Volume 314, number 3, 211–214 FEBS 11865 December 1992

Nucleotide sequence of the genes for ribosomal proteins HS15 and HSH from *Haloarcula marismortui*: an archaeon-specific gene cluster

Evelyn Arndt and Christina Steffens

Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestr. 73, D-1000 Berlin 33, Germany

Received 28 August 1992; revised version received 28 October 1992

The nucleotide sequences of the genes for two ribosomal proteins, HS15 and HSH, from the archaeon *Haloarcula marismortui*, have been determined. The genes were found in a cluster together with another open reading frame with a probable regulatory function. HS15 and HSH have counterparts in eucarya. HS15 is significantly homologous to S19 from frog (*Xenopus laevis*). HSH is related to S37 from yeast (*Saccharomyces cerevisiae*) and S27 from fly (*Drosophila melanogaster*), as well as to other members of the S27 family. Eubacterial counterparts were not found, suggesting that these proteins are 'extra proteins' that are absent in eubacterial ribosomes.

Ribosomal protein; Archaea; Evolution

1. INTRODUCTION

A molecular model of the ribosome is required to understand in detail the events taking place during protein biosynthesis. Currently, much effort is being invested to elucidate the three-dimensional structure of the ribosome and its subunits. The best crystals of the 50 S ribosomal subunit, diffracting to 3 Å resolution, were obtained from the halophilic archaeon, *Haloarcula marismortui*, and used for X-ray crystallography [1]. For the interpretation of these structural data, it is necessary to know the primary structures of all the macromolecules involved.

Furthermore, the comparison of the primary structures of the ribosomal proteins offers a promising tool for establishing phylogenetic relationships between the archaea, eucarya and eubacteria. In addition, these data may provide valuable information about important functional sites, since regions crucial for ribosomal assembly or the translational process are expected to be conserved throughout evolution [2,3]. For this purpose, complete sets of ribosomal protein sequences from all three kingdoms need to be evaluated.

In order to establish the ribosomal protein sequences of *H. marismortui*, protein chemical methods, as well as gene cloning techniques, were applied. So far, the primary structures of 45 ribosomal proteins have been established [4-7]. Advantage was taken of the fact that many ribosomal protein genes in *H. marismortui* are clustered in operon structures, as in the eubacterium, *Escherichia coli*, and that the gene order is quite conserved. We also characterized one operon which we proposed as an additional archaeal gene cluster in *H. marismortui*, comprising two ribosomal protein genes whose gene products are exclusively related to eukaryotic counterparts [8]. In this paper, we present a gene cluster comprising another two 'extra' ribosomal protein genes and an open reading frame, as well as characteristics of the resulting gene products.

Correspondence address: E. Arndt, Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestr. 73, D-1000 Berlin 33, Germany. Fax: (49) 8307 380.

2. EXPERIMENTAL

2.1. *Materials*

Sequencing vectors, enzymes and other materials were used as in [9].

2.2. *Cloning of the gene for ribosomal protein HS15*

The N- and C-terminal amino acid sequences of the HS15 protein [10] were backtranslated, and two oligonucleotides were chosen on the basis of the codon usage table for ribosomal protein genes of *H. marismortui* [9]. The two primers, complementary to the antisense and sense strands of the 5′ and 3′ ends of the coding sequence of the HS15 gene were synthesized by the solid-phase phosphoamidite method using an Applied Biosystems 380A oligonucleotide synthesizer. As shown in Fig. 1, the rightward oligonucleotide (N-oligo) included an *Eco*RI site followed by 28 bases of the expected coding sequence of the gene, including the initiator ATG codon. The leftward oligonucleotide (C-oligo) included a *Bam*HI site, the reverse complement of the stop-codon and 27 bases reverse complementary to the 3′ end of the gene. Three bases were added on each side of the recognition sites in order to obtain efficient cutting by the restriction endonucleases.

Genomic DNA of *H. marismortui* was used as the template for PCR reactions using *Taq* polymerase (Amersham, Braunschweig). The reaction volume was 100 μl containing 100 ng template DNA, 100 nmol of each N- and C-oligo, 200 μM of each of the 4 dNTPs, and 10 μl of 10 × reaction buffer supplied with the *Taq* polymerase. After denaturation of the DNA for 5 min at 94°C, 2 U *Taq* polymerase was added to the reaction mixture. The amplification was done with 30 repeated cycles of melting at 94°C for 2 min, hybridization at 47°C for 2 min and primer extension at 72°C for 2 min. Finally, the reaction

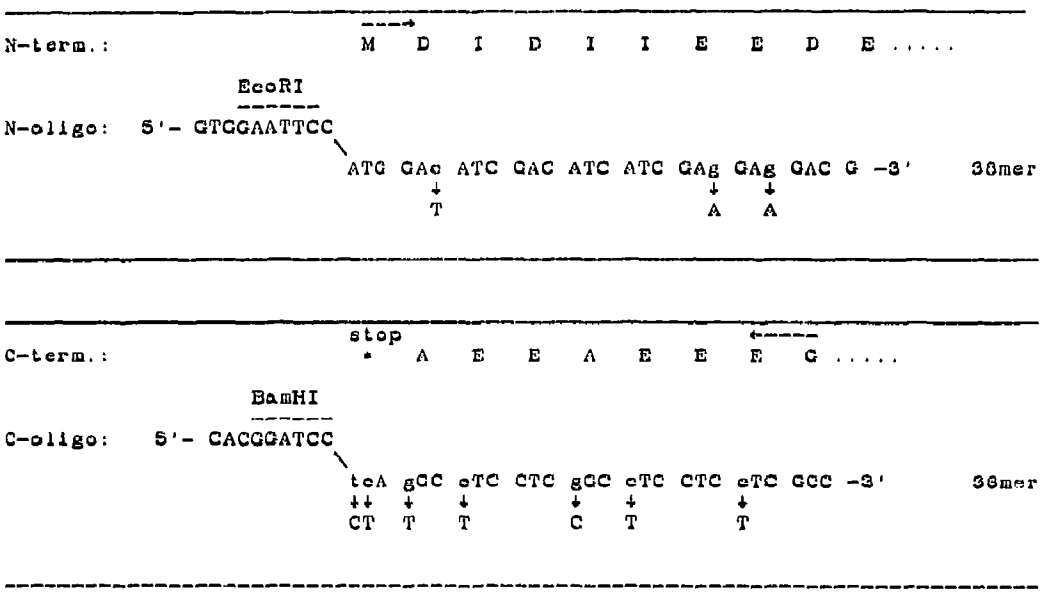

Fig. 1. Oligonucleotides for the HS15 gene used in the PCR reaction with chromosomal DNA of *H. marismortui*. The small letters in the sequence represent mismatches in the oligonucleotides, and the nucleotides which are present in the native gene are indicated by arrows.

mixture was heated for 20 min at 70°C and subsequently refrigerated. The PCR reaction mixture was separated on a 2% agarose gel. The band corresponding to the expected length of 328 bp was excised and electroeluted from the agarose gel. The DNA fragment was cut with *Eco*RI and *Bam*HI and cloned into the replicative form of M13mp10. Appropriate clones were sequenced by the dideoxy chain-termination method [11].

2.3. *Computer analysis*

The deduced amino acid sequences were compared with proteins in the NBRF protein data base (release 31.22, 1992) and the RIBO database (available in this Institute) using computer programs from the University of Wisconsin Genetic Computer group (version 6.2, 1990) [12] on a VAX/VMS computer. The program, ALIGN, was used to obtain maximal homology between related proteins [13]. The mutation data matrix was applied as scoring matrix, and a break penalty of 20 was employed.

2.4. *Nomenclature*

The protein, HS15, was named according to its migration on the two-dimensional gel electrophoresis system [14]. The arbitrary designation HSH is from [15].

3. RESULTS AND DISCUSSION

3.1. *Nucleotide sequence of the HS15 operon*

Sequencing of a M13 clone obtained by the procedure described in section 2 revealed that it encodes the HS15 gene. The protein sequence derived from the DNA se-

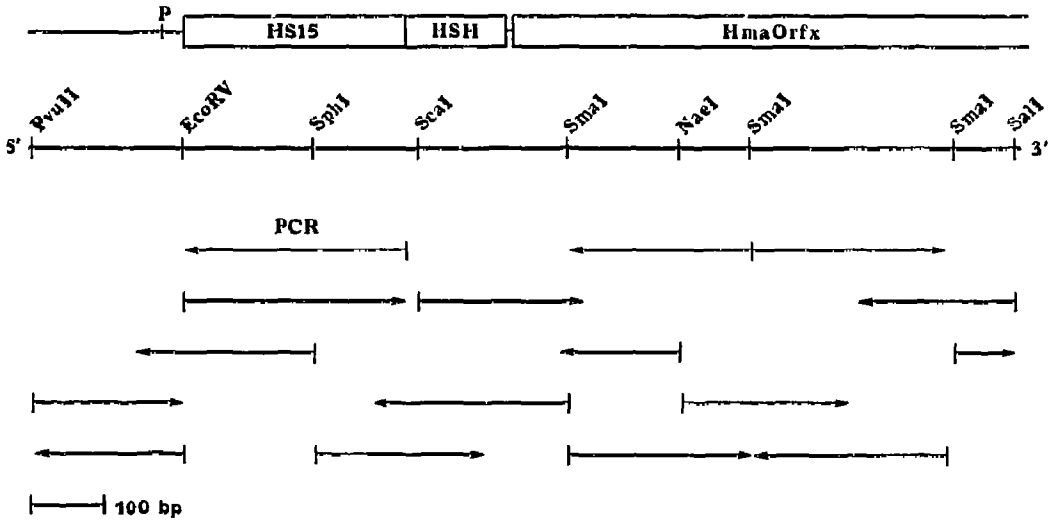

Fig. 2. Positions of genes and restriction endonuclease sites used for subcloning in the investigated DNA region. The nucleotide sequences of both DNA strands were determined, and all cloning sites were traversed as indicated by the arrows.

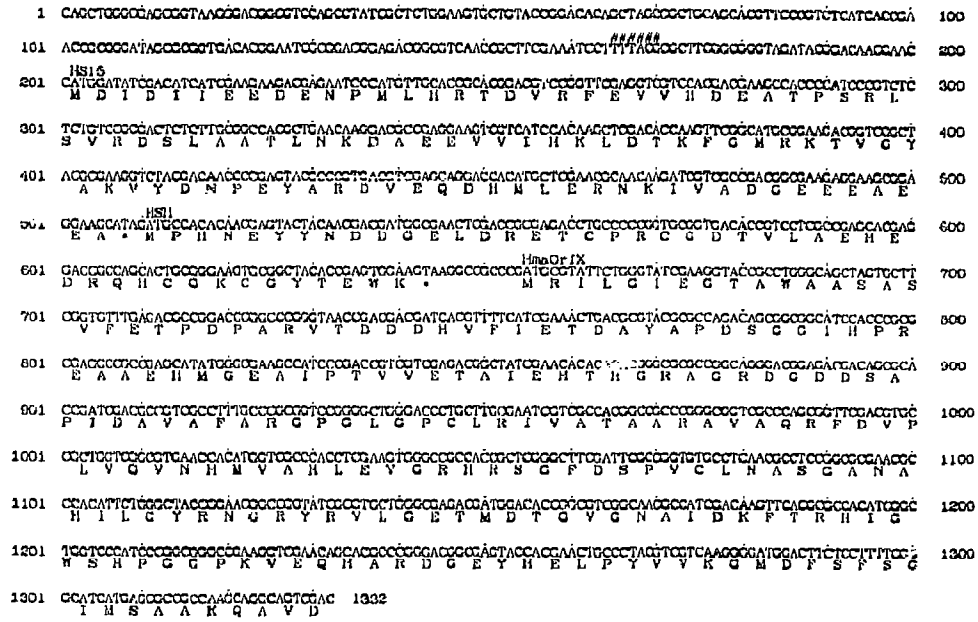

Fig. 3. Nucleotide sequence of a 1,332 bp genomic DNA region of *H. marismortui*. The deduced primary structures of the proteins are written below the DNA sequence. The gene for HS15 corresponds to positions 202–507, that for HSH to positions 511–642, and the HmaOrfx gene starts at position 655. From the amino acid sequence analysis, the initiator methionine of the gene sequence is not present in the mature protein HSH [15]. #, putative promoter structure.

quence was in complete agreement with the amino acid sequence reported by Kimura et al. [10].

The 328 bp insert DNA containing the HS15 gene was nicktranslated and used as a probe in Southern hybridization with digested genomic DNA of *H. marismortui* (not shown). Two hybridizing fragments, a ~1.3 kb *Sma*I–*Sph*I and a 950 bp *Sph*I–*Sal*I fragment were detected and chosen for cloning. Because the HS15 gene contains the rare cutting site for *Sph*I, these fragments offer the opportunity of determining the sequence of the 5′ as well as the 3′ region of the native HS15 gene: this allows the study of regulatory sequences such as promoter or terminator structures or additional genes within a gene cluster.

Genomic DNA was digested with *Sma*I/*Sph*I or *Sph*I/*Sal*I and separated on 0.8% agarose gels. The ~1.3 kb and ~950 bp fragment mixtures were excised, electroeluted from the agarose gel and cloned into the sequencing vectors, M13mp18 or M13mp19. Positive clones were identified by Southern hybridization of the *Sma*I–*Sph*I- or *Sph*I–*Sal*I-digested replicative forms. Clones containing the 5′ or 3′ part of the HS15 gene were selected and their subclones were sequenced according to the strategy shown in Fig. 2. The nucleotide sequence of 1,332 bp, comprising the HS15 gene, as well as two other open reading frames in its 3′ region, is shown in Fig. 3.

Surprisingly, the open reading frame which follows directly after the HS15 gene encodes another ribosomal protein gene, namely the HSH gene (see below). The HSH gene is followed by an open reading frame (HmaOrfX) which does not finish within the sequenced region. In the 5′ region of the HS15 gene, a DNA stretch was found (positions 165–170 in Fig. 3) which corresponds well with the consensus sequence of halobacterial promoters [9]. Thus, we suggest that HS15 is the first gene of this operon.

3.2. *Protein HS15*

The complete amino acid sequence of HS15 has been published [10]. In that work, the homology to a eukaryotic counterpart from *Xenopus laevis* (FrogL19) [16] was described. Corresponding proteins from eubacteria were not found. The protein sequence deduced from the gene is in complete agreement with that from the investigated protein sequence.

3.3. *Protein HSH*

The sequence of 16 N-terminal amino acid residues established by Edman degradation of the purified protein HSH has been published [15]. This N-terminal amino acid sequence (PGNKYYNDEGXLDPXT) agrees rather well with the N-terminal sequence of the protein predicted by DNA translation in this work. The complete HSH protein consists of 43 amino acid residues with a molecular mass of 5,071 Da and an isoelectric point of 4.55. Corresponding proteins are YeaS37 from *Saccharomyces cerevisiae* [17,18], FlyS27 from *Drosophila melanogaster* [19], as well as other members of the eukaryotic protein S27 family. HSH belongs to the group of archaeal ribosomal proteins for which no eubacterial counterparts have been found. In eucarya,

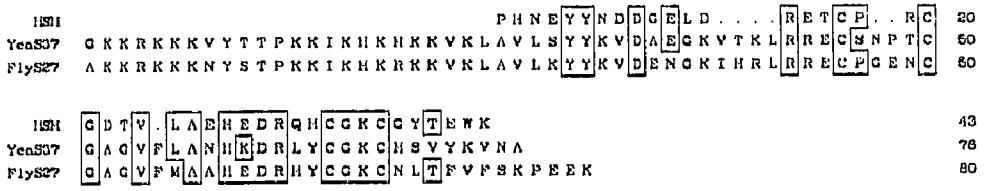

Fig. 4. Alignment of HSH from *H. marismortui* with the eukaryotic counterparts, YeaS37 from *S. cerevisiae* [17,18] and FlyS27 from *D. melanogaster* [19]. The proteins are aligned for maximal similarity. Identical residues are boxed, and gaps are marked by dots.

the S27 genes are associated with the ubiquitin genes and transcribed as tails of ubiquitin [18]. Fig. 4 presents the alignment of the homologous molecules. The protein pair, HSH/YeaS37, shares 44% identical amino acid residues and an alignment score of 5.18 S.D. units. The protein pair, HSH/FlyS27, shares 44% identical amino acid residues and an alignment score of 4.51. The N-terminal parts of the eukaryotic proteins are extended by 24 amino acid residues, which are predominantly basic. Interestingly, all 4 cysteines in the related molecules are conserved. A putative metal-binding, nucleic acid binding domain is formed by the motif C-X_{2-4}-C-X_{13-14}-C-X_2-C, suggesting that these proteins may bind to rRNA [20].

3.4. *HmaOrfx*

The putative gene product of the open reading frame (HmaOrfx), located 3′ of the HSH gene, was compared to the data bases. A corresponding gene has been described in *E. coli* (orf_x) [21]. In that work, it was shown that orf_x is expressed in exponentially growing *E. coli* cells. The orf_x is one of the orf's which was found in the 5′ region of the rpsU(EcoS21)-dnaG-rpoD operon (67 min in the *E. coli* genome). The authors suggested that these orf's may be involved in the regulation of the expression of the rpsU-dnaG-rpoD operon. Orf_x is located directly 5′ to the *E. coli* rpsU gene but is transcribed in the opposite direction.

Interestingly, the corresponding gene in *H. marismortui* is co-transcribed with two ribosomal protein genes. Thus it is likely that the gene product of HmaOrfx has some regulatory function in this operon. It is reasonable to propose that the HmaOrfx has the same function in archaeal and eubacterial cells, because it is conserved during evolution.

3.5. *Concluding remarks*

The gene cluster described in this paper is one of the additional gene clusters encoding 'extra ribosomal proteins' which are present in archaea but not in eubacteria [22-24]. Nevertheless, both ribosomal gene products, HS15 and HSH, have counterparts in eucarya. A gene for ubiquitin, which is located in front of the eukaryotic HSH counterparts, is missing in this gene arrangement.

Acknowledgements: We thank Dr. A.R. Subramanian and Dr. R. Brimacombe for critical reading of the manuscript.

REFERENCES

[1] von Böhlen, K., Makowski, I., Hansen, H.A.S., Bartels, H., Berkovitch-Yellin, Z., Zaytzev-Bashan, A., Meyer, S., Paulke, C., Franceschi, F. and Yonath, A. (1991) J. Mol. Biol. 222, 11–15.
[2] Wittmann-Liebold, B., Köpke, A.K.E., Arndt, E., Krömer, W., Hatakeyama, T. and Wittmann, H.G. (1990) in: The Ribosome: Structure, Function and Evolution (Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. and Warner, J.R. eds.) pp. 598–616, ASM Publications, Washington, DC.
[3] Matheson, A.T., Auer, J., Ramirez, C. and Böck, A. (1990) in: The Ribosome: Structure, Function and Evolution (Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D. and Warner, J.R. eds.) pp. 617–635, ASM Publications, Washington, DC.
[4] Kimura, M., Arndt, E., Hatakeyama, T., Hatakeyama, T. and Kimura, J. (1989) Can. J. Microbiol. 35, 195–199.
[5] Arndt, E., Scholzen, T., Krömer, W., Hatakeyama, T. and Kimura, M. (1991) Biochimie (Paris) 73, 657–668.
[6] Krömer, W.J. and Arndt, E. (1991) J. Biol. Chem. 266, 24573–24579.
[7] Scholzen, T. and Arndt, E. (1992) J. Biol. Chem. 267, 12123–12130.
[8] Bergmann, U. and Arndt, E. (1990) Biochim. Biophys. Acta 1050, 56–60.
[9] Arndt, E., Krömer, W. and Hatakeyama, T. (1990) J. Biol. Chem. 265, 3034–3039.
[10] Kimura, J., Arndt, E. and Kimura, M. (1987) FEBS Lett. 224, 65–70.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[13] Orcutt, B.C., Dayhoff, M.O. and Barker, W.C. (1982) NBR Report 820501-08710 National Biochemical Research Foundation, Washington, DC.
[14] Arndt, E., Breithaupt, G. and Kimura, M. (1986) FEBS Lett. 194, 227–234.
[15] Shoham, M., Dijk, J., Reinhardt, R. and Wittmann-Liebold, B. (1986) FEBS Lett. 204, 323–330.
[16] Amaldi, F., Beccari, E., Bozzoni, I., Luo, Z.X. and Pierandrei-Amaldi, P. (1982) Gene 17, 311–316.
[17] Özkaynak, E., Finley, D., Solomon, M.J. and Varshavsky, A. (1987) EMBO J. 6, 1429–1439.
[18] Finley, D., Bartel, B. and Varshavsky, A. (1989) Nature 338, 394–401.
[19] Lee, H., Simon, J.A. and Lis, J.T. (1988) Mol. Cell. Biol. 8, 4727–4735.
[20] Özkaynak, E., Finley, D., Solomon, M.J. and Varshavsky, A. (1987) EMBO J. 6, 1429–1439.
[21] Nesin, M., Lupski, J.R., Svec, P. and Godson, G.N. (1987) Gene 51, 149–161.
[22] Auer, J., Lechner, K. and Böck, A. (1989) Can. J. Microbiol. 35, 200–204.
[23] Scholzen, T. and Arndt, E. (1991) Mol. Gen. Genet. 228, 70–80.
[24] Ramirez, C., Louie, K.A. and Matheson, A.T. (1989) FEBS Lett. 250, 416–418.