JOURNAL OF BACTERIOLOGY, June 1996, p. 3232–3237 Vol. 178, No. 11
0021-9193/96/$04.00+0

Sequence of Plasmid pGT5 from the Archaeon *Pyrococcus abyssi*: Evidence for Rolling-Circle Replication in a Hyperthermophile

GAEL ERAUSO,[1] STEPHANIE MARSIN,[2] NADIA BENBOUZID-ROLLET,[2] MARIE-FRANCE BAUCHER,[2] TRISTAN BARBEYRON,[1] YVAN ZIVANOVIC,[2] DANIEL PRIEUR,[1] AND PATRICK FORTERRE[2]*

Station Biologique de Roscoff, UPR 9042 Centre National de la Recherche Scientifique, 29682 Roscoff-Cedex,[1] and Institut de Génétique et Microbiologie, URA 1354 Centre National de la Recherche Scientifique, Université Paris-Sud, 91405 Orsay-Cedex,[2] France

Received 14 November 1995/Accepted 20 March 1996

The plasmid pGT5 (3,444 bp) from the hyperthermophilic archaeon *Pyrococcus abyssi* GE5 has been completely sequenced. Two major open reading frames with a good coding probability are located on the same strand and cover 85% of the total sequence. The larger open reading frame encodes a putative polypeptide which exhibits sequence similarity with Rep proteins of plasmids using the rolling-circle mechanism for replication. Upstream of this open reading frame, we have detected an 11-bp motif identical to the double-stranded origin of several bacterial plasmids that replicate via the rolling-circle mechanism. A putative single-stranded origin exhibits similarities both to bacterial primosome-dependent single-stranded initiation sites and to bacterial primase (*dnaG*) start sites. A single-stranded form of pGT5 corresponding to the plus strand was detected in cells of *P. abyssi*. These data indicate that pGT5 replicates via the rolling-circle mechanism and suggest that members of the domain *Archaea* contain homologs of several bacterial proteins involved in chromosomal DNA replication. Phylogenetic analysis of Rep proteins from rolling-circle replicons suggest that diverse families diverged before the separation of the domains *Archaea*, *Bacteria*, and *Eucarya*.

The isolation and characterization of plasmids are prerequisites for the development of genetic studies on new groups of microorganisms. Plasmids also are essential tools for studying in vivo and in vitro mechanisms such as DNA replication, recombination, and repair (28). Bacterial plasmids have been extensively analyzed and used in molecular and genetic work. In contrast, information about plasmids in the domain *Archaea* (the third domain of life sensu Woese et al. [49]) is much more limited. Several plasmids in extreme halophiles and methanogens have been described, and a few of them were completely sequenced: pHV2 (6.3 kb) from *Haloferax volcanii* (7), pGRB1 (1.8 kb) from *Halobacterium* sp. strain GRB and its relatives pGN101 and pHSB1 (18, 19, 26), and pME2001 (4.4 kb) from *Methanobacterium thermoautotrophicum* Marburg (4). The minimal replication regions for pHH1 (~150 kb) and pNRC100 (~200 kb) from *Halobacterium salinarium* (33, 37) and pHK2 (10.5 kb) from *Haloferax* sp. strain Aa2.2 (24) were defined and sequenced. The plasmid pGRB1 and its relatives, as well as pHK2, encode homologous proteins that exhibit similarities with Rep proteins from rolling-circle (RC) replicons of the ϕX174 group (24, 25). In contrast, pNRC100 and pHV2 encode homologous Rep proteins (encoded by *repH*) unrelated to those of the RC replicons (33). Some of the haloarchaeal plasmids have been used successfully for the construction of shuttle vectors (23, 29, 30).

The number of plasmids described for thermophilic archaea is restricted compared with those described for halophiles and methanogens. The first extrachromosomal element detected in a member of the order *Sulfolobales*, pSB12 (15.5 kb) (50), was in fact the genome of a lysogenic virus, *Sulfolobus shibatae* virus 1 (SSV1) (for a review, see reference 38). Later, a bona fide plasmid (pDL 10) (7.7 kb) in *Desulfurolobus ambivalens* was described (52), and more recently, a variety of plasmids with sizes ranging from 4.7 to 45 kb have been isolated from *Sulfolobus* strains newly discovered in Iceland and Japan (40, 51).

For a long time, no plasmid from hyperthermophiles (optimal temperature, >90°C) was described. This was troublesome, since these organisms have attracted much interest both for their evolutionary significance and for their biotechnological potential (16, 21, 43). In recent years, *Pyrococcus furiosus* and other members of the order *Thermococcales* have been chosen by several laboratories as model organisms for the study of hyperthermophiles (2). These strains have a short generation time (~30 min) and can be plated on solid media containing Gelrite (13). We have recently reported the isolation of a small plasmid (3.4 kb), pGT5, from a new hyperthermophilic archaeon isolated from a deep-sea vent (12). This organism has now been described as the type strain of a new species, *Pyrococcus abyssi* GE5 (14). pGT5 is a stable multicopy plasmid (25 to 30 copies per chromosome) and has been isolated from cultures at different growth phases and in the temperature range from 75 up to 105°C (11). pGT5 could thus become an important tool to design cloning vectors for *Pyrococcus* spp. and to study DNA replication, repair, and recombination in hyperthermophilic archaea. We report here the complete nucleotide sequence of pGT5. Features of this sequence and the detection of specific single-stranded (ss) pGT5 DNA strongly suggest that this plasmid replicates via an RC mechanism and that *Archaea* contain homologs of some essential bacterial replication proteins.

MATERIALS AND METHODS

Plasmid isolation and sequencing. *P. abyssi* GE5 cells were cultivated in anaerobic conditions at 95°C in rich medium containing sulfur, as described previously (14). The plasmid pGT5 was purified on a cesium chloride gradient after sodium dodecyl sulfate lysis of the cells and drastic deproteinization (6). For pGT5 sequencing, two *Sau*3A fragments and two *Sac*I fragments encom-

* Corresponding author. Mailing address: Institut de Génétique et Microbiologie, URA 1354 Centre National de la Recherche Scientifique, Bâtiment 409, Université Paris-Sud, 91405 Orsay-Cedex, France. Phone: 33 1 69 41 74 89. Fax: 33 1 69 41 78 08. Electronic mail address: forterre@igmors.u-psud.fr.

passing the *Sau*3A junctions were cloned into the vectors pAT153 and M13mp18/19, respectively. Both strands were sequenced by primer walking (44) by the dideoxy chain termination method (39).

Sequence analyses. DNA and protein sequence analyses were performed with the Genetics Computer Group (University of Wisconsin Biotechnology, Madison, Wis.) software package version 8.0 and PC/GENE (IntelliGenetics Inc., Mountain View, Calif.). Secondary structures were identified with the Genetics Computer Group program RNA-FOLD. The nucleotide and protein sequence similarity searches in GenBank, EMBL, Swissprot, and the National Biomedical Research Foundation were done with the programs Fasta and BLAST. The coding probability of the open reading frames (ORFs) was tested by using the program TESTCODE (Genetics Computer Group) together with the programs FRAME (3) and COD-PROK (PC/GENE).

Identification of pGT5 ssDNA. The search for ssDNA of pGT5 was performed according to the method of Strauss et al. (44). The lysates were obtained by treating 50-ml suspensions of rapidly cooled cells with guanidine thiocyanate (10). The lysates were treated with RNases, and plasmidic DNA was precipitated with ethanol. The pellet was dissolved in 20 μl of Tris-EDTA, pH 7.5, and total DNA was analyzed by electrophoresis in 1% agarose gels containing 1 μg of ethidium bromide per ml. After being transferred to nitrocellulose membranes, DNA, with or without prior denaturation, was hybridized with pGT5 probes labelled using the Genius system (Boehringer, Mannheim, Germany). Double-stranded probes were labelled with digoxigenin (DIG)-dUTP incorporated by random priming of the whole pGT5 plasmid. To prepare ss pGT5 RNA probes, pGT5 was subcloned into the Bluescript plasmid (Stratagene), which contains specific promoters around the polylinker. RNA probes were labelled in vitro with DIG-UTP using T3 (plus-strand probe) or T7 polymerase (minus-strand probe).

To check the functionality of the pGT5 putative *sso* (ss origin), a 567-bp fragment (position 3108 to 3231) containing region RI was introduced in the ssDNA-overproducing plasmid pHV33Δ*HaeII* (46). This construct was used to transform *Escherichia coli* and *Bacillus subtilis*. The preparation of cell lysates and the identification of ssDNA were performed according to the method of Noirot-Gros and Ehrlich (34).

Nucleotide sequence accession number. The plasmid pGT5 from *P. abyssi* GE5 (3,444 bp) has been sequenced on both strands (GenBank accession number, U49503).

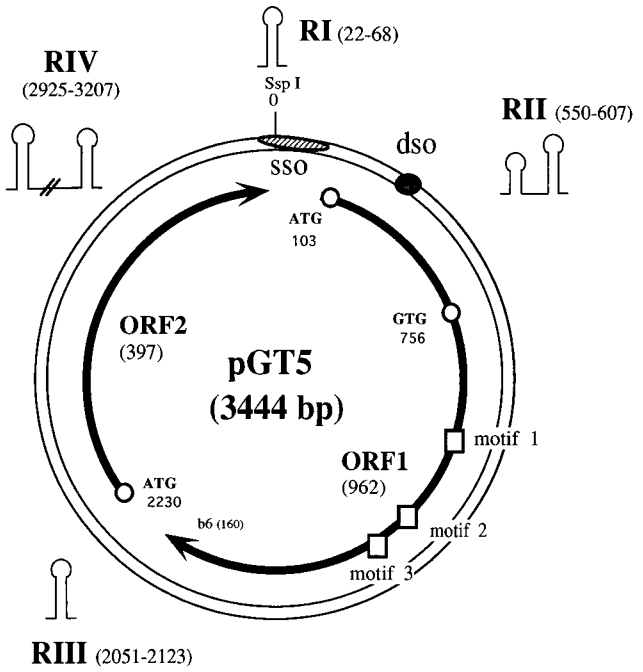

FIG. 1. Map of pGT5. The two major ORFs (1 and 2) are indicated. The two possible initiation codons for ORF1 are noted together with the three motifs (motifs 1 to 3) common to Rep proteins of RC plasmids. The putative *dso* and *sso* and the four regions rich in direct and inverted repeats (RI to RIV) are indicated.

RESULTS

Structural features of the pGT5 sequence. The sequence of the pGT5 plasmid from *P. abyssi* GE5 is numbered from the first base of the unique site *Ssp*I. The GC content (43.4%) is similar to that of the *P. abyssi* chromosomal DNA (44.7%). Numerous direct and inverted repeats can be detected. Many of them are clustered into four regions (RI to RIV) (Fig. 1). The putative hairpins produced by the inverted repeats in regions RI, RIII, and RIV have GC-rich stems (not shown), suggesting that they could be maintained in vivo at a high temperature. Three repeats, AGGTTC, CCTTGAG, and AT TTTG, are dispersed within the sequence of the first ORF (see below).

ORF analysis. Analysis of the six possible phases revealed 15 ORFs encoding putative polypeptides of at least 50 amino acids. The two major ORFs (ORF1 and -2) are on the same strand and in the same frame (Fig. 1). They are separated by the region RIII. These two ORFs cover 85% of the total sequence and encode putative basic proteins of 75 and 46 kDa, respectively. All the other ORFs are much smaller. Analysis of the coding capacity (42) of the pGT5 sequence and individual ORFs revealed only two regions with high coding probabilities, corresponding to ORF1 and -2 (not shown). However, this analysis suggests that ORF1 might not start at ATG in position 103 (ATG-103) but further downstream, at GTG-756. The search for translation and transcription signals also suggested that ORF1 and -2 are bona fide genes. Putative ribosome binding sites can be detected upstream of ATG-103 and GTG-756 and upstream of the ATG initiation codon of ORF2 (not shown).

ORF1 encodes a putative polypeptide with motifs characteristic of RC Rep proteins. Computer searches failed to detect similarities between any of the 15 putative ORFs and nucleic acid or protein sequences in the databases. Looking for known structural motifs by using PROSITE, we detected only two potential leucine zipper motifs in the C-terminal region of ORF1 (positions 568 to 589 and 575 to 596). However, visual inspection indicates that ORF1 contains three motifs (motif 1 to 3) which are present in the same arrangement in Rep proteins encoded by RC replicons related to the *E. coli* bacteriophage ϕX174 (25) (Fig. 2A). These Rep proteins are site-specific endonucleases and ligases which function in both initiation and termination steps of RC replication. Motif 2 includes two histidines which could be involved in metal ion coordination. Motif 3 contains the tyrosine required to cleave one of the two DNA strands at the origin of double-stranded DNA replication, generating a free 3′-OH primer for the DNA polymerase (17, 36). This tyrosine is also involved in the formation of a covalent phosphotyrosine link between DNA and the Rep protein.

The ϕX174 related RC replicons have been divided in two superfamilies by Koonin and Ilyina (27). Superfamily I, which includes ϕX174, is characterized by two tyrosines in motif 3, involved alternatively in the nicking-closing process. Motif 3 of pGT5 harbors a single tyrosine, indicating that it might belong to superfamily II (Fig. 2A). This superfamily can be further divided into three families: one of them comprises bacterial plasmids related to pC194 from *Staphylococcus aureus*, the second groups bacterial plasmids related to pMV158 from *Streptococcus agalactiae*, whereas the third one corresponds to plant geminiviruses (27). We noticed that the three pGT5 motifs most closely resemble those of the pC194 family (Fig. 2A). Indeed, Fig. 2B shows that the region surrounding motif 2 of pGT5 ORF1 can be aligned with the corresponding regions of Rep proteins from pC194 and related plasmids. In addition to the two histidines, four amino acids are conserved at the same positions in pGT5 and seven sequences of the pC194 subfamily. Such an alignment could not be obtained with sequences around motifs 2 of the Rep proteins from the

A)

		Motif 1	Motif 2	Motif 3
Superfamily I				
ΦX174	v	FDTLTLADD	LHFHAVHFM	FYVAKYVNKKSDM
pGRB1	a	MVTLTASST	VHIHLGVFV	AYLAAYMAGEYGS
pHK2	a	MLTFTASSV	SHLHVGVYF	SYLAAYMGGYTEE
pEE	b	FITLTLPPA	LHLHIVMVG	AYMGKYLSKGTQ
pRQ7	b	FLTLTSSLE	VHMHILFRG	RYMMKYLSKEMEG
Superfamily II				
pGT5	**a**	FVLTAPKDV	PHFHIDAIV	FFELKYASRKLFV
pC194	b	FLTLTVRNV	PHFHVLIAV	YEMAKYSGKDSDY
pMV158	b	FLLYPESIP	AHYHVLYIA	KKKHVYDKADIKL
ABMV	v	FLTYPQCSL	PHLHVLIQF	SDVKSYIDKDGDT

B)

pGT5	NLLFGFTINVHVTGDKNP---FEPHFHIDAIVTFICY--DK----SSTKWF
pC194	VIK-GYVRKLEITY-NKKRDDYNPHFHVLIAVNKSYFT-DKRYYISQQEWL
pUB110	NLV-GFMRATEVTIMNK-DNSYNQHMHVLVCVEPTYF-KNTENYVNQKQWI
pLAB1000	NLL-GYVRSTEITIMKN--GTYHQHMHVLLFVKPTYF-KDSANYINQAKWS
pST1	NLL-GYLRSVEVTHNEN-DKTYHPHIHVLMMVRPSYF-QSKKDYITQKEWS
pKYM	PVQ-GWIRATEVTRGKT-DGSAHPHFHCLLMVQPSWF-KGK-NYVKHERWV
pVA380	NVI-GYLRATEVTYS-TEHENYHPHLHVLLFVKSSYFTGNNTNYISQEEWT
pNostoc	FPAEGYIKTVEVTRGKTPDGSAHPHFHVLMMVKPSYF---GVGYLSQAKWV

FIG. 2. Identification of a putative Rep protein in pGT5. (A) Alignment of motifs 1, 2, and 3 of pGT5 ORF1 with a selection of similar motifs from Rep proteins of RC replicon superfamilies I and II (27). RC replicons are from viruses (v), *Archaea* (a), or *Bacteria* (b). Inside each subgroup, the amino acids present in both pGT5 and other replicons are shown as boldface characters. (B) Alignment of the region surrounding motif 2 of the pGT5 Rep protein with a selection of the corresponding regions in Rep proteins from plasmids of the pC194 family. Amino acids identical in pGT5 and other sequences are underlined. Amino acids present in all sequences are shown as boldface characters. The hosts are the following: *Halobacterium halobium* GRB (pGRB1), *H. volcanii* Aa2.2 (pHK2), cyanobacteria (pEE), *Thermotoga* sp. strain RQ7 (pRQ7), *S. aureus* (pC194), *S. agalactiae* (pMV158), Akutilon mosaic virus (ABMV), *S. aureus* (pUB110), *Lactobacillus hilgardii* (pLAB1000), *Streptococcus thermophilus* (pST1), *S. sonnei* (pKYM), and *Streptococcus ferus* (pVA380). For details, see references 9, 20, 27, and 41.

two other families included in superfamily II or with those of Rep proteins from superfamily I (not shown), despite the fact that the archaeal plasmids pGRB1 and pHK2 belong to superfamily I (Fig. 2A).

Identification of a sequence identical to that of RC *dso*. RC replicons exhibit two origins of replication, one to initiate replication of the double-stranded DNA, called the double-stranded origin (*dso*), and the other to initiate the conversion of the ss replication intermediate into double-stranded DNA, called the ss origin (*sso*) (17, 36). The *dso* is characterized by a specific DNA sequence that bears the cleavage site for the Rep protein. The sequence of pGT5 contains a stretch of 11 nucleotides identical to those of the *dso* of pC194 and of several related plasmids (Fig. 3A). This sequence is located between the two possible initiation codons of ORF1 (Fig. 1), in agreement with the typical organization of RC plasmids in which the *dso* is located just upstream of or within the *rep* gene and in the same strand (36). Fig. 3B shows an alignment of the pGT5 *dso* sequence with those of RC replicons from the different families of the ϕX174-related RC replicons. The number of conserved nucleotides with the *dso* sequences of RC replicons outside the pC194 family is reduced, in agreement with the assignment of pGT5 to this family. The only two nucleotides conserved in all *dso* sequences of the ϕX174 group are those on each side of the nick introduced by the Rep protein.

Regions surrounding the best-analyzed *dso* sequences of RC replicons are complex, being asymmetric and rich in short direct and inverted repeats (22, 48). These regions include, from 5′ to 3′, the termination site (overlapping the *dso*), the cleavage site, and the Rep binding site. Indeed, we detected one inverted repeat and two direct repeats around the putative *dso* of pGT5, as well as three overlapping imperfect direct repeats at the 3′ end of the pGT5 *dso* that are similar to repeats observed at the 3′ ends of other *dso* sequences and could correspond to the Rep protein binding site (not shown). Since the Rep proteins of RC plasmids studied up to now act as multimers (9, 47), the potential leucine zippers detected in the C-terminal region of the pGT5 Rep protein could be involved in oligomerization. A putative leucine zipper has been previously detected in the RepB protein of the RC plasmid pMV158 (9).

Possible *sso* similar to bacterial *ssi*. The *sso* sequences of RC replicons are typically long stretches of DNA (100 to 250 bp) with the potential to form extensive secondary structures (41). These *sso* sequences are recognized by host RNA polymerases which synthesize an RNA primer to initiate the replication of the complementary strand. *sso* sequences are located just upstream of the *dso*, so that replication of the ssDNA cannot start before completion of a first round of replication. Thus, region RI, located upstream of the putative *dso* of pGT5, was a good candidate to harbor the *sso* of pGT5 (Fig. 1). We failed to detect sequence similarities between RI and the two different types of *sso* (*palA* and BA3) typical of the pC194 family (36). However, it was possible to align the sequence surrounding the putative hairpin of RI with a particular group of bacterial sequences, called *ssi* (ss initiation) sequences, that can be used

A)

pGT5	GTTGGGTTATCTTG ATA
pC194	TTCTTTCTTATCTTG ATA
pBAA1	TCTTTTCTTATCTTG ATA
pLP1	TTTCTTCTTATCTTG ATA
pLpeI	TTTCCTCTTATCTTG ATA
pBC1	TTTTTTCTTATCTTG ATA

B)

pGT5	GGGTTTATCTTG ATA
pC194 (II)*	TTTCTTATCTTG ATA
pKYM (II)*	AATCCAAACTTG ATA
pMV158 (II)	GGGGCTACTACG ACC
pGRB-1 (I)	GCTGGTACGACG ACC
ΦX174 (I)	TCCCCCAACTTG ATA

C)

S1 S2

```
pGT5      5'ATTATCCTTTAGAAACGCGTTGGGTATTGGGGGAGCCCCCCTTAGTGGGGAGCTCCCCCTAAACACCC--CCAAGACAATAGGAAAC 3'
Rsc-13      TTTCAGTTATGGAAACGCCGGGAGCGGGGAAAACTTGCTTTTTCCCGTTTC-CGGGGATTGGACAACGAGCAACGCGA-AGGCGTC
R6-5        TTTCAGTTATTGAAACGCCGGGAGCGGGGAAAACTTGCTTTTTCCCGTTTC-CGGGG-TTGGACAACTGAGCAACGCGA-AGGCGTC
R100        TTTCAGTTATTGAAACGCCGGGAGCGGGGAAAACTTGCTTTTTCCCGTTTC-CGGGG-TTGGACAACTGAGCAACGCGA-AGGCGTC
                      ****                                                          * *  *
                    motif A                                                       motif B
```

FIG. 3. Identification of putative pGT5 *dso* and *sso*. Alignment of the sequence of the putative *dso* of pGT5 with those of *dso* sequences from different pC194-related plasmids (A) and *dso* sequences from plasmids of different RC superfamilies (B). Identical nucleotides in all sequences are shown as boldface characters; nucleotides identical to those in the pGT5 sequence are underlined. A space indicates the cleavage site. The superfamily (I or II) is given in parentheses, and plasmids from the pC194 family are marked with asterisks in panel B. The hosts not indicated in the legend to Fig. 2 are the following: *Lactobacillus plantarum* (pLP1), *Lactobacillus plantosus* (pLpe1), and *Bacillus coagulans* (pBC1). (C) Alignment of the region RI of pGT5 with those of *ssi* sequences detected in bacterial plasmids Rsc-13, R6-5, and R100. Nucleotides conserved in *ssi* sequences of plasmids pKYM, F, and pAMβ1 and bacteriophage ϕX174 are labelled with asterisks.

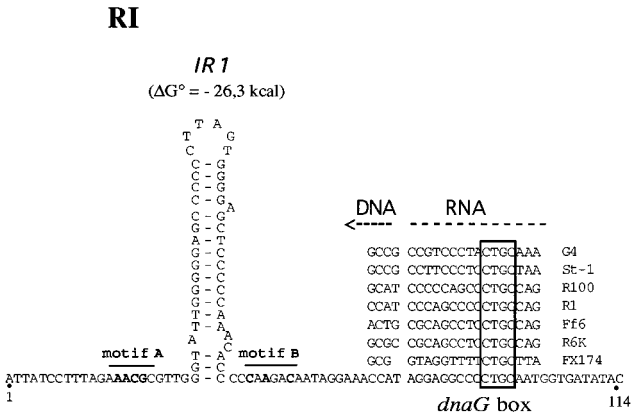

FIG. 4. Putative hairpin in the *sso* region of pGT5 and DNA primase start site. The alignment of primase binding sites of bacterial phages and plasmids is from Kornberg and Baker (28).

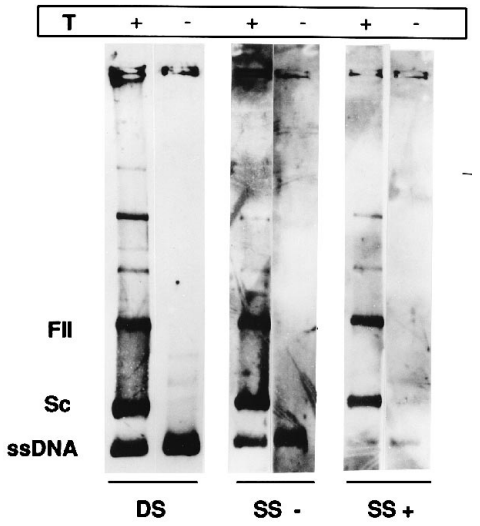

FIG. 5. Identification of the pGT5 ss replication intermediate. Purified DNAs from *P. abyssi* were run on agarose gels and transferred (T) with (+) or without (−) denaturation on nylon membranes. The different forms of the plasmid were detected with either a pGT5 double-stranded probe (DS) or an RNA probe corresponding to the minus strand (SS −) or to the plus strand (SS +). FII, open pGT5; Sc, supercoiled pGT5; ssDNA, ss pGT5.

to initiate DNA replication on an ss template. The best alignment was obtained between the RI region of pGT5 and the *ssi* sequences of Rsc-13, R6-5, and R100 (Fig. 3C). One could identify two DNA segments, S1 and S2, with 64 and 50% similarity on 17 and 24 nucleotides, respectively.

Initially, *ssi* sequences were detected in *E. coli* plasmids that replicate via a theta mode for their ability to complement the poor growth of M13 phage derivatives lacking their original *sso* sequence (35). In theta plasmids, they could be used to initiate lagging-strand DNA replication. Recently, similar *ssi* sequences have also been identified in the RC plasmid pKYM from *Shigella sonnei*, which belongs to the pC194 family (41), and in the theta replicating plasmid pAMβ1 (*ssiA*) from the gram-positive bacterium *Enterococcus faecalis* (5). The *E. coli ssi* sequences have been classified into several groups according to sequence similarities. Those resembling the RI region of pGT5 belong to the group I-B and originated from plasmids F-f2b (*ssiC*), R100 (*ssiB*), R6-5, and Rsc-13 (35, 41). Interestingly, the *ssi* sequences of this group can also be specifically aligned with the *sso* of ϕX174 (32) and with the *ssi* of pKYM and the *ssiA* of pAMβ1 (5, 41) (not shown). The alignment of the pGT5 sequence with these sequences allows the identification of two conserved motifs (A and B) inside segments S1 and S2, respectively, which are separated by a region without similarity (Fig. 3C). Interestingly, motifs A and B are located on both sides of the putative hairpin of RI, whereas the divergent region corresponds to the stem of the hairpin (Fig. 4). The latter is GC rich in pGT5, compared with the equivalent region in bacterial *ssi* sequences (Fig. 3C). This is reminiscent of the GC enrichment of the hairpin stems in stable RNAs from hyperthermophiles, strongly suggesting that the RI sequence forms a stable hairpin in vivo.

The *sso* of ϕX174 and group I *ssi* sequences are both assembly sites for the primosome in *E. coli* (32), whereas the *ssiA* of pAMβ1 is used by the *B. subtilis* primosome (5). Primosomes are complexes of several host proteins which load the bacterial primase, DnaG, onto the DNA to synthesize an RNA primer (28). We searched for a primase recognition sequence in the vicinity of the putative *sso* of pGT5. A perfect match was found between a sequence downstream of *sso* motif B and the priming site used by DnaG in *E. coli* (Fig. 4). This sequence includes the conserved DnaG box, 3′-CGTC-5′, which corresponds to the start site (first ribonucleotide, A), and a C residue located just downstream of the RNA-DNA transition point.

Identification of a pGT5 ssDNA replication intermediate. The replication of RC plasmids generates an ss intermediate (ssDNA) that can sometimes be detected in crude extracts of host cells. We have looked for pGT5 ssDNA in lysates of *P. abyssi* GE5, using a pGT5 probe. We observed by Southern blotting a band that migrated more rapidly than supercoiled pGT5 in agarose gels in the presence of ethidium bromide, as is expected for ssDNA (45) (Fig. 5). This band was transferred on the membrane without denaturation (Fig. 5) and disappeared after S1 nuclease treatment (not shown), demonstrating that it was made of ssDNA. The ssDNA produced in the course of the replication of the RC replicons corresponds to the plus strand, i.e., the strand which is cleaved by the Rep protein and displaced by the DNA polymerase. To confirm that the pGT5 ssDNA detected in *P. abyssi* was an RC replication intermediate, we verified that it hybridized specifically to the minus strand (defined in our case as the strand which does not contain the *sso*). RNA probes specific to each one of the two pGT5 strands were prepared. As expected for an RC replication intermediate, only the RNA probe corresponding to the minus strand hybridized with the pGT5 ssDNA (Fig. 5).

DISCUSSION

We report here the first sequence of a plasmid isolated from a hyperthermophilic archaeon, the plasmid pGT5 from *P. abyssi*. Sequence analysis revealed two major ORFs on the same strand and four regions rich in putative secondary structures. Our data strongly support the argument for an RC mechanism of replication for pGT5. The largest ORF encodes a basic protein exhibiting similarities with Rep proteins of RC plasmids of the pC194 family. A putative *dso* composed of 11 bases, identical to the *dso* of pC194 and its relatives, is located on the same strand (the plus strand) and upstream of the *rep* gene. The sequence of a putative *sso*, upstream of the *dso*, can be aligned with that of a bacterial *ssi* of group I-B (primosome dependent). It includes a putative GC-rich hairpin (as expected for a hyperthermophile) and a bacterial primase (*dnaG*) recognition sequence. Finally, *P. abyssi* cells contain an

ss form of pGT5 which corresponds to the plus strand, as expected for a bona fide RC replication intermediate.

The Rep protein of pGT5 can start at either ATG-103 or GTG-756. We favor the second possibility for the following reasons: (i) the region included between ATG-103 and GTG-756 has a low coding probability; (ii) the *dso* of pGT5 is located downstream of the ATG, and known *dso* sequences of plasmids from the pC194 family are located upstream of the *rep* gene (36); and (iii) the size of the Rep protein starting from GTG-756 (50 kDa) is more in the range of Rep proteins typical of this family.

Most genes in RC plasmids are oriented in the same direction as the replication gene (36). Indeed, the second-largest ORF of pGT5 (ORF2) is located on the plus strand, as is the Rep gene. This ORF could encode a protein involved in recombination (PRE) such as those detected in other plasmids of the pC194 family (36). The region RII, which contains two putative hairpins (Fig. 1) and is located between ATG-103 and GTG-756, could be involved in copy number regulation via antisense RNA interfering with the expression of the Rep gene, as reported for other RC plasmids (9, 36). We analyzed the transcription pattern of pGT5 to identify expressed genes and transcription signals. Several mRNAs hybridizing specifically with pGT5 probes were detected by Northern (RNA) blot analysis. However, these putative transcripts were present in very small amounts and usually degraded, preventing their identification and the determination of their start sites. This might be due to heat-induced hydrolysis of RNA at high temperatures in hyperthermophiles (16).

The lack of an efficient transformation method and genetic markers for *Pyrococcus* spp. has precluded a direct functional analysis of the different putative genetic elements identified by our sequence analysis. We checked for the activity of the putative pGT5 *sso* in *E. coli* and *B. subtilis*, using the system designed by te Riele and coworkers (46). A pGT5 DNA fragment containing region RI was cloned into the plasmid pHV33Δ*Hae*. This plasmid replicates partly via the RC mechanism but has no *sso*, leading to the accumulation of ssDNA in host cells. Insertion of pGT5 region RI did not reduce this accumulation in either *E. coli* or *B. subtilis* at either 37 or 48°C (not shown), indicating that pGT5 *sso* was not functional in these hosts. This is not really surprising since, for example, *ssiA* of pAMβ1 is functional in gram-positive bacteria but not in *E. coli* (5).

Our sequence analysis is critical in establishing a vector base on pGT5. The best site for the insertion of foreign DNA into pGT5 might be in between ORF1 and -2 or in ORF2, which might not be required for plasmid replication, since RC plasmids such as pGRB1 and pRQ7 contain a single ORF (18, 20). Recent experiments indicate that recombinant pGT5 containing foreign DNA inserted between ORF1 and -2 can be propagated into *P. furiosus* and related strains (1, 53).

The plasmid pGT5 could also become an interesting tool to study archaeal DNA replication at temperatures near the boiling point of water. In particular, the nature of its putative *sso* indicates that the components of the bacterial primosome, as well as the bacterial primase DnaG, have homologs in the *Archaea*. Since many archaeal proteins have close homologs in the *Eucarya*, our data suggest that primosomal proteins might also exist in the *Eucarya* but have not yet been detected. The existence of primosomal proteins in the last common ancestor of the three domains of life would be in agreement with the hypothesis that the initiation of replication of the bacterial chromosome by the primosomal protein PriA might have been an old mechanism predating that for the initiation of chromosome replication via the DnaA protein (31).

Phylogeny of RC plasmids. RC replicons are ubiquitous in the three domains of life. They have been tentatively classified according to the similarities of their Rep proteins (27). There is clearly no correlation between the phylogeny of RC replicons and that of their hosts. For example, different plasmids from gram-positive bacteria belong to different families of superfamily II, whereas archaeal and cyanobacterial plasmids are distributed between superfamilies I and II (Fig. 2). This could be explained either by the extensive lateral transfer of RC replicons between the different domains and bacterial kingdoms or by the formation of the various RC replicon superfamilies (or even families) before the divergence of the *Archaea*, *Bacteria*, and *Eucarya*. In favor of the latter hypothesis, Seery and coworkers reported that a phylogeny of Rep proteins restricted to the pC194 family is roughly in agreement with the bacterial phylogeny deduced from 16S rRNA analysis (41). They concluded that the bacterial hosts apparently do not exchange RC plasmids at a high frequency.

The finding that an archaeal plasmid and a family of eucaryotic viruses (geminiviruses) both belong to superfamily II of RC replicons is also interesting. Koonin and Ilyina suggested that geminiviruses evolved from RC plasmids of gram-positive bacteria (27). Our finding indicates that these viruses could have also evolved from archaeal plasmids, which is in agreement with the suggestion that *Archaea* and *Eucarya* are sister groups (49). However, another possibility is that viruses preceded plasmids in the course of evolution and that RC plasmids originated by reduction from RC viruses during the evolutionary period prior to the separation of the three domains (15). Further studies of pGT5 and other archaeal plasmids and viruses should help to decide between those alternative hypotheses.

ACKNOWLEDGMENTS

We are grateful to M. F. Noirot-Gros for kindly providing plasmids pHV33Δ*HaeII* and for her advice in testing the pGT5 *sso*. We thank M. Dyall-Smith and M. Holmes for communication of data before publication and J. Di Ruggiero for critical reading of the manuscript.

This work was supported by grants from the European Union Biotechnology program (BIO-CT93-02734) and the Association de la Recherche sur le Cancer (ARC).

REFERENCES

1. **Aagaard, C., I. Leveiev, R. N. Aravalli, P. Forterre, D. Prieur, and R. A. Garrett.** General vectors for archaeal hyperthermophiles: strategies based on a mobile intron and a plasmid. FEMS Microbiol. Rev., in press.
2. **Adams, M. W. W.** 1993. Enzymes and proteins from organisms that grow near and above 100°C. Annu. Rev. Microbiol. **47:**627–658.
3. **Bibb, M. J., P. R. Findlay, and M. W. Johnson.** 1989. The relationship between base composition and codon usage in bacterial genes and its use for simple and reliable identification of protein coding sequences. Gene **30:**157–166.
4. **Bokranz, M., A. Klein, and L. Meile.** 1990. Complete nucleotide sequence of plasmid pME2001 of *Methanobacterium thermoautotrophicum* strain Marburg. Nucleic Acids Res. **18:**363.
5. **Bruand, C., S. D. Ehrlich, and L. Jannière.** 1995. Primosome assembly site in *Bacillus subtilis*. EMBO J. **14:**2642–2650.
6. **Charbonnier, F., P. Forterre, G. Erauso, and D. Prieur.** 1995. Purification of plasmids from thermophilic and hyperthermophilic archaebacteria, p. 87–90. *In* F. T. Robb (ed.), Thermophiles, archaea: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. **Charlebois, R. L., M. L. Lam, S. W. Cline, and F. W. Doolittle.** 1987. Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaebacterium. Proc. Natl. Acad. Sci. USA **84:**8530–8534.
8. **De la Campa, A. G., G. H. del Solar, and M. Espinoza.** 1990. Initiation of replication of plasmid pLS1. The initiator protein RepB acts on two distant DNA regions. J. Mol. Biol. **213:**247–262.
9. **Del Solar, G., M. Moscoso, and M. Espinosa.** 1993. Rolling circle-replicating plasmids from gram-positive and Gram-negative bacteria: a wall falls. Mol. Microbiol. **8:**789–796.
10. **Di Ruggiero, J., and F. T. Robb.** 1995. RNA extraction from sulfur-utilizing

thermophilic archaea, p. 97–99. *In* F. T. Robb (ed.), Thermophiles, archaea: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

11. **Erauso, G.** 1994. Les Thermococcales du Bassin Nord-Fidjien, description d'une nouvelle espèce: *Pyrococcus abyssi* et caractérisation de son plasmide. Ph.D. thesis. Université de Bretagne occidentale, Brest, France.
12. **Erauso, G., F. Charbonnier, T. Barbeyron, P. Forterre, and D. Prieur.** 1992. Preliminary characterization of an ultrathermophilic archaebacterium with a plasmid isolated from a North Fiji basin hydrothermal vent. C. R. Acad. Sci. **314:**387–393.
13. **Erauso, G., A. Godfroy, G. Raguenes, and D. Prieur.** 1995. Plate cultivation techniques for strictly anaerobic, thermophilic, sulfur-metabolising archaea, p. 25–29. *In* F. T. Robb (ed.), Thermophiles, archaea: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. **Erauso, G., A. L. Reysenbach, A. Godfroy, J. R. Meunier, B. Crump, F. Partensky, J. A. Baross, V. Marteinsson, G. Barbier, N. R. Pace, and D. Prieur.** 1993. *Pyrococcus abyssi*, sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. Arch. Microbiol. **160:**338–349.
15. **Forterre, P.** 1992. New hypotheses about the origins of viruses, prokaryotes and eukaryotes, p. 221–234. *In* J. K. Trân Thanh Vân, J. C. Mounolou, J. Shneider, and C. McKay (ed.), Frontiers of life. Éditions Frontières, Gif-sur-Yvette, France.
16. **Forterre, P.** 1995. Thermoreduction, a hypothesis for the origin of prokaryotes. C. R. Acad. Sci. **318:**415–422.
17. **Gruss, A., and S. D. Ehrlich.** 1989. The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. Microbiol. Rev. **53:**231–241.
18. **Hackett, N. R., M. P. Krebs, S. Dassarma, W. Goebel, U. L. Rajbhandary, and H. G. Khorana.** 1990. Nucleotide sequence of a high copy number plasmid from *Halobacterium* strain GRB. Nucleic Acids Res. **18:**3408.
19. **Hall, M., and N. R. Hackett.** 1989. DNA sequence of a small plasmid from *Halobacterium* strain GN101. Nucleic Acids Res. **17:**10501.
20. **Harriott, O. T., R. Huber, K. O. Stetter, P. W. Betts, and K. M. Noll.** 1994. A cryptic miniplasmid from the hyperthermophilic bacterium *Thermotoga* sp. strain RQ7. J. Bacteriol. **176:**2759–2762.
21. **Herbert, R. A.** 1992. A perspective on the biotechnological potential of extremophiles. Trends Biotechnol. **10:**395–402.
22. **Higashitani, A., D. Greenstein, H. Hirokawa, S. Asano, and K. Horiuchi.** 1994. Multiple DNA conformational changes induced by an initiator protein precede the nicking reaction in a rolling circle replication origin. J. Mol. Biol. **237:**388–400.
23. **Holmes, M. L., S. D. Nuttall, and M. L. Dyall-Smith.** 1991. Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. J. Bacteriol. **173:**3807–3813.
24. **Holmes, M. L., F. Pfeifer, and M. Dyall-Smith.** 1995. Analysis of the halobacterial plasmid pHK2 minimal replicon. Gene **153:**117–121.
25. **Ilyina, T. V., and E. V. Koonin.** 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. **20:**3279–3285.
26. **Kagramanova, V. K., N. I. Derckecheva, and A. S. Mankin.** 1988. The complete nucleotide sequence of the archaebacterial plasmid pHSB from *Halobacterium*, strain SB3. Nucleic Acids Res. **16:**41–58.
27. **Koonin, E. V., and T. V. Ilyina.** 1993. Computer-assisted dissection of rolling circle DNA replication. BioSystems **30:**241–268.
28. **Kornberg, A., and T. Baker.** 1992. DNA replication. Freeman and Company, New York.
29. **Krebs, M. P., T. Hausst, M. P. Heyn, U. L. Rajbhandary, and H. G. Khorana.** 1991. Expression of the bacterioopsin gene in *Halobacterium halobium* using a multicopy plasmid. Proc. Natl. Acad. Sci. USA **88:**859–863.
30. **Lam, W., and W. F. Doolittle.** 1989. Shuttle vectors for the archaebacterium *Halobacterium volcanii*. Proc. Natl. Acad. Sci. USA **86:**5478–5482.
31. **Masai, H., T. Asai, Y. Kubota, K. I. Arai, and T. Kogoma.** 1994. *Escherichia coli* PriA protein is essential for inducible and constitutive stable replication. EMBO J. **13:**5338–5345.
32. **Masai, H., N. Nomura, Y. Kubota, and K. Arai.** 1990. Roles of the ϕX174 type primosome- and G4 type primase-dependent primings in initiation of lagging and leading strand syntheses of the DNA replication. J. Biol. Chem. **265:**15124–15133.
33. **Ng, W.-L., and S. DasSarma.** 1993. Minimal replication origin of the 200-kilobase *Halobacterium* plasmid pNRC100. J. Bacteriol. **175:**4584–4596.
34. **Noirot-Gros, M. F., and S. D. Ehrlich.** 1994. Detection of single-stranded plasmid DNA. Methods Mol. Genet. **3:**370–379.
35. **Nomura, N., H. Masai, M. Inuzuka, C. Miyazaki, E. Ohtsubo, T. Itoh, S. Sasamoto, M. Matsui, R. Ishizaki, and K. I. Arai.** 1991. Identification of eleven single-strand initiation sequences (ssi) for priming DNA replication in the F, R6K, R100 and ColE2 plasmids. Gene **108:**15–22.
36. **Novick, R. P.** 1989. Staphylococcal plasmids and their replication. Annu. Rev. Microbiol. **43:**537–565.
37. **Pfeifer, F., and P. Gharaman.** 1993. Plasmid pHH1 of *Halobacterium salinarium*: characterization of the replicon region, the gas vesicle gene cluster and insertion elements. Mol. Gen. Genet. **238:**193–200.
38. **Reiter, W. D., W. Zillig, and P. Palm.** 1988. Archaebacterial viruses. Adv. Virus Res. **34:**143–188.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
40. **Schleper, C., I. Holz, D. Janekovic, J. Murphy, and W. Zillig.** 1995. A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. J. Bacteriol. **177:**4417–4426.
41. **Seery, L. T., N. C. Nolan, P. M. Sharp, and K. M. Devine.** 1993. Comparative analysis of the pC194 group of rolling-circle plasmids. Plasmid **30:**185–196.
42. **Staden, R., and A. D. McLachlan.** 1982. Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res. **10:**141–156.
43. **Stetter, K. O.** 1992. Life at the upper temperature border, p. 195–219. *In* J. K. Trân Thanh Vân, J. C. Mounolou, J. Shneider, and C. McKay (ed.), Frontiers of life. Éditions Frontières, Gif-sur-Yvette, France.
44. **Strauss, E., J. Kobori, G. Siu, and L. Hood.** 1986. Specific-primer directed DNA sequencing. Anal. Biochem. **154:**353–360.
45. **te Riele, H., B. Michel, and S. Ehrlich.** 1986. Single-stranded plasmid DNA in *Bacillus subtilis* and *Staphylococcus aureus*. Proc. Natl. Acad. Sci. USA **83:**2541–2545.
46. **te Riele, H., B. Michel, and S. Ehrlich.** 1986. Are single-stranded circles intermediates in plasmid DNA replication? EMBO J. **5:**631–637.
47. **Thomas, C. D., D. F. Balson, and W. V. Shaw.** 1990. *In vitro* studies of the initiation of staphylococcal plasmid replication. J. Biol. Chem. **265:**5519–5530.
48. **Wang, P.-Z., S. J. Projan, V. Henriquez, and R. Novick.** 1993. Origin recognition specificity in pT181 plasmids is determined by a functionally asymmetric palindromic DNA element. EMBO J. **12:**45–52.
49. **Woese, C. R., O. Kandler, and M. L. Wheelis.** 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA **87:**4576–4579.
50. **Yeats, S., P. McWilliam, and W. Zillig.** 1982. A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. EMBO J. **1:**1035–1038.
51. **Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendorfer, and J. K. Kristjansson.** 1994. Screening for Sulfolobales, their plasmids and their viruses in Icelandic solfataras. Syst. Appl. Microbiol. **16:**609–628.
52. **Zillig, W., S. Yeats, I. Holz, A. Bock, F. Gropp, M. Rettenberger, and S. Lutz.** 1985. Plasmid-related anaerobic autotrophy of the novel archæbacterium *Sulfolobus ambivalens*. Nature (London) **313:**789–791.
53. **Zivanovic, Y.** Unpublished results.