Volume 246, number 1,2, 13-16 FEB 06932 March 1989

The nucleotide sequence of the genes coding for the S19 and L22 equivalent ribosomal proteins from *Halobacterium halobium*

Alexander S. Mankin

Belozersky Laboratory of Molecular Biology and Bioorganic Chemistry, Moscow State University, 119899 Moscow, USSR

Received 18 January 1989

The primary structure of the two ribosomal protein genes of archaebacterium *Halobacterium halobium* has been determined. The encoded polypeptides are homologous to the *Escherichia coli* ribosomal proteins S19 and L22. The two genes constitute part of an operon whose organization is analogous to that of the 'S10' operon of *E. coli*.

Archaebacteria; Ribosomal protein; Evolution; Gene structure; Halobacteria

1. INTRODUCTION

Archaebacteria are recognized as a third evolutionary kingdom besides eubacteria and eukaryotes [1]. Therefore, speculations on the origin and evolution of the genetic apparatus could hardly be treated as valuable unless the knowledge of the structure and expression of the archaebacterial genome reaches at least the level of our understanding of eubacterial and eukaryotic genome strategies.

So far, nucleotide sequencing has been the most straightforward approach to archaebacterial gene studies. In the halophilic branch of archaebacteria, the primary structures of a number of stable RNA genes from different species are now available [2-5]. However, only a few protein genes from halobacteria have been sequenced [6-9]. Thus, the investigation of other halobacterial sequences could illuminate the peculiarities of gene expression. A comparison of halobacterial proteins with their analogs from the non-halophilic organisms might also contribute to an understanding of the basic principles of halotolerance.

In the course of our study of the halobacterial genome structure and expression, we have cloned a cluster of ribosomal protein (r-protein) genes. The complete primary structure of the two genes from this cluster is reported in this paper.

Correspondence address: A.S. Mankin, Belozersky Laboratory of Molecular Biology and Bioorganic Chemistry, Moscow State University, 119899 Moscow, USSR

2. METHODS

The cloning of the *Halobacterium halobium* r-protein gene cluster inserted as a 2.2 kb *Pst*I-*Pst*I fragment into the pUC19 vector has been described previously [10]. A partial sequence of the insert was reported in our recent paper [11].

To determine the complete primary structure of the insert DNA, the *Hin*dIII-*Nco*I and *Nco*I-*Eco*RI subfragments (see fig.1) were isolated from the gel and treated with Klenow DNA polymerase in the presence of four deoxyribonucleoside triphosphates. The resulting fragments with blunt ends were ligated into the M13mp10 vector linearized with *Sma*I. The single-stranded DNA from the recombinant clones was isolated and sequenced by a conventional technique [12]. The *Nco*I site was overlapped by dideoxy sequencing of the initial recombi-

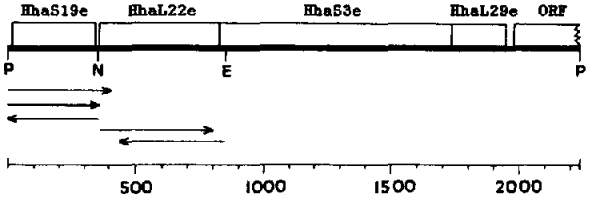

Fig.1. Physical map of the cloned r-protein gene cluster and strategy of sequencing of the HhaS19e and HhaL22e r-protein genes. Restriction sites shown correspond to *Pst*I (P); *Eco*RI (R); *Nco*I (N). The nucleotide scale is shown below.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

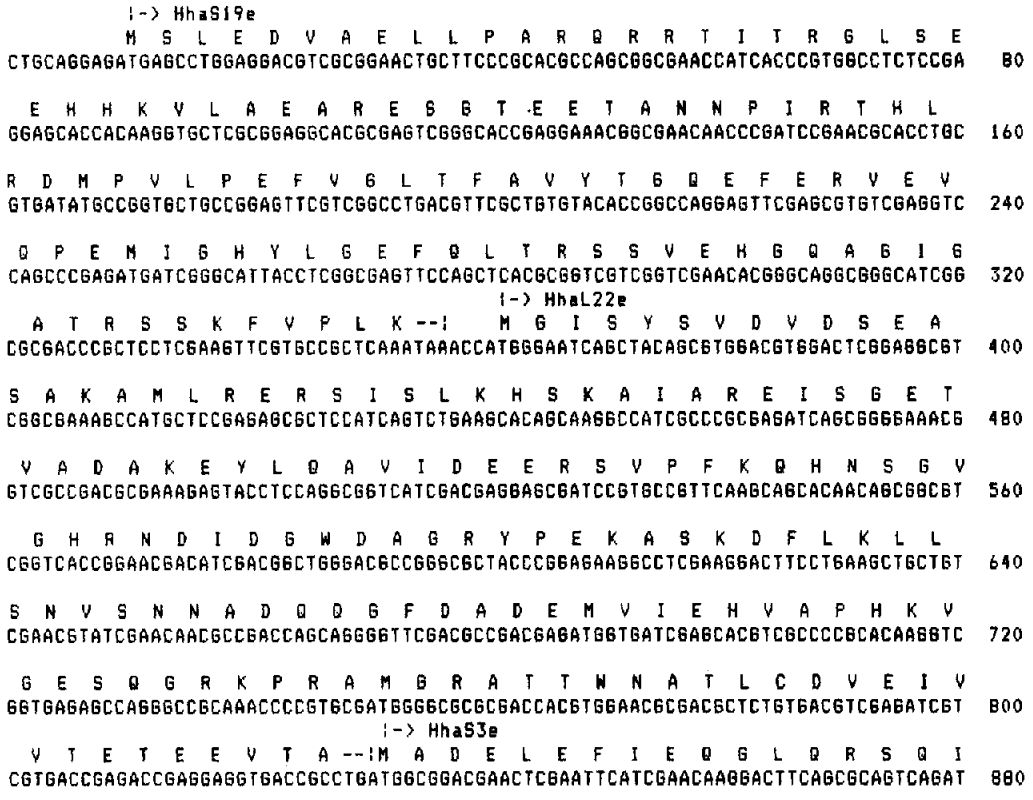

Fig.2. Nucleotide sequence of the HhaS19e and HhaL22e ribosomal protein genes of *H. halobium*. Encoded amino acids are shown above the central nucleotide of the corresponding codons.

nant plasmid from the reverse sequencing primer according to the described procedure [13].

3. RESULTS

The cloning of the 2.2 kb *Pst*I-*Pst*I fragment of the *H. halobium* DNA comprising several r-protein genes was described by Spiridonova et al. [10]. Recently, we reported complete sequences of the genes coding the HhaS3e and HhaL29e proteins (representing *Halobacterium halobium* r-proteins equivalent to the corresponding *E. coli* ones), as well as the 3′-terminal sequence of the HhaL22e gene [11].

The sequencing strategy for the rest of the insert is shown in fig.1 and the deduced primary structure is given in fig.2. Two open reading frames (ORFs), 345 and 468 bp long, were found within this region. The protein encoded in the first ORF has a molecular mass of 12.9 kDa and exhibits significant (25%) homology to the eubacterial r-protein S19 (fig.3A). Therefore, in accordance with the accepted nomenclature, it will be referred to as HhaS19e. The second ORF codes for a 17.1 kDa polypeptide. The previous analysis of the C-terminal sequence of this protein [11] revealed its homology to the eubacterial r-protein L22. A comparison of its complete sequence with the *E. coli* r-protein L22 confirms this observation (fig.3B).

4. DISCUSSION

This paper completes the sequence analysis of the *H. halobium* 2.2 kb DNA fragment that includes the structural genes for HhaS19e, HhaL22e, HhaS3e and HhaL29e ribosomal proteins, as well as an unassigned ORF (see fig.1).

4.1. *Operon structure*

As follows from the S1 nuclease mapping data

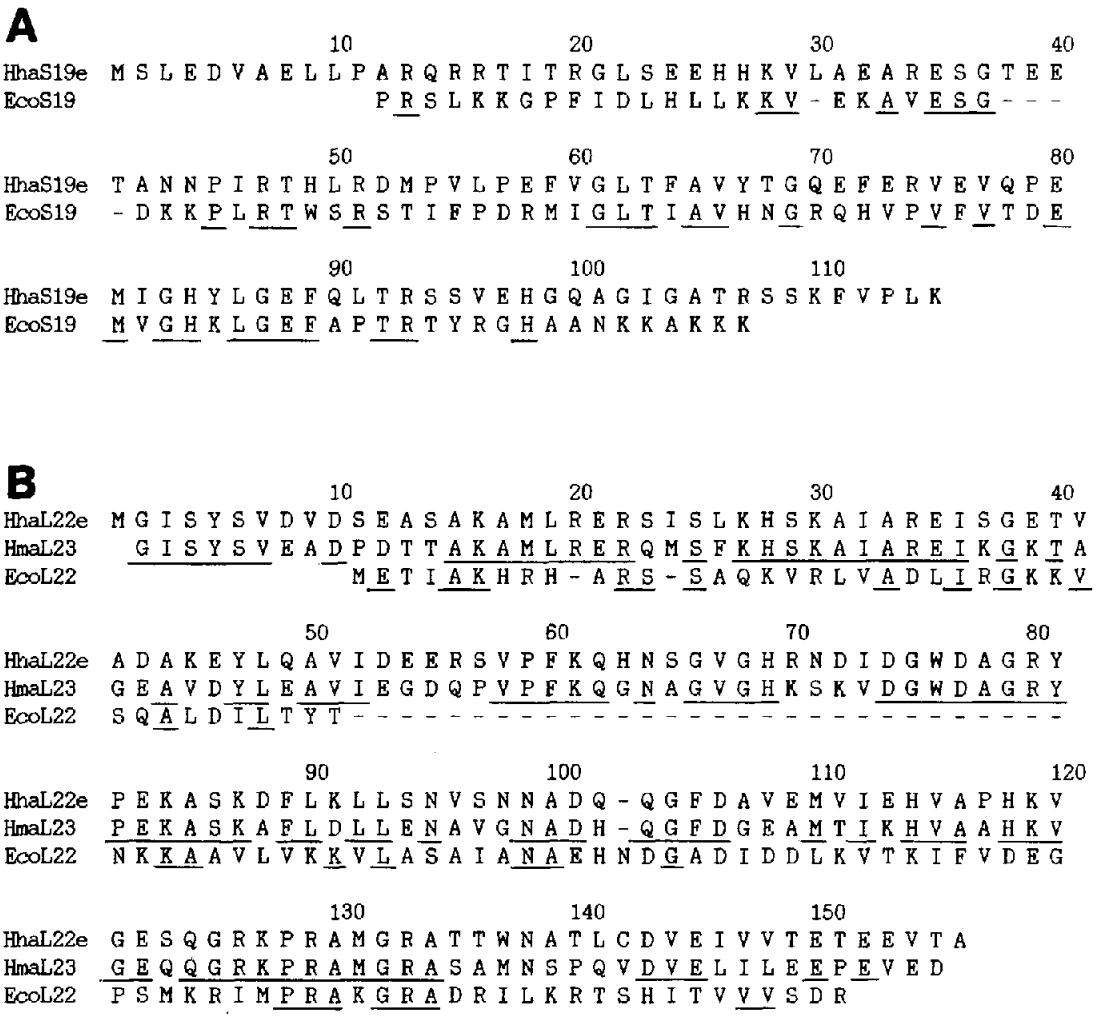

Fig.3. Alignment of the HhaS19e sequence with the *E. coli* S19 r-protein (A) and of the HhaL22e sequence with the primary structures of corresponding proteins from the archaebacterium *H. marismortui* and the eubacterium *E. coli* (B). Initial alignment of the *H. marismortui* and *E. coli* sequences is taken from [17].

[11], all four r-protein genes are cotranscribed as a part of an operon, with the promoter located beyond the 5′ border of the cloned region, while the incomplete downstream ORF seems to be transcribed from its own promoter.

The order of the genes in this r-protein transcription unit closely resembles that in the so-called 'S10' operon of *E. coli*, in which genes are linked in the order S10-L3-L4-L23-L2-S19-L22-S3-L16-L29-S17 [14]. An analogous connection of the L22, S3 and L29 r-protein genes was also found in the methanogenic archaebactrium *Methanococcus vannielii* [15]. The conservation of the r-protein operon structure in archaebacteria and eubacteria deserves interest, assuming a deep phylogenetic branching of the two kingdoms. Moreover, it is surprising, considering the high frequency of genetic rearrangements in halobacteria [16]. Thus, it seems that r-protein operon organization is influenced by a strong selective pressure preventing the rearrangement of the constituent cistrons.

4.2. *Primary structure comparison*

A high degree of homology (60%) exists between r-protein HhaL22e and an r-protein from *H. marismortui* designated as L23 [17]. The homology

of these proteins with the *E. coli* L22, though significant, is much less pronounced, which confirms the concept of the third evolutionary kingdom. In the case of HhaS19e, no analogous archaebacterial sequence has been elucidated yet, whereas sequences of several eubacterial and chloroplast S19-like proteins have been obtained from the NBRF and r-protein data banks.

It should be mentioned that no homology was found between either of the two sequenced proteins and any eukaryotic sequence. This might reflect either a low degree of similarity, or, more plausibly, a mere absence of the corresponding eukaryotic sequences in the data banks.

Though the primary structures of the homologous archaebacterial and eubacterial r-proteins diverge widely, some regions have been well conserved in the course of evolution. These are regions around positions 65 and 95 in the case of HhaS19e and around positions 100 and 130 in HhaL22e. Corresponding sequences might be crucial for the main functions of these r-proteins.

4.3. *Translation initiation signals*

Previously, it was shown that a very well defined Shine-Dalgarno sequence with a 7 and 9 nucleotide complementarity to the 16 S 3′ end [18] was located upstream from the respective ATG codons of HhaS3e and HhaL29e genes [11]. However, no Shine-Dalgarno sequence could be found upstream from the HhaL22e gene. In the case of the HhaS19e gene, only 9 bp are present within the cloned region upstream from the potential initiator ATG. Thus, it is impossible to make any conclusions about the existence of a Shine-Dalgarno sequence for this gene.

HhaL22e, HhaS3e and HhaL29e genes are present within the same operon and are expected to be expressed in equimolar amounts, and so the drastic difference between ribosome binding sites of these genes is amazing indeed. As proposed by P. Dennis, the Shine-Dalgarno mechanism is often not implied in the initiation of translation of the 5′-proximal genes in halobacterial protein operons. The absence of the corresponding sequence in the HhaL22e gene, an internal gene in this transcription unit, suggests that some 'alternative' translation initiation mechanisms could work in this case as well. However, the details of translation initiation in halobacteria are still to be elucidated.

Acknowledgements: I am grateful to Dr Kagramanova for fruitful discussions and to Dr Tanaka for sending data from the r-protein data base. The help of Dr Spiridonova and Dr Akhmanova is greatly appreciated. This work was supported by the constant interest of Dr Baratova and Professor Bogdanov.

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] Mankin, A.S. and Kagramanova, V.K. (1986) Mol. Gen. Genet. 202, 152–161.
[4] Hui, I. and Dennis, P.P. (1985) J. Biol. Chem. 260, 899–906.
[5] Leffers, H., Kjems, J., Ostergaard, N.L. and Garrett, R.A. (1987) J. Mol. Biol. 195, 43–61.
[6] Dunn, R., McCoy, J., Simsek, M., Majumdar, A., Chang, S.H., RajBhandary, U.L. and Khorana, H.G. (1981) Proc. Natl. Acad. Sci. USA 88, 6744–6748.
[7] Blanck, A. and Oesterhelt, D. (1987) EMBO J. 6, 265–273.
[8] Itoh, T., Kumazaki, T., Sugiyama, M. and Otaka, E. (1988) Biochim. Biophys. Acta 949, 110–118.
[9] Betlach, M., Friedman, J., Boyer, H.W. and Pfeifer, F. (1984) Nucleic Acids Res. 12, 7949–7959.
[10] Spiridonova, V.A., Kagramanova, V.K., Baratova, L.A., Golova, Yu., Chekulaeva, L.N. and Mankin, A.S. (1988) Bioorg. Khim. 14, 116–118.
[11] Spiridonova, V.A., Akhmanova, A.S., Kagramanova, V.K., Köpke, A.K.E. and Mankin, A.S. (1989) Can. J. Microbiol. 35, in press.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Hattori, M. and Sakaki, Y. (1986) Anal. Biochem. 152, 232–238.
[14] Lindahl, L. and Zengel, J.M. (1986) Annu. Rev. Genet. 20, 297–326.
[15] Auer, J., Lechner, K. and Böck, A. (1989) Can. J. Microbiol. 35, in press.
[16] Sapienza, C. and Doolittle, W.F. (1982) Nature 295, 384–389.
[17] Hatakeyama, T., Hatakeyama, T. and Kimura, M. (1988) FEBS Lett. 240, 21–28.
[18] Kagramanova, V.K., Mankin, A.S., Baratova, L.A. and Bogdanov, A.A. (1982) FEBS Lett. 144, 177–180.