986) Biosystems 19, 163–172.

[4] Matheson, A.T. and Yaguchi, M. (1982) Zentralbl. Bacteriol. Abt. 1 Orig. B Hyg. C3, 192–199.

[5] Kimura, M. and Kimura, J. (1987) FEBS Lett. 210, 91–96.

[6] Kimura, J., Arndt, E. and Kimura, M. (1987) FEBS Lett. 224, 65–70.

[7] Yonath, A., Saper, M.A. and Wittmann, H.G. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp.112–127, Springer, Heidelberg.

[8] Makowski, I., Frolow, F., Saper, M.A., Shoham, M., Wittmann, H.G. and Yonath, A. (1987) J. Mol. Biol. 193, 819–822.

[9] Nomura, M. (1970) Bacteriol. Rev. 34, 228–277.

[10] Stöffer-Meilicke, M. and Stöffler, G. (1987) Biochimie 69, 1049–1064.

[11] Isono, S. and Isono, K. (1975) Eur. J. Biochem. 50, 483–488.

[12] Arndt, E., Breithaupt, G. and Kimura, M. (1986) FEBS Lett. 194, 227–234.

[13] 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.

[14] Inglis, A.S. (1983) Methods Enzymol. 91, 324–332.

[15] Balian, G., Click, E.M. and Bornstein, P. (1971) Biochemistry 10, 4470–4478.

[16] Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, vol.5, suppl.3, National Biomedical Research Foundation, Washington, DC.

[17] Kamp, R. and Wittmann-Liebold, B. (1980) FEBS Lett. 121, 117–122.

[18] Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinizaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. and Sugiura, M. (1986) EMBO J. 5, 2043–2049.

[19] Teem, J.L., Abovich, N., Kaufer, N.F., Schwindinger, W.F., Warner, J.R., Levy, A., Woolford, J., Leer, R.J., Van Raamsdonk-Duin, M.M.C., Mager, W.H., Planta, R.J., Schultz, L., Friesen, J.D., Fried, H. and Rosbash, M. (1984) Nucleic Acids Res. 12, 8295–8312.

[20] Rhoads, D.D. and Roufa, D.J. (1985) Mol. Cell. Biol. 5, 1655–1659.

[21] Chen, I.-T., Dixit, A., Rhoads, D.D. and Roufa, D.J. (1986) Proc. Natl. Acad. Sci. USA 83, 6907–6911.

[22] Kimura, J. and Kimura, M. (1987) J. Biol. Chem. 262, 12150–12157.

[23] Hatakeyama, T. and Kimura, M. (1988) Eur. J. Biochem. 172, 703–711.