Organization and Nucleotide Sequence of a Gene Cluster Coding for Eight Ribosomal Proteins in the Archaeabacterium Halobacterium marismortui*

Evelyn Arndt‡, Wolfgang Krömer, and Tomomitsu Hatakeyama
From the Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 73, D-1000 Berlin 33 (Dahlem), West Germany

A DNA fragment containing the genes for the eight ribosomal proteins HmaL3, H1L6, HmaL23, HmaL2, HmaS19, HmaL22, HmaS3, and HmaL29 from Halobacterium marismortui has been cloned and sequenced. The organization of this gene cluster in general corresponds to the S10 operon of Escherichia coli although there exists some differences between them. The sequence analysis of the 5′- and 3′-region of the gene cluster revealed three open reading frames (orf1, orf2, and orf3) which do not code for any ribosomal protein whose structure is known. A putative promoter is located upstream of orf1. Out of the eight ribosomal proteins five have counterparts in eubacteria only, two in both eubacteria and eukaryotes, and one is exclusively related to an eukaryotic ribosomal protein.

Since Woese and Fox (1) postulated three kingdoms of organisms (eubacteria, eukaryotes, and archaeabacteria), phylogenetic relationships between them have been examined by comparative studies especially on components of the protein synthesis machinery. In addition to the intensive investigations on the primary structures of ribosomal proteins from eubacteria, the amino acid sequences of ribosomal proteins from the extreme halophilic organism Halobacterium marismortui have been determined (3). Sequence comparisons revealed that some halobacterial proteins are related to eubacterial and/or eukaryotic ribosomal proteins, but in some cases there are significant differences between them. Here, we present the nucleotide sequence and the organization of the H. marismortui ribosomal protein gene cluster which corresponds to the S10 operon of Escherichia coli. Simplicities and differences between the operons of these eu- and archaeabacterial species are described. Furthermore, the primary structures of the ribosomal proteins deduced from the nucleotide sequences are compared with those of other organisms.

EXPERIMENTAL PROCEDURES

Materials—M13mp10/11 replicative forms and the 32P-sequencing mixtures were from Pharmacia LKB Biotechnology Inc. DNA polymerase I, nuclease S1, calf intestine phosphorylase, and T4 polynucleotide kinase were from Bethesda Research Laboratory. [γ-32P]ATP (3000 Ci/mmol), [α-32P]dATP (400 Ci/mmol), and [α-35S]dATP (1000 Ci/mmol) were from Amersham Corp. The λ EMBL 3 BamHI arms cloning and packaging system was from Promega.

Oligonucleotide Synthesis, Labeling, and Hybridization—On the basis of the amino acid sequence of the ribosomal protein HmaL23 (H1L23) (6), a mixed oligonucleotide with 26 bases (Tables I) was synthesized by the solid-phase phosphoramidite method using an Applied Biosystems 380A oligonucleotide synthesizer. The sequence of the oligonucleotide mixture was designed by the ribosomal protein database included in the NBRF mutation data matrix was used as scoring matrix, and a break penalty of 20 was employed.

DNA Sequence Analysis—Restriction fragments to be sequenced were ligated into M13mp10 or mp11 vectors. The sequence analysis was carried out by the dideoxy chain termination method (12) with a 17-mer primer from Pharmacia.

Computer Analysis—The computer analysis was carried out by the programs from the University of Wisconsin Genetics Computer Group (Version 5.3, July 1988) (13) on a VAX/VMS computer. The alignment program for archaebacterial ribosomal proteins have been carried out in order to compare not only their sequences but also their gene organization with those of other organisms (7–10). From these studies it became clear that the gene organization in archaebacteria is essentially similar to that in eubacteria, although some differences in this respect exist between the two kingdoms.

Here, we present the nucleotide sequence and the organization of the H. marismortui ribosomal protein gene cluster which corresponds to the S10 operon of Escherichia coli. Similarities and differences between the operons of the two eu- and archaeabacterial species are described. Furthermore, the primary structures of the ribosomal proteins deduced from the nucleotide sequences are compared with those of other organisms.

The abbreviations used are: bp, base pair; kb, kilobase pair.

† To whom correspondence should be addressed. Fax: 49-30-8307380.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM/EMBL Data Bank with accession number(s) J05222.
**RESULTS AND DISCUSSION**

**Cloning of the HmaL23 Gene from Genomic DNA**—After Southern hybridization with the 26-mer oligonucleotide for HmaL23 described under “Experimental Procedures,” a single band was detected for each restriction digest of *H. marismortui* DNA (not shown). DNA fragments of ~0.45 kb in size which were detected in the SalI digest were cloned into the replicative form of M13mp10. Positive clones were identified by Southern hybridization of the SalI-digested replicative forms. One of the hundred tested clones gave a positive signal in the autoradiography and was sequenced. In this way the presence of the HmaL23 gene could be identified.

**Screening of the Genomic Library**—15-20 kb fragments of partial digested chromosomal DNA from *H. marismortui* were ligated into EMBL 3 BamHI arms and packaged with “Packagen” extracts. The cloned ~0.46 kb SalI fragment containing the HmaL23 gene was nick-translated and hybridized to the EMBL 3 genomic library from *H. marismortui* under highly stringent conditions in the presence of 50% formamide (11, 16). The positive clone PP-7 with a 1.6 kb *H. marismortui* DNA insert of about 15 kb was selected for further study. Figure 1 shows the 6338-bp region carrying several ribosomal protein genes which were characterized by restriction fragment analysis and sequence determination. Digestion with Smal, PoulI, PotI, and SalI yielded relatively large fragments used for mapping studies. Smaller fragments for subcloning and sequencing were obtained by digesting large fragments with AluI, RalI, Sall, and Sau3AI.

**Nucleotide Sequence Data**—Figure 2 shows the nucleotide sequence (6338 bp) of the PP-7 insert DNA which has been determined so far. By comparison with the partial or complete amino acid sequences determined from the purified proteins, we identified eight ribosomal protein genes and three other open reading frames in this region, namely the genes for the ribosomal proteins HmaL3 (HL1; positions 885–1898), HL6 (positions 1905–2642), HmaL23 (HL25; positions 2642–2896), HmaL2 (HL4; positions 2902–3621), HmaS19 (HS18; positions 3621–4040), HmaL22 (HL33; positions 4046–4510), HmaS3 (HS1; positions 4513–5424), HmaL29 (HL33; positions 5424–5636), and the open reading frames orf1 (positions 36–677), orf2 (positions 677–877), and orf3 (from position 5639 on).

In this gene cluster the translation start codon ATG of orf2, HmaL23, HmaS19, HmaS3, HmaL29, and orf3 overlap with the stop codons of the preceding genes.

Methionine residues specified by initiator ATG appear to be cleaved off during or after translation except for HL6 since N terminal methionines were not found in these proteins by N-terminal sequence analysis.

The 3′-terminal region of the 16S rRNA of *H. marismortui* is 3′-UCCUCCACU. Therefore, if there is an interaction between the 3′-region of 16S rRNA and the translation initiation site of mRNA as shown in eubacteria (19), the “ideal” sequence for this site would be AGGAGGTGA. Comparison of the sequence regions around the ATG start codon demonstrates that some of these regions are in good agreement with the ideal sequence. For example, the upstream region of the ATG start codon for HmaL23 contains 7 nucleotides which are complementary to the 3′-end of the 16S rRNA. On the other hand, others such as the 5′-regions for HL6 and HmaL22 show very weak agreement with the consensus sequence. In the *E. coli* S10 operon, the ribosomal binding sites also vary in extent and location (51), but an apparent lack of this region (e.g. before HmaL22) is observed here for the first time. However, the function of these complementary sequences has not yet been established in halobacteria.

**Gene Organization**—Figure 3 illustrates the comparison of the organization of the investigated ribosomal protein genes from *H. marismortui* and the S10 operon from *E. coli*. The order of the genes in *H. marismortui* essentially agrees with that of the corresponding genes of *E. coli*. The genes for the L3, “L4” (HL6), L23, L2, S19, L22, S3, and L29 equivalent proteins have the same order in both gene clusters. Similarities of the ribosomal protein gene operons between archaeabacteria and eubacteria have also been revealed for the “A” protein operon of *Halobacterium halobium* (7), *Halobacterium cutirubrum* (6), Methanococcus vannielli (20, 21), and *Sulfolobus solfataricus* (22), as well as for the “S10 operon” (20, 23) and the “spc operon” of *M. vannielli* (10). On the other hand, the following differences were found. 1) There are two open reading frames (orf1 and orf2) upstream of the HmaL3 gene and one open reading frame (orf3) downstream of the HmaL29 gene. It is apparent that these open reading frames correspond to neither the HmaS10 nor the HmaS17 gene since the amino acid sequences deduced from these open reading frames drastically differ from the amino acid sequences determined by the Edman degradation method using the purified proteins.2) In *M. vannielli*, the gene for the protein which is homologous to EcoS10 has been found next to the “str operon” which is located about 30 kb away from the “S10-sp operon region” (10). It seems likely that this situation is also true for *H. marismortui*. 3) Based on the location of the gene within the cluster, protein HL6 is expected to be equivalent to EcoL4. However, the amino acid homology between these two proteins is so weak (see below) that this conclusion cannot be drawn with certainty. 4) The gene for the protein which is homologous to EcoL18 is missing in this halobacterial operon. A similar observation was made for the equivalent region in *H. halobium* (O).

**Codon Usage**—Table II shows the frequency of codon usage in the *H. marismortui* ribosomal protein genes described in this paper as well as in the HmaS15 gene reported previously (11). In general, the G/C rich codons (especially those with G or C in the third position) are highly preferred. This is consistent with the results obtained for the genes of other halophilic organisms (8, 24, 25). The data in Table II also show the high content of acidic amino acid residues (glutamic acid and aspartic acid) as well as valine, alanine, and glycine in the ribosomal proteins of *H. marismortui*. Interestingly, these are all amino acids whose codons begin with G. This may reflect the situation that a selection pressure for G has occurred also at the first position during evolution as an

---

**Table I**

Oligonucleotide probe for the isolation of the HmaL23 gene

<table>
<thead>
<tr>
<th>Position</th>
<th>Protein sequence</th>
<th>mRNA sequence</th>
<th>26-mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Met-Asn-Asp-Met-Asp-Phe-Gln-Asn-Lys</td>
<td>5′-AUG AAC GAC AUG GAC UUC CAG AAC AA-3′</td>
<td>3′-TAC TGT CAT GTC TGT TT-5′</td>
</tr>
<tr>
<td>23</td>
<td>U</td>
<td>U</td>
<td>A</td>
</tr>
</tbody>
</table>

by the prefix “Hma,” e.g. HmaL2, HmaL3, etc. When this is not the case, the proteins are designated by the prefix “H,” e.g. HL6, according to their positions on the two-dimensional gel electrophoresis (6, 18).

---

* M. Bronombre, personal communication.
* M. Kimura, personal communication.
adaptation mechanism to high salt conditions. Similarly, C is used for the first positions for leucine and arginine in preference to U. The total G + C content within the structural genes is 66% and the preference for G and C in the third positions of the codons is 83%.

Promoter and mRNA Start—The determination of the mRNA starting point was carried out by S1 nuclease mapping analysis performed on total RNA extracted from exponentially growing cells, as described under "Experimental Procedures." The assay revealed that transcription of a major mRNA begins 35 nucleotides upstream of the ATG start codon of orf1. There are three promoter-like structures "S10P1," "S10P2," and "S10P3" upstream of this region. Figure 4 shows a comparison of these putative promoter regions with 24 other promoter sequences in halophilic bacteria. Among these sequences there are no conserved "-10" and "-35"-regions which were found for most eubacterial promoters (37). However, an AT-rich region (box A) which resembles the TATA-box in eukaryotes (38) was found in a distance of 22-33 bp from the transcriptional starting points.

Accordingly, it seems that mainly "S10P1" functions as a promoter for the H. marismortui gene cluster since it is located 22 bp upstream from the transcriptional initiation site. This would correspond to the situation with other archaeabacterial promoters.

Comparison of the Protein Structures—The molecular masses of the ribosomal proteins HmaL3, HL6, HmaL2, HmaS19, HmaL22, HmaS3, and HmaL29 and the putative proteins translated from orf1 and orf2 were calculated to be 37,127, 26,423, 9,471, 25,267, 15,865, 16,812, 33,593, 7,749, 23,172, and 7,062, respectively. No homologous protein in the data bank has so far been found for the putative gene products encoded by orf1, orf2, and orf3.

In the following, the structures of those proteins are discussed which are described here for the first time.

HmaL3—Comparison of the deduced amino acid sequences of the H. marismortui protein HmaL3 with the sequences of all ribosomal proteins in the data bank revealed that HmaL3 is homologous to EcoL3 (identity 28%, alignment score 10.4 S.D. units) as well as to the yeast ribosomal protein YeaL3 (identity 34%, alignment score 34.3 S.D. units). Figure 5 shows the alignment of the three proteins by the program ALIGN. As seen from the boxes, identical residues are not clustered but distributed over the entire length of the protein chains. Interestingly, HmaL3 exhibits similarities to either EcoL3 or YeaL3 with respect to the introduced gaps. Positions 60-70, 139-185, and the N-terminal region are common to HmaL3 and YeaL3, and positions 114-126 to HmaL3 and EcoL3. This observation apparently suggests that the halo-bacterial protein has intermediate features between yeast and E. coli. In other words, HmaL3 may have retained its original structure most among these three homologous proteins. As has been observed for other H. marismortui ribosomal proteins showing homology with both eubacterial and eukaryotic counterparts (3), HmaL3 also shows a higher homology to the eukaryotic protein than to the eubacterial counterpart. The homology between EcoL3 and YeaL3 is the lowest based on the alignment score (6.3 S.D. units) although the number of identical amino acids (29%) is approximately the same (28%) as with HmaL3 and EcoL3.

HL6—As mentioned above, protein HL6 has little homology to EcoL4 although the two proteins are located at the equivalent positions within the respective gene clusters (Fig. 3). As shown in Fig. 6a, many gaps are required for the best alignment of these proteins. The percentage of identical amino acids and the alignment score are only 21% and 2.7 S.D. units, respectively.

On the other hand, HL6 is homologous to a protein in eukaryotes, namely XL1B from Xenopus laevis. 34% identical amino acids and an alignment score of 24 S.D. units demonstrate the high homology between these proteins as shown in Fig. 6b by the alignment of both proteins. XL1B is much longer (396 versus 246 amino acid residues). HL6 fits very well with the N-terminal part of XL1B. There is only a very weak homology between EcoL4 and XL1B (identity 20%, alignment score 0.5 S.D. units).

Based on these findings one may speculate that HL6 is an extra protein which is homologous only to an eukaryotic ribosomal protein and that the gene for the true protein "HmaL4" is translocated to another gene locus on the H. marismortui genome or is lacking at all. There is, however, also the possibility that EcoL4 has undergone drastic structural changes during evolution, leading to only a weak homology to HL6 and XL1B.

HmaL2—As shown in Fig. 7 HmaL2 has a relatively high degree of homology to the archaeabacterial ribosomal protein L2 from M. vanielii (MvaL2) as well as to the eubacterial proteins EcoL2 and BstL2. The best conservation of the sequences was found between the archaeabacterial proteins HmaL2 and MvaL2 with 59% identical residues and an alignment score of 66.9 S.D. units. About 40% identity with both E. coli and Bacillus stearothermophilus ribosomal proteins and alignment scores of 31.5 and 27.3 S.D. units indicates that this protein family is one of the most conserved ribosomal proteins among archaeabacterial and eubacterial ribosomal proteins, suggesting a critical role in the structure and/or function of the ribosome. The eubacterial L2 proteins have been found to be involved in the peptidyltransferase activity (45, 46). Furthermore, the central region (excluding the N-terminal and C-terminal 40 residues) was shown to interact with the 23 S rRNA (47). The occurrence of most of the conserved amino acid residues in the central region and the lack of the N-terminal 40 residues in both archaeabacterial proteins indicates that the functionally important region is present in the eubacterial and in the archaeabacterial ribosomal proteins.

HmaS19—The amino acid sequence of HmaS19 is homologous to the eubacterial protein S19 as shown in Fig. 8. This is obvious from the number of identical amino acids, 37% for HmaS19 and EcoS19 and 34% for HmaS19 and BstS19 as well as from their alignment scores with 14.1 and 9.5 S.D. units, respectively. The N-terminal serine residue of HmaS19...
Ribosomal Protein Gene Cluster of H. marismortui

H. marismortui

E. coli

FIG. 3. Organization of the gene cluster of H. marismortui in comparison with the corresponding E. coli S10 operon (26).

TABLE II
Codon usage in the H. marismortui ribosomal protein genes HmaL3, HmaL6, HmaL23, HmaL2, HmaS19, HmaL22, HmaS3, and HmaL29

The values are numbers of amino acids (total 1729).

<table>
<thead>
<tr>
<th>Codon</th>
<th>HmaL3</th>
<th>HmaL6</th>
<th>HmaL23</th>
<th>HmaL2</th>
<th>HmaS19</th>
<th>HmaL22</th>
<th>HmaS3</th>
<th>HmaL29</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UUC</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>UUA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UUG</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CUU</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>CUC</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CUA</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>CUG</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>GGU</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GGU</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GGA</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>GGG</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

is blocked by an acetyl group. The N terminal region of the archaeabacterial protein is considerably longer in comparison to the eubacterial counterparts as shown by the alignment (Fig. 8).

HmaS3—Although the number of identical residues between HmaS3 and the eubacterial proteins EγS3 and BαS3 is relatively low (27%), the alignment scores (16.3 and 17.1...
FIG. 4. Promoter sequences of the halophilic archaeabacteria *H. cutirubrum* (*H. cu.*), *H. halobium* (*H. ha.*), *H. volcanii* (*H. vo.*), *Halococcus morrhuae* (*H. mo.*), and *H. marismortui* (*H. ma.*). The underlined small characters mark the first transcribed nucleotide. The underlined nucleotides next to the box show conserved regions between groups of promoters. Most sequences are also listed in Ref. 27. Asterisk (*) indicates promoter sequences found in this paper.

S.D. units) gave significantly high values to prove the homology between HmaS3 and the two eubacterial proteins. The alignment of these proteins in Fig. 9 shows that there is a very highly conserved region (positions 50-59 of HmaS3) among the molecules. Another remarkable region of homology is from position 88 to 97 of HmaS3 especially with regard to conservative amino acid replacements, e.g., R-93 in HmaS3 to K in EcoS3/BstS3 and I-94 in HmaS3 to L in EcoS3/BstS3. Interestingly, HmaS3 is about 100 amino acids longer at the C terminus than the eubacterial proteins EcoS3 and BstS3, and there is a very high number of acidic amino acids concentrated in this region (positions 188-304). The extraordinary acidic nature of the C-terminal part of HmaS3 could be the result of an adaptation to high salt concentrations as has been suggested for the *H. halobium* HhaS3 protein (9).

FIG. 5. Alignment of HmaL3 from *H. marismortui* with the homologous proteins YeaL3 from yeast (39) and EcoL3 from *E. coli* (40). The proteins are aligned for maximal homology. Identical residues are boxed.

FIG. 6. a, alignment of HL6 from *H. marismortui* with EcoL4 from *E. coli* (41). The alignment demonstrates the weak homology of both proteins in which many gaps are required to obtain maximal homology. b, alignment of HL6 from *H. marismortui* with the homologous protein from *X. laevis* XL1B (42).
Ribosomal Protein Gene Cluster of H. marismortui

Concluding Remarks—Among the ribosomal proteins encoded in the cluster described in this paper, two proteins (HmaL3 and HmaL23) were found to be homologous to both eubacterial and eukaryotic counterparts, whereas the proteins HmaL29, HmaL2, HmaS19, HmaS3, and HmaL22 are only homologous to eubacterial counterparts. The protein HL6 (orfl, orf2, and orf3) also code for ribosomal proteins or homologous proteins EcoS3 from E. coli (50) and BstS3 from B. stearothermophilus.6

Fig. 8. Alignment of HmaS19 from H. marismortui with the homologous proteins EcoS19 from E. coli (48) and BstS19 from B. stearothermophilus (49).

Fig. 9. Alignment of HmaS3 from H. marismortui with the homologous proteins EcoS3 from E. coli (50) and BstS3 from B. stearothermophilus.

Acknowledgments—We thank Drs. H. G. Wittmann and B. Wittmann-Liebold for their interest and encouragement.

REFERENCES

Organization and nucleotide sequence of a gene cluster coding for eight ribosomal proteins in the archaebacterium Halobacterium marismortui.
E Arndt, W Krömer and T Hatakeyama


Access the most updated version of this article at http://www.jbc.org/content/265/6/3034

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/6/3034.full.html#ref-list-1