Volume 240, number 1,2, 21–28 FEB 06492 November 1988

The primary structures of ribosomal proteins L16, L23 and L33 from the archaebacterium *Halobacterium marismortui*

Tomomitsu Hatakeyama, Tamiko Hatakeyama and Makoto Kimura

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

Received 3 August 1988

The complete amino acid sequences of ribosomal proteins L16, L23 and L33 from the archaebacterium *Halobacterium marismortui* were determined. The sequences were established by manual sequencing of peptides produced with several proteases as well as by cleavage with dilute HCl. Proteins L16, L23 and L33 consist of 119, 154 and 69 amino acid residues, and their molecular masses are 13538, 16812 and 7620 Da, respectively. The comparison of their sequences with those of ribosomal proteins from other organisms revealed that L23 and L33 are related to eubacterial ribosomal proteins from *Escherichia coli* and *Bacillus stearothermophilus*, while protein L16 was found to be homologous to a eukaryotic ribosomal protein from yeast. These results provide information about the special phylogenetic position of archaebacteria.

Amino acid sequence; Ribosomal protein; Sequence comparison; (*H. marismortui*)

1. INTRODUCTION

Elucidation of the detailed structure of ribosomes is a prerequisite for understanding the molecular mechanism of protein biosynthesis. The most direct way of determining the tertiary structure of a biological macromolecule is by its crystallization followed by X-ray analysis. This approach has been applied to whole ribosomes from *Escherichia coli*, as well as to 50 S ribosomal subunits from *Bacillus stearothermophilus* [1]. Recently, large and well-ordered three-dimensional crystals of 50 S ribosomal subunits from *Halobacterium marismortui* have been obtained by taking advantage of their unusual stability under high salt conditions. These crystals diffract to 6 Å in a synchrotron beam [2]. Since chemical information, such as the primary structure of constituent proteins, is essential to solve the phase problem by heavy-atom derivatives, we are presently determining the amino acid sequences of proteins isolated from the large subunits of *H. marismortui*. We have previously presented the complete amino acid sequences of the three 50 S ribosomal proteins L25, L29 and L31 [3].

Furthermore, our study is beginning to shed light on the evolutionary relationship of ribosomal proteins from halobacteria to those from eubacteria and eukaryotes. Previous comparative studies on the primary structures of ribosomal proteins from *H. marismortui* and other organisms [3–8] indicated that the amino acid sequences of several halophilic ribosomal proteins are at least in some regions similar to those of eubacterial or eukaryotic organisms in spite of the extremely different ionic conditions in which the ribosomes function, and that halophilic ribosomal proteins appear in general to be more related to their eukaryotic than to their eubacterial counterparts.

Here, we describe the complete amino acid sequences of proteins L16, L23 and L33, and present their relationships to eubacterial and eukaryotic ribosomal proteins.

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

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

2. MATERIALS AND METHODS

2.1. *Preparation of ribosomal proteins*

Ribosomal proteins L16, L23 and L33 were obtained from 50 S ribosomal subunits of *H. marismortui* as described [3].

2.2. *Sequence determination*

Proteins were digested with trypsin (Merck), chymotrypsin (Merck), *Staphylococcus aureus* V8 protease (Miles) and lysylendopeptidase (Wako) in 0.2 M *N*-methylmorpholine acetate buffer (pH 8.1) at 37°C for 5 h, at an enzyme/substrate ratio of 1:50 (w/w). Cleavage of peptides with dilute HCl was carried out for 2 h at 108°C in 0.01 M HCl according to Inglis [9]. The resulting peptides were separated by reverse-phase HPLC on a Vydac C_{18} column (4.6 × 250 mm) with a linear gradient of acetonitrile in 0.1% trifluoroacetic acid. Eluates were monitored by the absorbance at 220 nm.

Amino acid analyses were performed on an HPLC system using *o*-phthaldialdehyde as a derivatized reagent [10]. The amino acid sequences were determined using the DABITC/PITC double-coupling method [11]. C-terminal sequences of the proteins were examined using carboxypeptidase Y (Boehringer Mannheim) as in [3].

2.3. *Computer analysis*

The amino acid sequences were compared with those of other ribosomal proteins in the NBRF Protein Sequence Data Base (release 13, July, 1987) as well as in our own files [3] by the computer programs RELATE and ALIGN [12]. Hydropathic analysis was performed using the program HYDROP in the UWGCG (University of Wisconsin Genetic Computer Group, Version 5.2, February 1988).

3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

The complete amino acid sequences of ribosomal proteins L16, L23 and L33 of *H. marismortui* are shown in fig.1. The proteins were primarily digested with either trypsin or lysylendopeptidase, and each digest was separated by reverse-phase HPLC on a Vydac C_{18} column. Amino acid sequences of separated peptides were analyzed by the DABITC/PITC double-coupling method [11]. Alignments of these peptides were established by amino acid sequences of overlapping peptides obtained by *S. aureus* protease digestion. As shown in fig.1, a combination of these results provided most of the sequence information for the proteins. However, for some regions, further digestions with chymotrypsin, thermolysin or cleavage using dilute HCl were required to obtain all overlaps of the peptides. The C-terminal sequences of these proteins were confirmed by carboxypeptidase Y digestion as described in [3] except for protein L16 which exhibited resistance against the digestion.

Proteins L16, L23 and L33 consist of 119, 154 and 69 amino acid residues with molecular masses of 13538, 16812 and 7620 Da, respectively. The amino acid compositions calculated from these sequence data were in good agreement with those obtained from the hydrolysates of the intact proteins.

In order to examine the structural features of the halophilic ribosomal proteins, the hydropathic analysis based on the amino acid sequences thus determined was performed according to Kyte and Doolittle [13]. This analysis revealed the general hydrophilic character of these proteins, as shown in fig.2. In particular, the N-terminal portions (positions 1–24 and 30–40) and the C-terminal portion (positions 100–119) of L16, as well as the C-terminal region (positions 120–130) of L23, are calculated to be strongly hydrophilic. These regions may play an important role in the protein's overall stability in a high salt environment.

3.2. *Comparison of the amino acid sequences with those from other organisms*

The complete amino acid sequences of proteins HL16, HL23 and HL33 from *H. marismortui* were compared by a combination of the two computer programs RELATE and ALIGN with those of other ribosomal proteins. This analysis showed that protein HL23 gave a rather high score (alignment score 6.7) with the eubacterial protein *E. coli* EL22. As shown in fig.3, protein HL23 can be aligned with EL22, but has a long insertion in the middle region (positions 51–80). The two molecules share 26 identical residues (24% identity), and in particular, most of the amino acid residues in positions 127–133 of HL23 are identical to those in positions 87–93 of EL22.

Protein HL33 is significantly homologous to eubacterial proteins EL29 from *E. coli* and BL29 from *B. stearothermophilus*, giving alignment scores of 5.3 and 12.0, respectively. Interestingly, as shown in table 1, the homology between HL33 and BL29 (43.9%) is much higher than that between HL33 and EL29 (31.6%). The former value is almost as high as that for the two eubacterial proteins EL29 and BL29 (45.6%). The greater similarity of HL33 to BL29 than to EL29 is also evident from the fact that the region corresponding to the sequence from 43 to 48 in *E. coli* is not pre-

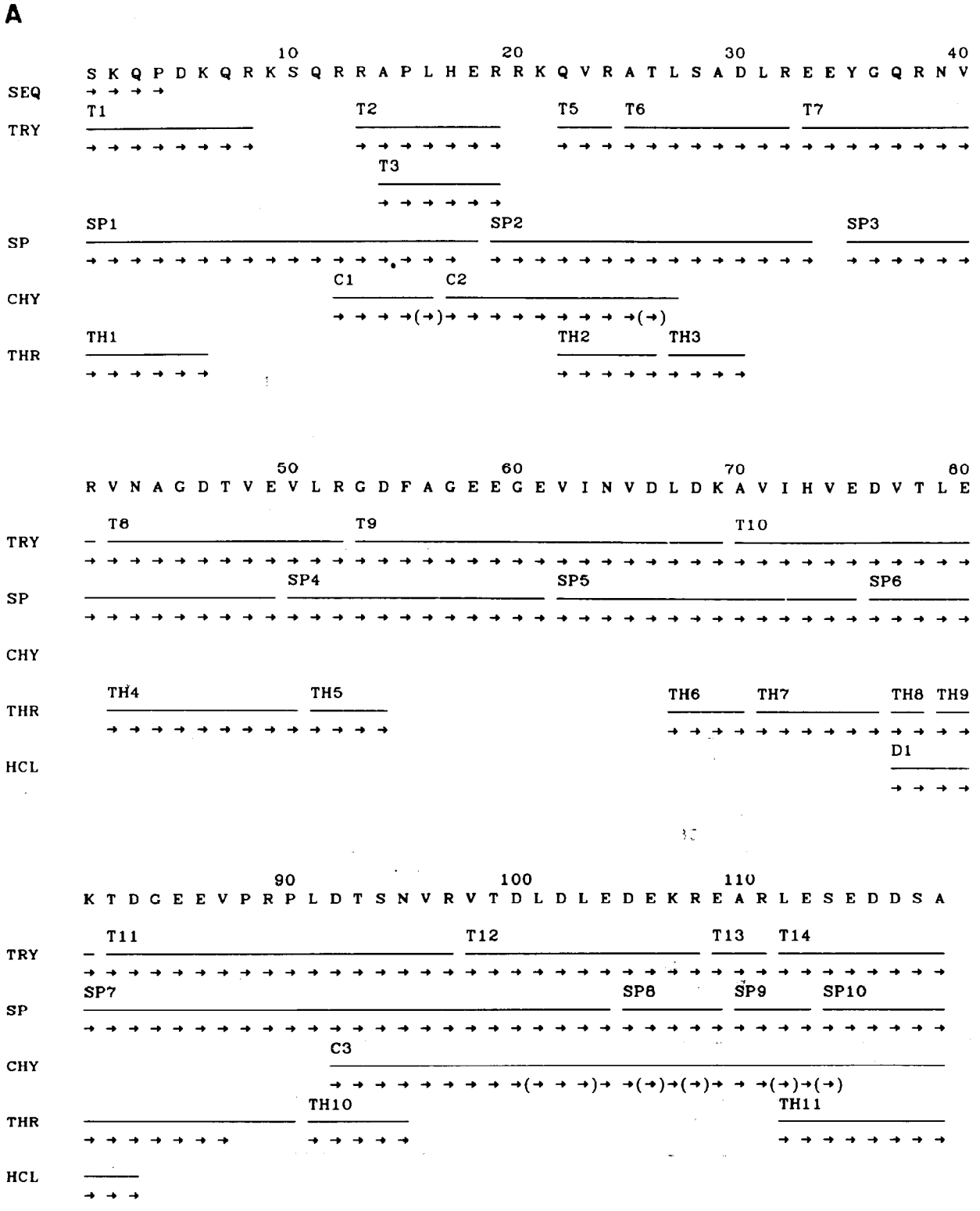
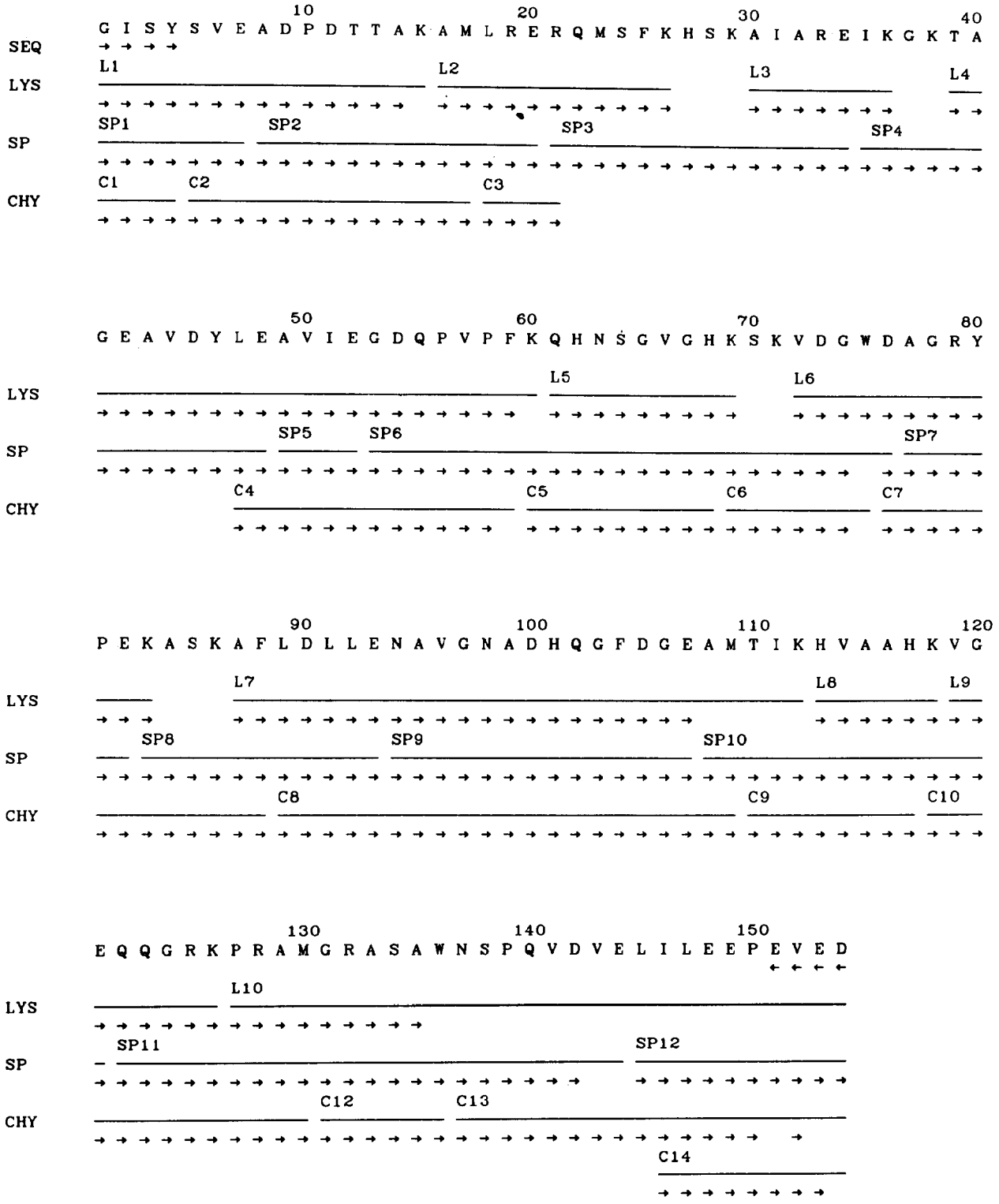


Fig.1A. See page 25 for legend.

B


SEQ
LYS
SP
CHY
L1
L2
L3
L4
SP1
SP2
SP3
SP4
C1
C2
C3
L5
L6
SP5
SP6
SP7
C4
C5
C6
C7
L7
L8
L9
SP8
SP9
SP10
C8
C9
C10
L10
SP11
SP12
C12
C13
C14

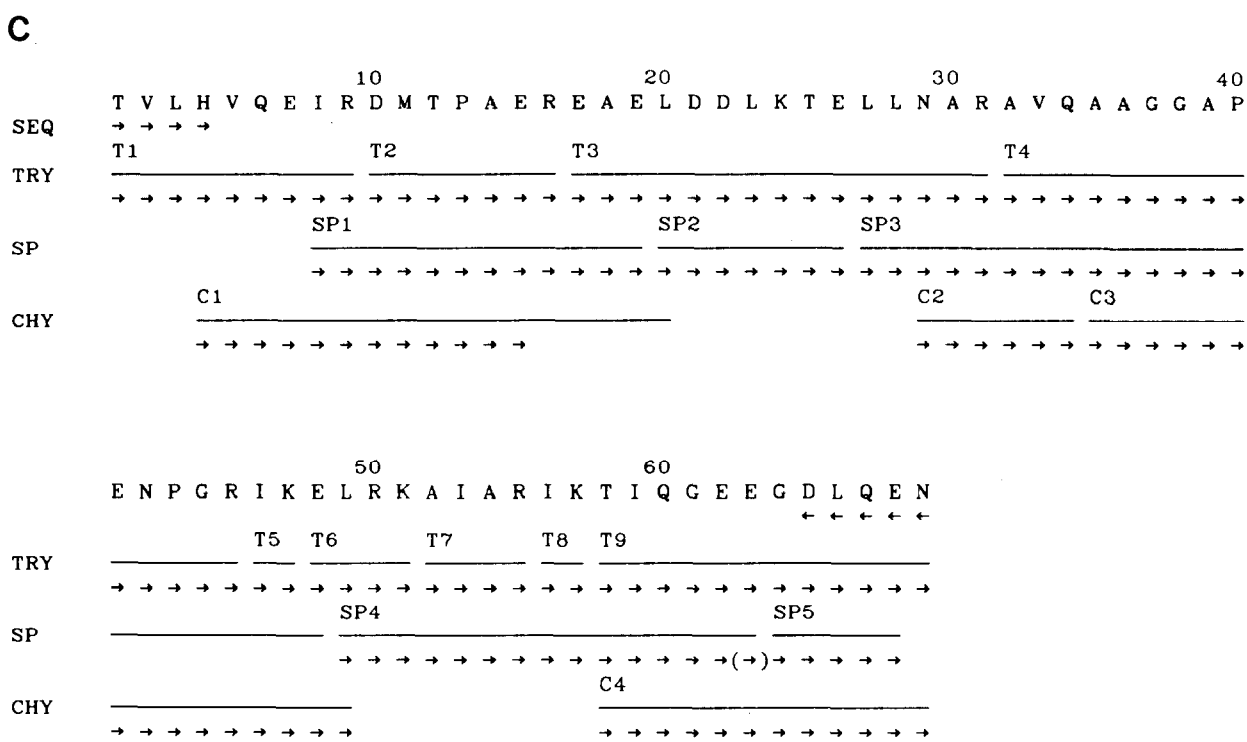


Fig.1. Amino acid sequences of proteins L16 (A), L23 (B) and L33 (C). Arrows show sequencing by the DABITC/PITC double-coupling method [11] and digestion with carboxypeptidase Y [3]. SEQ indicates Edman degradation of intact proteins. TRY, LYS, SP, CHY and THR indicate peptides derived from the digestion with trypsin, lysylendopeptidase, *S. aureus* protease, chymotrypsin and thermolysin, respectively. HCL indicates dilute HCl cleavage.

sent in both HL33 and BL29 (fig.3). A somewhat higher similarity between ribosomal proteins from halophilic bacteria to Gram-positive than to Gram-negative eubacteria has also been found previously [4,6].

There is no indication that proteins HL23 and HL33 are homologous to any of the eukaryotic ribosomal proteins whose primary structures are so far known. Since only a limited number of ribosomal proteins from eukaryotes have been sequenced, it is possible that for proteins HL23 and HL33 their counterparts can be identified when the sequences of more eukaryotic ribosomal proteins have become available.

When protein HL16 was compared to ribosomal proteins from other organisms, it was found that it is related only to the eukaryotic ribosomal protein yeast YL33 (alignment score 7.5) and not to any known ribosomal protein from eubacteria. Since all ribosomal proteins from the Gram-negative *E. coli* and two-thirds of those from the Gram-positive *B. stearothermophilus* have been sequenced [18], it is possible but not very likely that a homology will be found between HL16 and a *B. stearothermophilus* ribosomal protein whose sequence is still not known.

The present data and our previous results [3–8] demonstrate that the ribosomal proteins from the archaebacterium *H. marismortui* are heterogeneous with respect to their homology to ribosomal proteins from other organisms: (i) Proteins HS12, HS15 and HL16 are homologous to ribosomal proteins from eukaryotes whereas proteins HS17, HS11, HL23 and HL33 show homology to eubacterial ribosomal proteins. (ii) For proteins HS6, HL29 and HL31 no homology, either to eubacteria or to eukaryotes, has yet been found but for the reasons mentioned above, i.e. insufficient sequence data for ribosomal proteins from eukaryotes, it is still possible that a homology of these proteins to those of eukaryotes can be demonstrated. (iii) Finally, it has been revealed

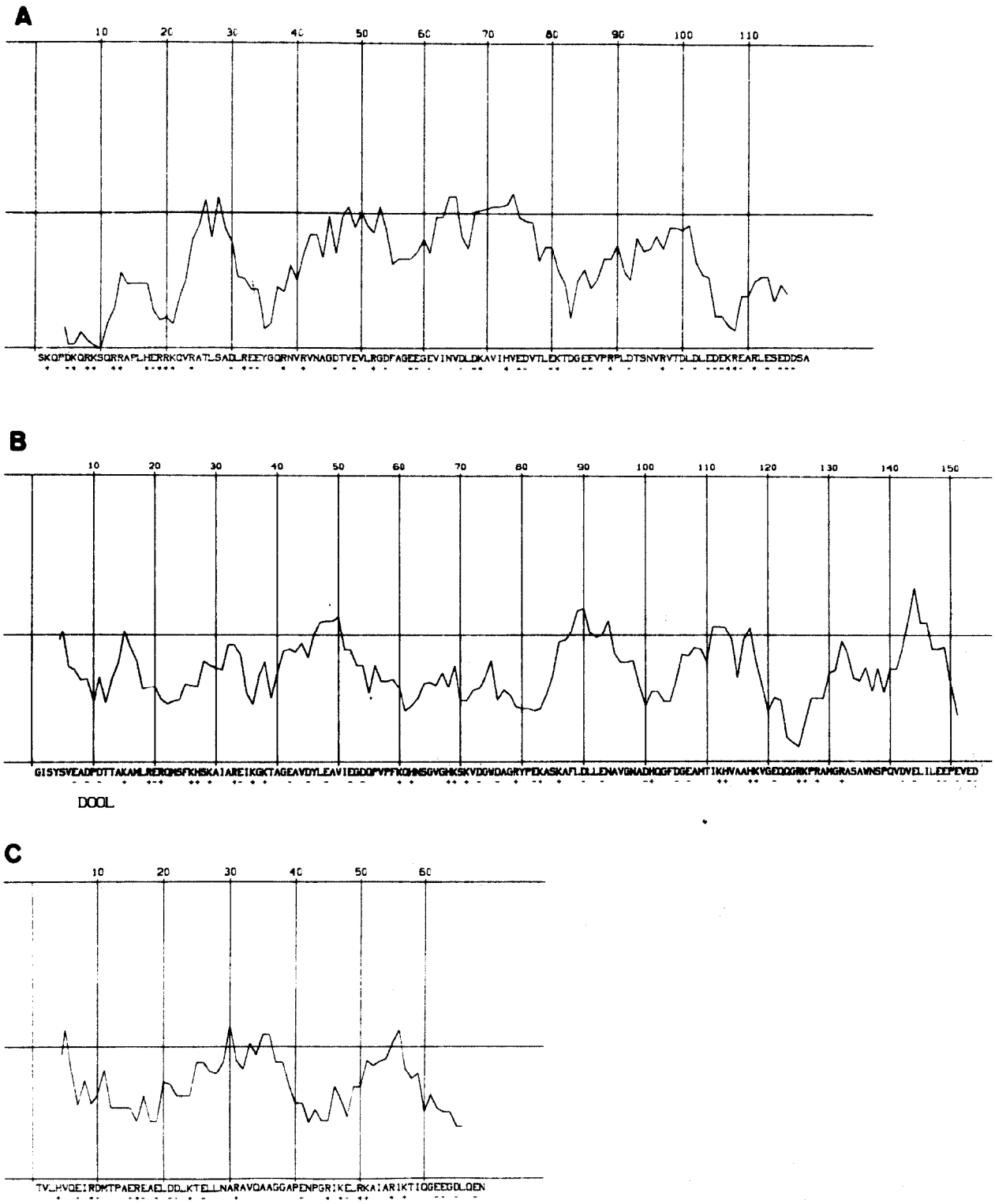


Fig.2. Hydropathic profiles for proteins L16 (A), L23 (B) and L33 (C).

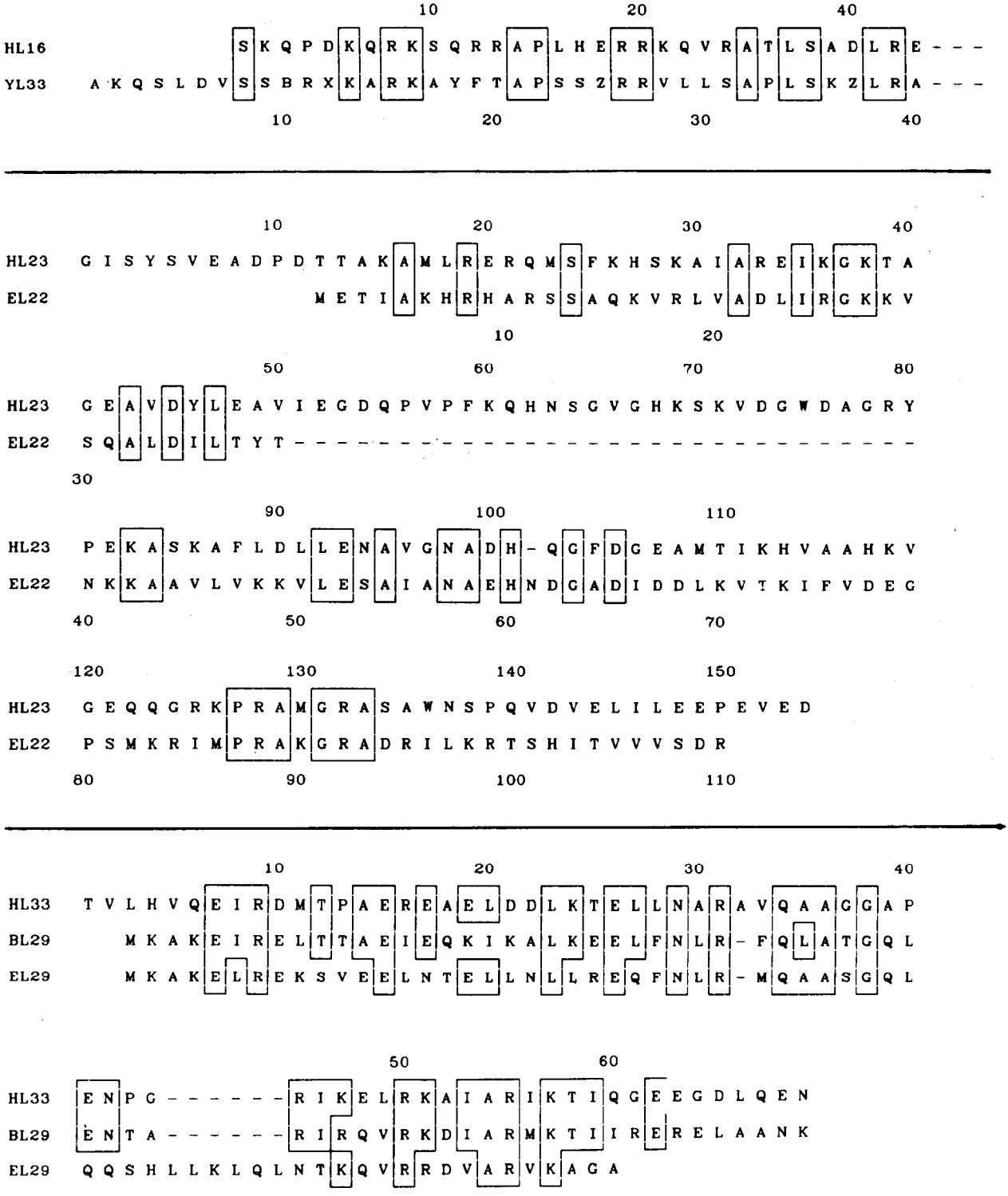


Fig.3. Comparison of the amino acid sequences of *H. marismortui* L16 (HL16), L23 (HL23) and L33 (HL33) with those of homologous proteins from yeast (YL33) [14], *E. coli* (EL22 and EL29) [15,16] and *B. stearothermophilus* (BL29) [17]. Residues identical with those of halobacterial proteins are enclosed in boxes.

Table 1

Degree of homology for each pair of *H. marismortui* L33 (HL33) homologous proteins

Protein	HL33	EL29	BL29
HL33	–	5.3	12.0
EL29	31.6%	–	11.3
BL29	43.9%	45.6%	–

Definitions given in legend to fig.3. The lower left-hand triangle presents values for the identity (%); the upper right-hand triangle gives the alignment scores (in SD units)

that proteins HS14, HS16, HS19 and HL25 are homologous to both eubacteria and eukaryotes [3,4,8].

Acknowledgements: We would like to thank Dr H.G. Wittmann for his interest and encouragement, and Dr J. McDougall for critical reading of the manuscript.

REFERENCES

[1] 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.
[2] Makowski, I., Frolow, F., Saper, M.A., Shoham, M., Wittmann, H.G. and Yonath, A. (1987) J. Mol. Biol. 193, 819–822.
[3] Hatakeyama, T. and Kimura, M. (1988) Eur. J. Biochem. 172, 703–711.
[4] Kimura, J. and Kimura, M. (1987) J. Biol. Chem. 262, 12150–12157.
[5] Kimura, J., Arndt, E. and Kimura, M. (1987) FEBS Lett. 224, 65–70.
[6] Kimura, M. and Langner, G. (1984) FEBS Lett. 175, 213–218.
[7] Arndt, E., Breithaupt, G. and Kimura, M. (1986) FEBS Lett. 194, 227–234.
[8] Kimura, M. and Kimura, J. (1988) FEBS Lett., submitted.
[9] Inglis, A.S. (1983) Methods Enzymol. 91, 324–332.
[10] Ashman, K. and Bosserhof, A. (1985) in: Modern Methods in Analytical Protein Chemistry (Tschesche, H. ed.) pp.151–171, De Gruyter, Berlin.
[11] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[12] Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure, vol.5, suppl.3, National Biomedical Research Foundation, Washington, DC.
[13] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[14] Otaka, E., Higo, K. and Itoh, T. (1984) Mol. Gen. Genet. 195, 544–546.
[15] Wittmann-Liebold, B. and Greuer, B. (1980) FEBS Lett. 121, 105–112.
[16] Bitar, K.G. (1975) Biochim. Biophys. Acta 386, 99–106.
[17] Kimura, M., Kimura, J. and Ashman, K. (1985) Eur. J. Biochem. 150, 491–497.
[18] Wittmann-Liebold, B. (1986) in: Structure, Function and Genetics of Ribosomes (Hardesty, B. and Kramer, G. eds) pp.326–361, Springer, Heidelberg.