Volume 216, number 2, 217–220 FEB 04745 June 1987

Simple repetitive sequences in the genomes of archaebacteria

R.P. Vashakidze and D.A. Prangishvili

Institute of Molecular Biology and Biological Physics, Georgian SSR Academy of Sciences, Gotua str. 14, Tbilisi 380060, USSR

Received 19 March 1987

Stretches of simple sequences poly(dG-dT)·poly(dC-dA), poly(dG-dA)·poly(dC-dT), poly(dG)·poly(dC) and poly(dA)·poly(dT), the occurrence of which is a characteristic feature of eukaryotic genomes, are found in the genomes of archaebacteria *Halobacterium halobium* and *Sulfolobus acidocaldarius*. In *S. acidocaldarius* these sequences constitute a considerable portion of the genome; they belong to a class of repetitive sequences dispersed throughout the genome, being transcribed and found in RNAs of different lengths.

Nucleotide sequence; Genome sequence; Archaebacteria

1. INTRODUCTION

Short stretches of simple sequences consisting of one or two tandemly repeated nucleotides, for example poly(dA)·poly(dT), poly(dG-dT)·poly(dC-dA) [Further we write homopolymeric sequences as two nucleotides from both strands, for example AA/TT instead of poly(dA)·poly(dT). For higher order simple sequences we write the repeat unit from both strands, for example GT/CA instead of poly(dG-dT)·poly(dC-dA).], constitute ubiquitous repetitive components of eukaryotic genomes [1,2]. They are dispersed throughout the genome [2], found in introns [3,4] and other untranslated regions [5–7] and transcribed into RNA [8]. The functions of these sequences, if any, may be related to chromatin folding [9], recombination [10,11] or regulation of gene activity [12]. Attempts to find simple sequences in genomes of prokaryotes have had negative results [8]. However, these efforts were confined to eubacteria, one of the two major phylogenetic groups of prokaryotes. For an understanding of the origin of simple sequences in genomes it seems useful to look for them in archaebacteria, the second phylogenetic group of prokaryotes, believed to have diverged from eubacterial and eukaryotic lineages at the earliest stages of cellular evolution [13,14]. Here, we show that genomes of archaebacteria contain simple sequences, which have the same characteristics as those of eukaryotic genomes.

Correspondence address: R.P. Vashakidze, Institute of Molecular Biology and Biological Physics, Georgian SSR Academy of Sciences, Gotua str. 14, Tbilisi 380060, USSR

2. MATERIALS AND METHODS

Sulfolobus acidocaldarius strain DSM 639 was obtained from W. Zillig, Martinsried. Cells were grown at 75°C and pH 2.5 as in [15]. *Drosophila melanogaster* line RC was obtained from M. Evgen'ev, Moscow.

DNAs from *S. acidocaldarius* and *D. melanogaster* were purified as in [16]. *Escherichia coli* and *Staphylococcus aureus* DNAs were provided by D. Chinchaladze, Tbilisi, *Halobacterium halobium* DNA by A. Mankin, Moscow, and phage ϕH DNA by H. Schnabel, Martinsried.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

The chloroform-phenol method as described in [17] was used for preparation of total RNA from *S. acidocaldarius*.

The polydeoxyribonucleotides poly(dG-dT)·poly(dC-dA), poly(dG-dC)·poly(dC-dG), poly-(dG-dA)·poly(dC-dT), poly(dT)·poly(dA) and poly(dG)·poly(dC) were purchased from PL Biochemicals, WI.

DNAs and polydeoxyribonucleotides were ^{32}P labelled by nick translation using one ^{32}P-labelled nucleotide.

Cleavage of DNA, agarose gel electrophoresis and transfer to nitrocellulose filters of DNA fragments were performed as described in [18], and agarose gel electrophoresis and transfer to nitrocellulose filters of RNA as in [8].

For construction of a genomic library, *S. acidocaldarius* DNA was digested with *Hin*dIII and ligated to the vector pBR322 which had been cleaved with the same restriction enzyme; the ratio of fragment ends to ends of the vector was 50:1. After transformation about 10% of the clones were recombinant. In total about 2000 recombinant clones were obtained.

An *Eco*RI-*Hin*dIII digest of phage λ DNA was used as a marker for determining the lengths of DNA fragments.

The hybridization conditions for different simple sequence probes were as follows [19]: for GT/CA, GC/CG and GA/CT: 5 × SSC, 50% formamide, 37°C; for AA/TT: 2 × SSC, 5% formamide, 37°C; for GG/CC: 2 × SSC, 50% formamide, 45°C. Filters were finally washed in 2 × SSC, 50% formamide, 0.1% SDS at 37°C.

Hybridization with bacterial clones was performed as in [20], and dot hybridization as in [21].

DNA sequencing was carried out as described in [22].

3. RESULTS AND DISCUSSION

The archaebacteria are divided into two main subgroups, one comprising methanogens, extreme halophiles and thermoacidophiles *Thermoplasma acidophilum* and *Thermococcus celer*, and the other consisting of sulfur-dependent extreme thermophiles [23]. Looking for simple sequences in the genomes of archaebacteria, we examined genomes of representatives of both subgroups, the extreme halophile *H. halobium* and sulfur-dependent thermophile *S. acidocaldarius*. As probes we used the ^{32}P-labelled synthetic polynucleotides GT/CA, GC/CG, GA/CT, AA/TT and GG/CC. They were hybridized to dot blots of DNAs from the two archaebacteria and *Halobacterium* phage φH. The hybridization was also carried out with *D. melanogaster* DNA and DNAs from two eubacteria, Gram-negative *E. coli* and Gram-positive *St. aureus*. The hybridization conditions employed were such that runs of simple sequences shorter than 20 base pairs (bp) should give no signal, because their melting point would be below the hybridization temperature [19]. The results (see fig.1) show that stretches of GG/CC, GA/CT and GT/CA are present in all archaebacterial and eukaryotic species. AA/TT was detected in *D. melanogaster* DNA; from archaebacterial species only in *S. acidocaldarius* DNA. GC/CG is present in eukaryotic DNA but not in archaebacterial DNAs. As expected, neither type of simple sequence was detected in the DNAs of eubacteria.

Simple sequences of *S. acidocaldarius* DNA were studied in detail. A library of cloned restriction fragments of *S. acidocaldarius* DNA was probed by colony hybridization with ^{32}P-labelled

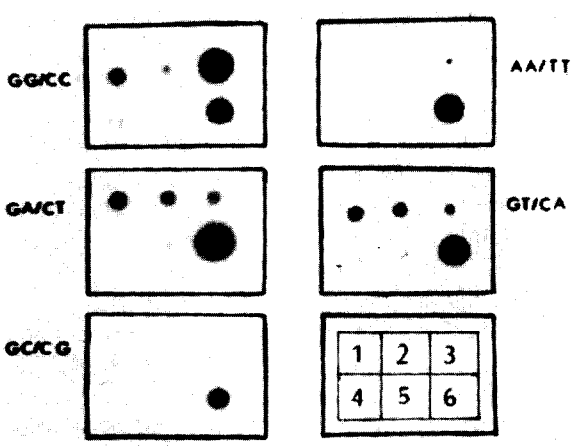

Fig.1. Hybridization of ^{32}P-labelled GG/CC, AA/TT, GA/CT, GT/CA and GC/CG to dot blots of DNAs from archaebacteria, eubacteria and eukaryotes. DNAs from different organisms (2 μg each) were present on filters in the following order: dots (1) *Halobacterium* phage φH; dots (2) *H. halobium*; dots (3) *S. acidocaldarius*; dots (4) *E. coli*; dots (5) *St. aureus*; dots (6) *D. melanogaster*.

polynucleotides. About 1.65% of clones gave a positive signal with GG/CC, 1.2% with GA/CT, 1.15% with GG/CC and 0.15% with AA/TT. These findings suggest that all types of simple sequences belong to a class of repetitive elements of the *S. acidocaldarius* genome.

The distribution of simple sequences in *S. acidocaldarius* DNA was studied by Southern hybridization experiments. ^{32}P-labelled GA/CT, GG/CC, CT/CA and AA/TT were hybridized with fractionated *Bam*HI fragments of DNA. As a result, a smear and occasionally some bands could be observed in each case, except for AA/TT which hybridized with a single fragment 6 kbp in length (see fig.2). Such a picture is consistent with the notion that simple sequences are rather interspersed in the genome.

The nucleotide sequence of *S. acidocaldarius* DNA fragments containing stretches of simple sequences was studied using the example of clone S219 DNA. The clone gave positive signal when hybridized to [^{32}P]AA/TT. Insertion of 3.5 kbp from this clone was digested with the restriction enzyme *Rsa*I and fragments obtained were subcloned in the *Hin*cdII site of pUC19. DNA from only one of these subclones (SpUA3) hybridized with [^{32}P]AA/TT. SpUA3 DNA sequencing was carried out using the dideoxy procedure of Sanger. The sequence of the AA/TT element is presented in fig.3. The length of the element is 18 nucleotides; in position 14 A is substituted for G.

To determine whether simple sequences are transcribed in archaebacteria, ^{32}P-labelled GA/CT, GG/CC, GT/CA and AA/TT were hybridized with blots of fractionated *S. acidocaldarius* total RNA. As seen in fig.4, each type of simple sequence present in the genome is transcribed into RNA. The occurrence of AA/TT in RNAs of different lengths may reflect post-

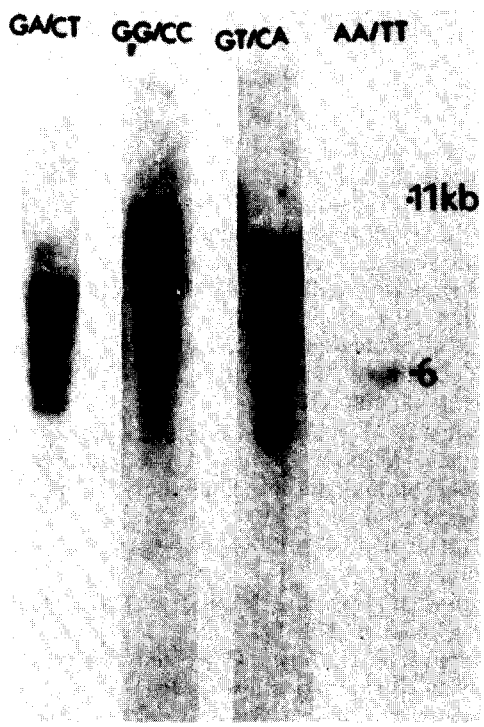

Fig.2. Hybridization of ^{32}P-labelled GA/CT, GG/CC, GT/CA and AA/TT to genomic blots of *S. acidocaldarius*. DNA was digested with *Bam*HI; the resulting fragments were fractionated on 1% agarose gels and transferred to nitrocellulose filters.

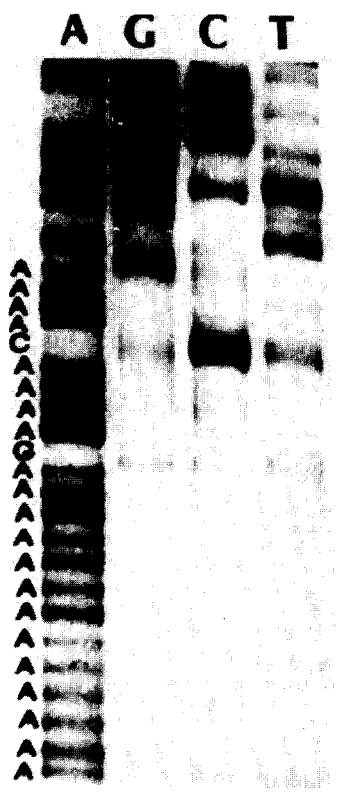

Fig.3. Radioautography of a portion of sequencing gel for clone SpUA3 DNA, containing a stretch of AA/TT.

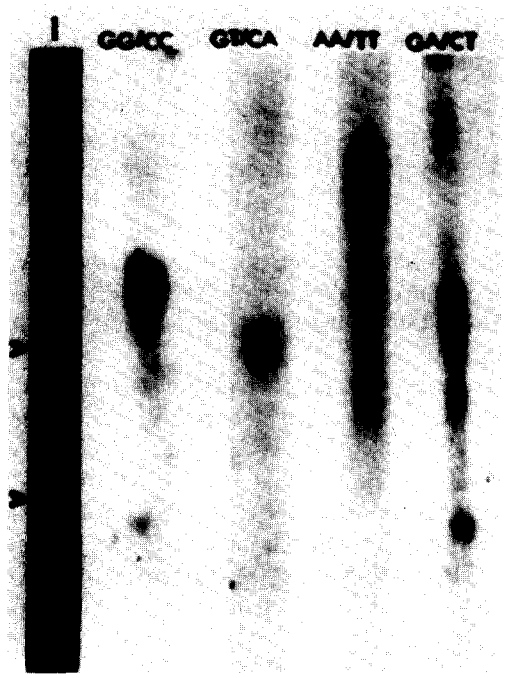

Fig.4. Hybridization of ^{32}P-labelled GG/CC, GT/CA, AA/TT and GA/CT to blots of total RNA from *S. acidocaldarius*. The gel, stained with ethidium bromide, is presented in lane 1.

transcriptional polyadenylation of RNAs [24] rather than transcription of corresponding genomic sequence.

The present data indicate that simple sequences of archaebacterial genomes display the same characteristics as those of eukaryotic genomes. In both cases simple sequences belong to a class of repetitive elements dispersed throughout the genome, and are transcribed and present in RNAs of different lengths.

So far, as simple sequences were revealed only in genomes of eukaryotes, their origin was attributed either to slippage mechanisms or to an unequal cross-over taking place at randomly occurring runs of these sequences [2]. The discovery of simple sequences in genomes of archaebacteria makes other suggestions concerning their origin also possible. One of them might relate the origination of simple sequences to the earliest stages of genome formation.

REFERENCES

[1] Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) Proc. Natl. Acad. Sci. USA 79, 6465–6469.
[2] Tautz, D. and Renz, M. (1984) Nucleic Acids Res. 12, 4127–4138.
[3] Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) Proc. Natl. Acad. Sci. USA 79, 5901–5905.
[4] Kvist, C., Roberts, L. and Dobberstein, B. (1983) EMBO J. 2, 245–254.
[5] Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnsteil, M.L. (1978) Cell 14, 655–671.
[6] Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033–1044.
[7] Hamada, H., Petrino, M.G., Kakunaga, T., Siedman, M. and Stollar, B.D. (1984) Mol. Cell. Biol. 4, 2610–2621.
[8] Tautz, D. and Renz, M. (1984) J. Mol. Biol. 172, 229–235.
[9] Johnson, D. and Morgan, R. (1978) Proc. Natl. Acad. Sci. USA 75, 1637–1641.
[10] Rogers, J. (1983) Nature 305, 101–102.
[11] Stringer, J.R. (1985) Mol. Cell. Biol. 5, 1247–1259.
[12] Hamada, H., Seidman, M., Howard, B.H. and Gorman, C.M. (1984) Mol. Cell. Biol. 4, 2622–2630.
[13] Woese, C.R. and Fox, G.E. (1977) Proc. Natl. Acad. Sci. USA 74, 5088–5090.
[14] Woese, C.R. (1981) Sci. Am. 244, 94–106.
[15] Brock, T.D., Brock, K.M., Beiley, R.T. and Weiss, R.L. (1972) Arch. Microbiol. 84, 54–68.
[16] Kolchinsky, A.M., Vashakidze, R.P. and Mirzabekov, A.D. (1980) Mol. Biol. (USSR) 14, 1098–1107.
[17] Tu, J., Prangishvili, D., Huber, H., Wildgruber, G., Zillig, W. and Stetter, K.O. (1982) J. Mol. Evol. 18, 109–114.
[18] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[19] Tautz, D. (1983) Dissertation, Universität Tübingen.
[20] Grunstein, M. and Wallis, J. (1979) Methods Enzymol. 68, 379–389.
[21] Kafatos, F.C., Jones, C.W. and Efstratiadis, A. (1979) Nucleic Acids Res. 7, 1541–1552.
[22] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[23] Woese, C.R. and Olsen, G.J. (1986) Systems Appl. Microbiol. 7, 161–177.
[24] Oshima, T., Ohba, M. and Wakagi, T. (1984) Orig. Life 14, 665–669.