Volume 12 Number 16 1984 Nucleic Acids Research

Putative promoter region of rRNA operon from archaebacterium *Halobacterium halobium*

A.S.Mankin, N.L.Teterina, P.M.Rubtsov*, L.A.Baratova and V.K.Kagramanova

Belozersky Laboratory of Molecular Biology and Bioorganic Chemistry, Moscow State University, 119899 Moscow, and *Institute of Molecular Biology, USSR Academy of Sciences, 117984 Moscow, USSR

Received 5 June 1984; Accepted 20 July 1984

ABSTRACT

The 100 bp sequence from the begining of the 16S rRNA gene of archaebacterium Halobacterium halobium and the adjacent 800 bp upstream sequence were determined. Four long (80 bp) direct repeats were found in the region preceeding the structural gene of the 16S rRNA. These repeats are proposed to constitute the promoter region of the rRNA operon of H.halobium.

INTRODUCTION

Archaebacteria, the recently discovered third primary line of descent, are attracting great interest. This is not surprising because an elucidation of the workings of the archaebacterial cellular machinery (replication, transcription, translation) and a comparison with similar processes in the two other primary evolutionary kingdoms - eubacteria and eukaryotes - may prove useful for an understanding of the most general principles of cellular organization.

Some data on the structure and function of the translational apparatus in archaebacteria is available (1-3,see 4 for review), whereas there is scarce information concerning the transcription process in this group of organisms. Schnabel and coworkers (5) have investigated the subunit composition of RNA polymerases from different archaebacteria, and found it to be more similar to eukaryotic RNA polymerases rather than to eubacterial ones. So far nothing is known about the transcription initiation signals in the archaebacterial DNA. The transcription initiation point has been determined only for a single archaebacterial gene of bacterio-opsin from H.halobium (6). However, the upstream DNA sequence does not contain any defined elements similar to those

involved in the transcription initiation by the eubacterial and eukaryotic RNA polymerases.

In order to investigate the primary structure of the putative promoter region of other archaebacterial genes, we have sequenced the 800 bp region adjacent to the 5'-end of the 16S rRNA gene in the archaebacterium H.halobium. This DNA segment should presumably contain the external transcribed spacer and the transcription initiation site of the single rRNA operon of this organism. A rather long sequence (80 bp) was found to be repeated four times in the putative promoter region of the H.halobium rRNA operon and it could be a candidate for the archaebacterial promoter.

MATERIALS AND METHODS

Restriction endonucleases BamHI, Sau3A and TagI were purchased from "Biopreparat" (Vilnius,USSR). Restriction endonuclease KpnI was kindly provided by Dr.L.Gening; restriction endonuclease AccI, Klenow Fragment of E.coli DNA polymerase I and phage M13 mp8 (replicative form) were gifts of Dr.A.Metspalu. Dideoxy NTPs were from "P-L Biochemicals,Inc." The 17-mer universal primer for M13 dideoxy sequencing was synthesized and kindly provided by Dr.B.Chernov.

Restriction nuclease digestion, isolation of DNA fragments from agarose gel and cloning procedures were performed exactly as described in (7).

Isolation and characterization of the recombinant plasmid pHT 6, containing the BamHI - HindIII fragment of H.halobium rDNA cloned in the pBR 322 plasmid, will be described elsewhere (manuscript in preparation).

The sequencing was performed by dideoxy chain terminating primer extention technique according to Sanger et al.(8). The BamHI - KpnI fragment was excised from the insert of the recombinant plasmid pHT 6 and purified by electrophoresis in a 1% agarose gel. The total digests of this fragment with restriction endonucleases Sau3A or TagI were cloned into the BamHI or AccI site, respectively, of the phage M13 mp8 replicative form. The sequencing procedure was carried out according to the "BRL M13 Cloning/'Dideoxy' Sequencing Instruction Manual".

RESULTS

The recombinant plasmid pHT 6 was selected from the library of H.halobium DNA digested with BamHI and HindIII restriction endonucleases and cloned in pBR 322 vector (detailes on the cloning and selection will be published elsewhere). Its schematic structure is shown in fig.1. This plasmid carries a 4.5 kbp insert of H.halobium DNA that includes the 5'-half of the 16S rRNA gene.

The KpnI - BamHI fragment (about 1.6 kbp) derived from the insert of this plasmid (see fig.1) was isolated, totally digested with Sau3A or TaqI restriction endonucleases and recloned into phage M13 mp8. The nucleotide sequence of the 800 bp region adjacent to the 5'-end of the 16S rRNA gene and of the 100 nucleotide portion of this gene was determined by the dideoxy technique (8). The sequencing strategy is shown in fig.1 and the deduced primary structure of the region is given in fig.2.

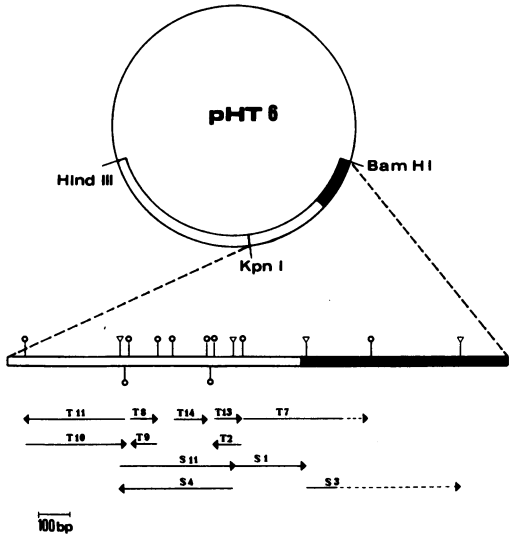

FIGURE 1. The physical map of pHT 6 recombinant plasmid and strategy of sequencing the KpnI - BamHI fragment. Single line indicates pBR 322 sequences; open box represents H.halobium sequences; black box corresponds to the 16S rRNA structural gene. Sau3A (▽) and TaqI (o) restriction sites are marked. M13 mp8 recombinant clones used for dideoxy sequencing are designated by 'T' for TaqI fragments and by 'S' for Sau3A fragments.

```
                                                        TCG ACGGTGTTTT

-780  ATGTGCCCCA CCACTCGGAT GAGATGCGAA CGACGTGAGG TGGCTCGGTG CACCCGACGC

-720  CACTGATTGA CGCCCCCTCG TCCCGTTCGG ACGGAACCCG ACTGGGTTCA GTCCGATGCC

-660  CTTAAGTACA ACAGGGTACT TCGGTGGAAT GCGAACGACA ATGGGGCCGC CCGGTTACAC

-600  GGGTGGCCGA CGCATGACTC CGCTGATCGG TTCGGCGTTC GGCCGAACTC GATTCGATGC

-540  CCTTAAGTAA TAACGGGTGT TCCGATGAGA TGCGAACGAC AATGAGGCTA TCCGGTTCGT

-480  CCGGGTGGCT GATGCATCTC TTCGACGCTC TCCATGGTGT CGGTCTCACT CTCAGTGAGT

-420  GTGATTCGAT GCCCTTAAGT AATAACGGGC GTTACGAGGA ATTGCGAACG ACAATGTGGC

-360  TACCTGGTTC TCCCAGGTGG TTAACGCGTG TTCCTCGCCG CCCTGGTGGG CAAACGTCAC

-300  GCTCGATTCG AGCGTGATTC GATGCCCTTA AGTAATAACG GGGCGTTCGG GGAAATGCGA

-240  ACGTCGTCTT GGACTGATCG GAGTCCGATG GGTTTATGAC CTGTCGAACT CTACGGTCTG

-180  GTCCGAAGGA ATGAGGATTC CACACCTGCG GTCCGCCGTA AAGATGGAAT CTGATGTTAG

-120  CCTTGATGGT TTGGTGACAT CCAACTGGCC ACGACGATAC GTCGTGTGCT AAGGGACACA

- 60  TTACGTGTCC CCGCCAAACC AAGACTTGAT AGTCTTGGTC GCTGGGAACC ATCCCAGCAA

     pAUUCCGGUUG AUCCUGCCGG AGGUCAUUGC UAUUGGGGUC CGAUUUAGCC AUGCUAGUUG
+  1  ATTCCGGTTG ATCCTGCCGG AGGTCATTGC TATCGGAGTC CGATTTAGCC ATGCTAGTTG

     CACGAGUUCA UACUCGUGGC GAAAAGCUCA GUAACACGUG
+ 60  TGCGGGTTTA GACCCGCAGC GGAAAGCTCA GTAACACGTG
```

FIGURE 2. The primary structure of the 900 bp segment from the KpnI - BamHI region of the pHT 6 plasmid (the non-coding DNA strand is shown). The nucleotide sequence of the 5'-portion of 16S rRNA from related archaebacterium Halobacterium volcanii (1) is given in italics above homologous DNA region. The direct repeats in the putative promoter region are underlined. The sequence corresponding to the 5'-terminal T_1-oligonucleotide of H.halobium 16S rRNA (11) is boxed.

DISCUSSION

Hofman et al.(9) and Neumann et al.(10) have shown that there is only one copy of 5S, 16S and 23S rRNA structural genes in the genome of the archaebacterium H.halobium linked similarly to all eubacterial rDNAs investigated so far, i.e. 5'-16S-23S-5S-3'.

We have cloned the HindIII - BamHI 4.5 kbp fragment of H.halobium DNA that contained the 5'-half of the 16S rRNA gene (see fig.1). The KpnI - BamHI region of this insert includes the 5'-half of the 16S rRNA gene and approximately 900 bp of the upstream sequence. If the H.halobium rRNA operon is indeed organized similarly to the eubacterial one, then the KpnI - BamHI

fragment should contain the external transcribed spacer and the promoter region of the ribosomal operon.

Localization of the 5'-end of the 16S rRNA gene. To localize the 5'-end of the 16S rRNA gene, we drew upon three different facts. Firstly, the M13 recombinant clones S1, T7 and S3 (see fig.1) gave positive signals when hybridized with H.halobium ^{32}P-16S rRNA, while those located "to the left" from the T7 clone did not hybridize. Therefore, the 5'-end of the 16S rRNA gene is localized within the TaqI - Sau3A 200 bp fragment common to clones S1 andT7, before the starting point of the S3 clone. Secondly, Magrum et al.(11) have previously shown the structure of the 5'-terminal T_1-oligonucleotide of H.halobium 16S rRNA to be pAUUCCG. This sequence is present only once in the determined primary structure of the 900 bp fragment of H.halobium chromosome and lies within the clones T7 and S1 (boxed in fig.2). Thirdly, starting from this oligonucleotide, the sequence we have determined is highly homologous to the 5'-portion of archaebacterium 16S rRNA from Halobacterium volcanii determined by Gupta et al. (1). The combination of these facts proves that H.halobium 16S rRNA gene starts from A +1 in fig.2.

The external transcribed spacer. All known rRNA operons of eubacteria and eukaryotes include an external transcribed spacer - the transcribed segment of the operon preceeding the 16S (18S) rRNA gene. The size of this spacer varies from 200 - 300 bp in eubacteria (12-14) to several hundreds bp in eukaryotes (15).

In this work we have not localized experimentally the transcription initiation point of H.halobium ribosomal operon and thus we do not know either the starting point, or the size of the external transcribed spacer of the archaebacterial rRNA operon. However, we suppose that a transcription of the operon initiates upstream to the position -127 (fig.2). This proposition is based on the following observation. Having determined the nucleotide sequence downstream the 3'-end of the 16S rRNA gene (manuscript in preparation), we found a very strong complementarity of the regions flanking the 16S rRNA gene: (-76) - (-127) / (+37) - (+86) (fig.3). A similar complementarity of the external transcribed spacer and of the first internal spacer was initially found by Young and Steitz (16) for

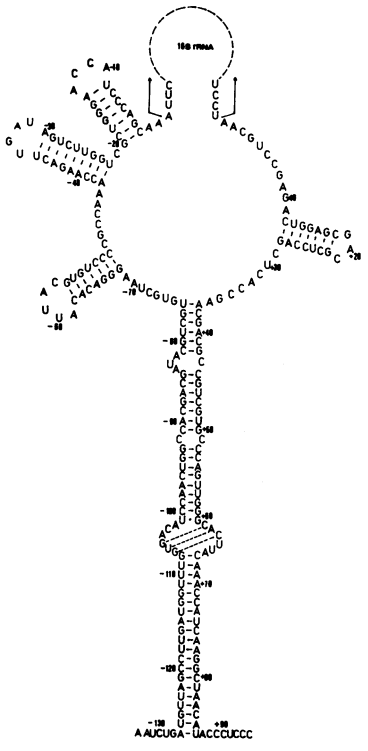

FIGURE 3. Complementarity of H.halobium 16S rRNA flanking regions.

E.coli ribosomal operon and later for other ribosomal operons of eubacterial and eukaryotic origin and was thought to serve as a signal for the processing ribonucleases.

The complementarity of the regions (-76) - (-127) and (+37)-(+86) seems too extensive to be of occasional origin and we believe that at least 127 nucleotides preceeding the gene of the 16S rRNA in H.halobium ribosomal operon are included in the external transcribed spacer.

A putative promoter region. A computer investigation of the sequence has revealed the presence of four rather long direct repeats (fig.4) separated from each other by about 40 nucleotides; 200 nucleotides separate the first repeat from the begining of the 16S rRNA gene. Two regions of the repeated sequence are perfectly homologous (boxed in fig.4), while the other parts of the repeats share a significant though not perfect homology. It is interesting that upstream to the fourth repeat one can find a stretch of

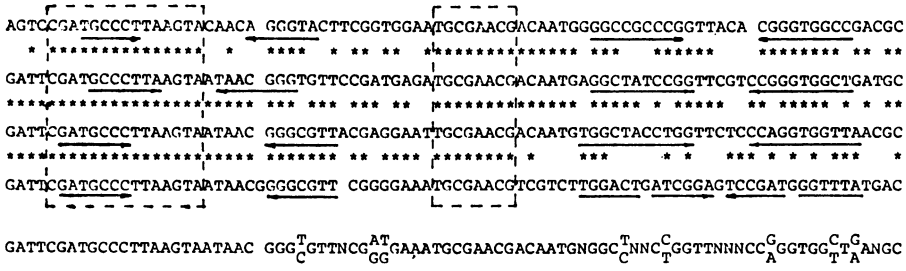

FIGURE 4. Four direct repeats from the promoter region of H.halobium rRNA operon. Identical nucleotides in the neibouring repeats are indicated by asterisks. The sequences of perfect homology are boxed. Inverted sequences in the direct repeats are shown by arrows. The stretch below represents the consensus sequence.

nucleotides (-747) - (-757) very similar to the second region of perfect homology (see fig.2). Some homology could also be found immediately downstream, though upstream no homology with the repeated sequence was detected.

We propose that these repeated sequences correspond to the promoters (or some parts of promoters) of the rRNA operon in archaebacterium H.halobium. This suggestion is confirmed by several observations.

i) Recently Huet et al. (17) have found that antibodies raised against eukaryotic RNA polymerases I (A) and II (B) could also interact with RNA polymerases from different archaebacteria and (to a much less extent) with eubacterial (E.coli) RNA polymerase. These authors have proposed the common evolutionary origin of RNA polymerases from the three primary kingdoms. The homology of RNA polymerases in turn means the homology of transcription initiation signals encoded in the DNA of organisms from different lines of descent. Indeed one can find several "homologous" features in the promoter regions recognized by eubacterial and eukaryotic RNA polymerases (RNA polymerase II in the later case). These are: an AT-rich consensus sequence separated by 5-10 (eubacteria) or 20-25 (eukaryotes) nucleotides from the starting point of transcription; an existence of the second consensus sequence located at (-35) (eubacteria) or (-75)-(-80) (eukaryotes) position; inverted repeats are present both in eubacterial and eukaryotic promoters; in eukaryotes and eubacteria transcription starts

```
-717  TGATTGACGC
      **   *****
-597  TGGCCGACGC
      *    ******
-462  TCTTCGACGC
      ** *** ***
-329  TCCTCGCCGC

      T(C/G)NTCGACGC
```

FIGURE 5. Homology of the sequences located about 30 bp upstream from the main direct repeats shown in fig.4. The consensus sequence is given below.

more often from A or G rather than from C or T (18,19). Taking into account the homology of RNA polymerases, one may expect to find similar features in the archaebacterial promoters. The direct repeats we have found in the putative promoter region of the H.halobium rRNA operon do reveal some of these features. Thus one can find the AT-rich region - TAATAA in the case of the first three repeats - which is much similar to the TATAAT Pribnow box of eubacteria or the TATA box of eukaryotes. It should be noted that this is the only region containing six consecutive A-T pairs from (-1) up to (-780) position in the sequence we determined. Yet it is somewhat more difficult to find the second consensus sequence because it could be located outside the borders of the repeat. Indeed, a close investigation of the interrepeat spacers revealed the presence of rather homologous stretches of nucleotides (shown in fig.5) separated by approximately 50 bp from the TAATAA consensus sequence. This homology may indicate the presence of the second consensus sequence. Such "promoter feature" as the presence of inverted sequences is also characteristic of the repeats in the putative promoter region of H.halobium rRNA operon (see fig.4).

ii) To our knowledge, the bacterio-opsin gene from H.halobium is the only archaebacterial gene sequenced so far for which the transcription starting site was determined (6). Comparison of the nucleotide sequence near the transcription initiation site of the bacterio-opsin gene with the putative promoter region of H.halobium rRNA operon does not reveal any extensive homology. This is not surprising because in the case of eukaryotic or eubacterial promoters the similarity is found only for two short segments located rather far from each other (see previous section). Some versions of alignment of bacterio-opsin and rRNA operon promoter sequences are possible.

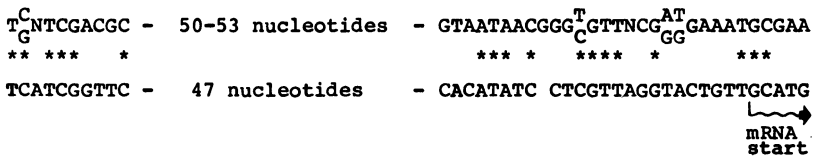

FIGURE 6. Alignment of the consensus sequence of the putative H.halobium rRNA operon promoter (upper stretch, this work) with the transcription initiation region of H.halobium bacterio-opsin gene (lower stretch,(6)). Identical nucleotides are indicated by asterisks.

One version is shown in fig.6. It is remarkable for the homology of the sequences just surrounding the transcription start, in the AT-rich region and in the region of the second consensus sequence. If this version of alignment is indeed of functional significance, then the second inverted repeat (fig.4) could reflect the existence of a hairpin not far from the begining of the transcripts. It is noteworthy that some homology could be found between these hairpins and the transcription termination sequence for the bacterio-opsin gene. It is possible that these hairpins could reflect the presence of attenuation signals in the H.halobium rRNA operon.

iii) Ribosomal operons belong to the most extensively expressed operons in the living cell. In many cases it was shown that promoters of rRNA operons are multiplicated (13,14,20-22). Multiplication of promoters could serve to increase the level of rRNA genes expression. The archaebacterium H.halobium is remarkable for having a single copy of rRNA operon (in contrast, the majority of eubacteria and eukaryotes have several copies of ribosomal operons); that is why there should exist a mechanism to compensate such a low dosage of rRNA genes in this organism. The four tandemly repeated promoters could be a part of an apparatus designed for an increasing of the rRNA operon transcription level.

The lack of experimental results leaves no ground for an unequivocal conclusion whether the repeats found in the putative promoter region of H.halobium rRNA operon indeed correspond to the multiple promoters and if they do, whether all the promoters are functionally active in the initiation of transcription. We

hope that the development of the homologous cell-free system for the in vitro transcription would help to resolve these questions.

ACKNOWLEDGEMENTS

We thank Professor A.Bogdanov for fruitful discussions and careful reading of the manuscript, Dr. K.Skryabin for support of this work, Drs. V.Mankin and A.Mironov for help in computing the results and Ms. O.Primak for technical assistance.

REFERENCES

1. Gupta,R., Lanter,J.M. and Woese,C.R. (1983) Science 221, 656-659.
2. Kagramanova,V.K., Mankin,A.S., Baratova,L.A. and Bogdanov,A.A. (1982) FEBS lett. 144, 177-180.
3. Mankin,A.S., Kagramanova,V.K., Belova,E.N., Teterina,N.L., Baratova,L.A. and Bogdanov,A.A. (1982) Biochem.Int. 5, 719-726.
4. Matheson,A.T. and Yaguchi,M. (1981) in International Cell Biology 1980-1981, Schweiger,H.G. Ed., pp.103-110, Springer-Verlag, Berlin.
5. Schnabel,R., Thomm,M., Gerardy-Schaln,R., Zillig,W., Stetter, R.O.and Huet,J. (1983) EMBO J. 2, 751-755.
6. Das Sarma,S., RajBhandary,U.L. and Khorana,H.G. (1984) Proc. Natl.Acad.Sci. USA 81, 125-129.
7. Maniatis,T., Fritch,E.F. and Sambrook,J. (1982) Molecular Cloning, Cold Spring Harbor Lab.
8. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc.Natl.Acad. Sci. USA 74, 54-63.
9. Hofman,J.D., Lau,R.U. and Doolottle,W.F. (1979) Nucl.Acids Res. 7, 1321-1333.
10. Neumann,H., Gierl,A., Tu,J., Leibrock,J., Staiger,D. and Zillig,W. (1983) Mol.Gen.Genet. 192, 66-72.
11. Magrum,L.J., Luehrsen,K.R. and Woese,C.R. (1978) J.Mol.Evol. 11, 1-8.
12. Young,R.A. and Steitz,J.A. (1979) Cell 17, 225-234.
13. Gilbert,S.F., deBoer,H.A. and Nomura,M. (1979) Cell 17, 211-224.
14. Stewart,G.C. and Bott,K.F. (1983) Nucl.Acids Res. 11,6289-6300.
15. Long,E.O. and Dawid,I.B. (1980) Ann.Rev.Biochem. 49, 727-764.
16. Young,R.A. and Steitz,J.A. (1978) Proc.Natl.Acad.Sci. USA 75, 3593-3597.
17. Huet,J., Schnabel,R., Seutenac,A. and Zillig,W. (1983) EMBO J. 2, 1291-1294.
18. Artemiev,I.V., Vassiliev,G.V. and Gurevitch,A.I. (1983) Bioorg.Khim.(USSR) 9, 1544-1557.
19. Breathnach,R. and Chambon,P. (1981) Ann.Rev.Biochem. 50, 349-383.
20. Glaser,G. and Cashel,M. (1979) Cell 16, 111-121.
21. Seiki,M., Ogasawara,N. and Yoshikawa,H. (1982) Proc.Natl. Acad.Sci. USA 79, 4285-4289.