

Novel RepA-MCM proteins encoded in plasmids pTAU4, pORA1 and pTIK4 from *Sulfolobus neozealandicus*

BO GREVE,¹ SUSANNE JENSEN,¹ HOA PHAN,¹ KIM BRÜGGER,¹ WOLFRAM ZILLIG,² QUNXIN SHE¹ and ROGER A. GARRETT^{1,3}

¹ Danish Archaea Centre, Institute of Molecular Biology, University of Copenhagen, Sølvgade 83H, DK-1307 Copenhagen, Denmark

² Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany

³ Corresponding author (garrett@mermaid.molbio.ku.dk)

Received September 2, 2004; accepted November 8, 2004; published online November 29, 2004

Summary Three plasmids isolated from the crenarchaeal thermoacidophile *Sulfolobus neozealandicus* were characterized. Plasmids pTAU4 (7,192 bp), pORA1 (9,689 bp) and pTIK4 (13,638 bp) show unusual properties that distinguish them from previously characterized cryptic plasmids of the genus *Sulfolobus*. Plasmids pORA1 and pTIK4 encode RepA proteins, only the former of which carries the novel polymerase–primase domain of other known *Sulfolobus* plasmids. Plasmid pTAU4 encodes a mini-chromosome maintenance protein homolog and no RepA protein; the implications for DNA replication are considered. Plasmid pORA1 is the first *Sulfolobus* plasmid to be characterized that does not encode the otherwise highly conserved DNA-binding PlrA protein. Another encoded protein appears to be specific for the New Zealand plasmids. The three plasmids should provide useful model systems for functional studies of these important crenarchaeal proteins.

Keywords: chemotaxis, crenarchaeal plasmid, DNA replicase, MCM protein.

Introduction

Over the past decade, a major effort has been made to investigate a diverse collection of plasmids and viruses associated with the genera *Sulfolobus* and *Acidianus* of the crenarchaeal order Sulfolobales (Zillig et al. 1994, 1998, Prangishvili et al. 1998, 2001). This work has underpinned a program to elucidate basic molecular and cellular mechanisms that are characteristic of the Archaea, in particular of the kingdom Crenarchaeota. Whereas viruses from the Crenarchaeota show considerable morphological and genomic diversity (Prangishvili et al. 2001, Prangishvili and Garrett 2004), many of the plasmids have been grouped into a few related families. For example, the cryptic plasmids pRN1, pRN2, pDL10, pHEN7 and pSSVx have been assigned to a single pRN family on the basis of a conserved genomic region (Keeling et al. 1998, Arnold et al. 1999, Kletzin et al. 1999, Peng et al. 2000). Three pRN-type plasmids (pXQ1, pST1 and pST3) are also found inte-

grated within *Sulfolobus* chromosomes (Peng et al. 2000, She et al. 2002). In addition, all the known crenarchaeal conjugative plasmids, pNOB8, pING1, pARN2, pARN3, pHVE14 and pKEF9, share extensive genome conservation (She et al. 1998, Stedman et al. 2000, Greve et al. 2004).

The conserved region of the pRN plasmids encodes a RepA protein, the N-terminal region of which exhibits novel DNA polymerase and primase activities (Lipps et al. 2003, 2004). Moreover, a partial homolog of this N-terminal region is encoded, separately, in the self-transmissible plasmids of *Sulfolobus* (Greve et al. 2004). The conserved pRN plasmid region also encodes a DNA-binding protein, PlrA, which has a DNA regulatory function (Kletzin et al. 1999, Lipps et al. 2001b) and is the most highly conserved of the *Sulfolobus* plasmid-encoded proteins (Garrett et al. 2004). A CopG family protein is also encoded (Lipps et al. 2001a). One pRN plasmid, pSSVx from *Sulfolobus islandicus* REY15/4, has the unusual property that it can spread as a virus satellite in the presence of the fusellovirus SSV2, a process probably facilitated by two proteins of fusellovirus origin that are encoded in the non-conserved region of pSSVx (Arnold et al. 1999).

Recently, three small, high-copy-number cryptic plasmids, pTAU4, pORA1 and pTIK4, were isolated from *Sulfolobus neozealandicus* strains (Zillig et al. 1998, this work). The strain TIK4/2, harboring pTIK4, can outcompete other *Sulfolobus* strains, swarms on Gel-rite plates and shows chemotaxis, although it is unknown whether plasmid-encoded products facilitate this process (Zillig et al. 1998). Studies with DNA–DNA hybridization and 16S rRNA sequence analyses show that the *S. neozealandicus* strains are fairly heterogeneous (Zillig et al. 1998). Moreover, chromosomal fragmentation analyses on the plasmid-harboring strains of *S. neozealandicus* suggest that they have diverged considerably from *Sulfolobus* strains found in other geographical locations including Iceland, Japan and the Naples region of Italy (P. Redder and R.A. Garrett, unpublished observations). Given the divergent properties of the *S. neozealandicus* hosts, we sequenced and analyzed the plasmids in order to examine further the diversity of *Sulfolobus* plasmids.

Materials and methods

Plasmid preparation and DNA sequencing

Sampling, enrichment and plating of *Sulfolobus* strains were performed as described previously (Zillig et al. 1996, 1998). Plasmids were isolated from *S. neozealandicus* strains 32/4 (pTAU4), 43/1 (pORA1) and 4/2 (pTIK4). Plasmid DNA was isolated on Qiagen columns (QIAfilter, Plasmid Maxi Kit, Qiagen, Westburg, Germany) and digested with *EcoRI* to confirm the plasmid identity. Each plasmid was sonicated and fragments of about 2 kb were selected for generating shotgun libraries in pUC18. Cloned DNA was isolated in a Biorobot 8000 (Qiagen) and sequenced in MegaBACE 1000 Sequencers (Amersham Biotech, Amersham, U.K.). Plasmid sequences were assembled by Sequencher 3.1.2 (Gene Codes, Ann Arbor, MI). Ambiguous sequences were resolved by primer walking.

Comparative sequence analyses

Open reading frames (ORFs) encoding more than 40 amino acids were found with the program MUTAGEN (Brügger et al. 2003). Searches for sequence matches were made against public databases with BLAST2 (Altschul et al. 1997). Open reading frame comparisons and pI determinations were performed with MUTAGEN. Open reading frame sequences were also checked for conserved motifs and protein family relationships (<http://pfam.wustl.edu/> and <http://www.ncbi.nlm.nih.gov/COG/new/>). Transmembrane regions were identified with TMHMM (<http://www.cbs.dtu.dk/services>). Signal peptide sequences were predicted with SignalP (<http://www.cbs.dtu.dk/services>). Amino acid repeat sequences were identified with RADAR (Heger and Holm 2000). Direct and inverted repeats of nucleotide sequences were detected with DNA Strider (Marck 1988). Sequences were aligned with T-Coffee (Notredame et al. 2000).

Results and discussion

Libraries were generated for each of the three plasmids in the cloning vector pUC18. A fivefold plasmid sequence coverage was generated and any remaining sequence ambiguities were resolved by primer walking on the plasmid DNA. The sizes of the three plasmids are: pTAU4, 7,192 bp; pORA1, 9,689 bp; and pTIK4, 13,638 bp. Both sizes and sequences are consistent with restriction digest patterns generated on agarose gels (data not shown). The G+C contents of the plasmids are: pTAU4, 34.8%; pORA1, 37.8%; and pTIK4, 35.1%. Sequence accession nos. are: pTAU4, AJ852505; pORA1, AJ862826; and pTIK4, AJ852506.

Initially, all ORFs larger than 40 amino acids were located with MUTAGEN (Brügger et al. 2003). Regions of 50 bp upstream of each putative start codon were then screened for TATA-like and BRE sequence motifs, and the -5 to -15 bp upstream region was screened for Shine-Dalgarno (S-D) motifs. If these sequence motifs were undetected, then the analysis was repeated on the next potential downstream start codon. The sequence region immediately downstream from the puta-

tive stop codon was also examined for potential T-rich terminator sequences (Reiter et al. 1988).

Gene maps are presented for each plasmid in Figure 1. Most of the predicted genes are single and carry AUG, GUG or UUG start codons. Only a few pairs of genes appear to be arranged in operons. They include the RepA, or mini-chromosome maintenance (MCM), genes coupled with upstream CopG genes, as well as ORFA, a putative CopG paralog in pORA1, and the adjacent ORF206 and ORF181 in pTIK4 (Figure 1). For the single genes and the first gene in an operon, transcripts appear leaderless, whereas an S-D sequence precedes the second gene in each operon. This follows the pattern observed earlier for single genes and first genes of operons in *Sulfolobus* chromosomes (Tolstrup et al. 2000, Torarinsson et al. 2004). Most putative ORFs are transcribed in one direction for pTAU4 but not for pORA1 and pTIK4. The ORFs that yielded sequence matches in public databases are listed in Table 1 and are considered below.

RepA proteins of pORA1 and pTIK4

The ORF866 from pORA1 shows good sequence matches to putative RepA proteins encoded in pRN plasmids, primarily in the N-terminal and central regions, whereas ORF1053 of pTIK4 exhibits significant sequence similarity over the central and downstream regions (26/48% identity/similarity to positions 170 to 849 of ORF866). The N-terminal region (positions 90 to 280) of the pTIK4 protein, which carries a 120-aa extension, shows a significant sequence match to some bacterial DNA replicases (Brüggemann et al. 2003).

The N-terminal region of the RepA protein of plasmid pRN1 has been examined experimentally and its crystal structure has been determined. It carries novel DNA polymerase and primase activities and exhibits a DNA-binding domain and an active site subdomain (Lipps et al. 2003, 2004). A sequence alignment of this N-terminal region is provided for the RepA proteins of the pRN plasmids (pRN1, pSSVx, pIT3), pORA1 and pTIK4 (Figure 2). This shows that the DNA-binding domain and the active site subdomain of the pRN RepA, amino acids critical for function and a zinc binding site, are conserved in the pORA1 protein but not in that of pTIK4 (Lipps 2003, Lipps et al. 2004). Moreover, the homologous region of the separate ORF106, conserved in the self-transmissible plasmids of *Sulfolobus* (Greve et al. 2004), which is included in the alignment (Figure 2), carries the conserved DNA-binding domain but not the active site subdomain. The latter function may be supplied by another of the small proteins encoded in the same operon as ORF106 (Greve et al. 2004).

The central regions of ORF866/RepA and ORF1053/RepA (~150 aa) show sequence similarity to the RepA homolog of pRN1, and the helicase superfamily III (Koonin 1993). For the pRN1 RepA protein, the Walker A motif was shown to be essential for ATPase activity, which was strongly stimulated in the presence of duplex DNA, although efficient unwinding of DNA was not detected (Lipps et al. 2003). This domain also gave a significant sequence match to ORF671 (unknown func-

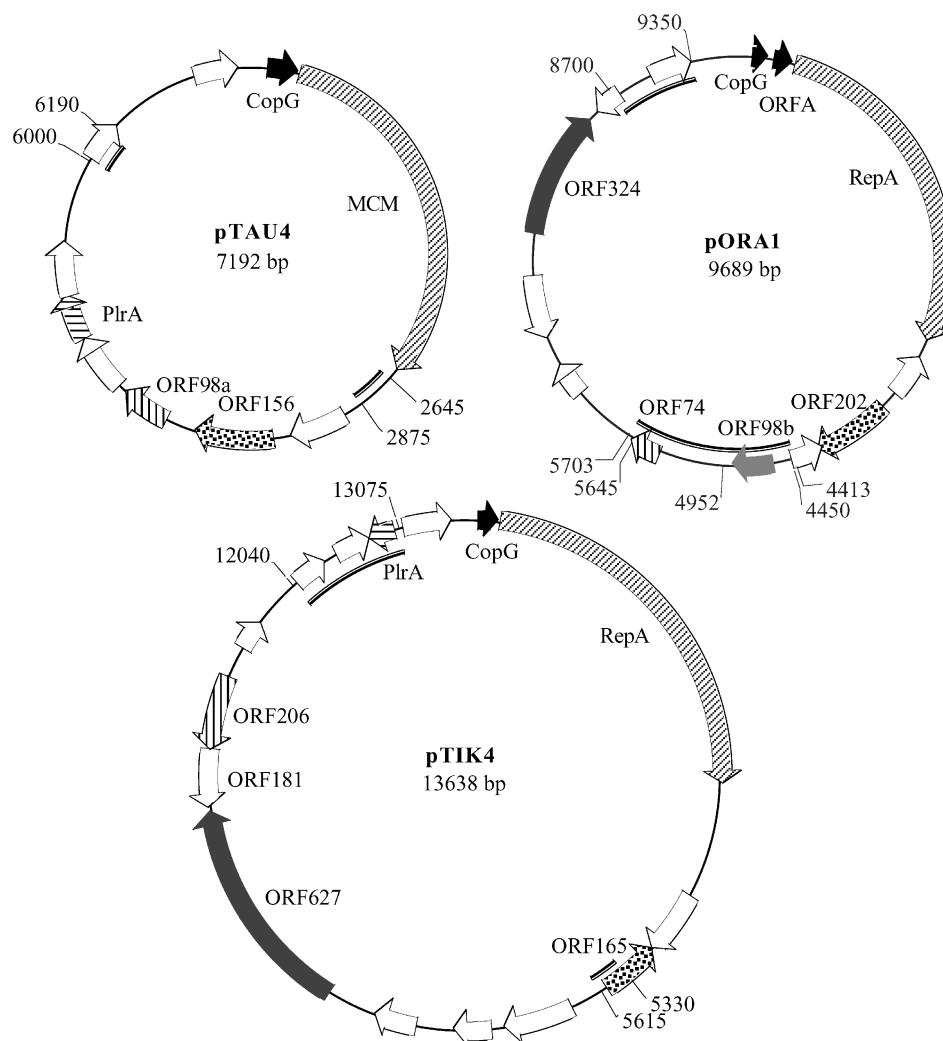


Figure 1. Circular gene maps of plasmids pTAU4, pORA1 and pTIK4. Open reading frames (ORFs) presented in either black, or identical patterns, are homologous. Gray-shaded ORFs yield matches in public sequence databases (see Table 1). Regions exhibiting a high concentration of direct or inverted repeats, or both, are drawn inside the circular genome with a thick black line and the nucleotide positions bordering the regions are given. The positions of five identical 18-bp palindromic repeats in pORA1 are indicated between positions 4413 and 5703. Position 1 of the genome sequence immediately precedes the promoter region of the CopG gene in each plasmid.

tion) in the *Sulfolobus* conjugative plasmid pKEF9 (19/45% identity/similarity over about 150 aa) (Greve et al. 2004).

Mini-chromosome-maintenance protein of pTAU4

Plasmid pTAU4 is the only cryptic plasmid of *Sulfolobus* that does not encode a putative RepA protein. However, it does encode a protein, ORF759, which is homologous to MCM proteins encoded in archaeal chromosomes and proteins MCM2 to MCM7, which are involved in initiation of DNA replication in eukaryotes (Tye et al. 1999). Sequence similarity to the archaeal chromosomal proteins is high, including those of *Sulfolobus solfataricus* (42/61% identity/similarity), *Aeropyrum pernix* (35/54% identity/similarity) and *Methanocaldococcus jannaschii* (26/45% identity/similarity), and the pTAU4 protein differs significantly only in that it carries an additional N-terminal region (60 aa) of unknown function.

Alignment of the MCM protein of pTAU4 with matches in the GenBank/EMBL databases reveals that the central region (positions 180 to 670) is well conserved in all MCM proteins with 30–46% identity and 50–65% similarity. Although the N-terminal 50 amino acids of most proteins show little significant conservation, the C-terminal 90 amino acids are, with few

exceptions, conserved among the archaeal proteins, including the pTAU4 protein (data not shown). The protein also carries the conserved Walker A-, Walker B- and C-motifs which are characteristic of the MCM proteins and essential for ATP binding and hydrolysis (Koonin 1993; Figure 3).

The MCM protein of pTAU4 shows low sequence similarity to ORF866/RepA of pORA1 (14/23% identity/similarity over positions 200–866). This observation is consistent with earlier bioinformatic analyses, which demonstrated that archaeal and eukaryal MCM proteins, and plasmid-encoded RepA proteins, are all members of the AAA+ class of P-loop ATPases, which are related to the helicase superfamily III enzymes which either unwind or open DNA (or RNA) duplexes (Koonin 1993, Iyer et al. 2004). Thus, the apparent replacement of a gene for RepA by one for an MCM protein in pTAU4 reinforces the view that there is a functional homology between the two classes of proteins.

CopG family protein

A small ORF immediately upstream from the putative RepA and MCM genes is conserved in the three plasmids. An alignment is presented in Figure 4A. The closest matches are be-

Table 1. Open reading frame (ORF) analysis of the plasmids pTAU4, pORA1 and pTIK4. Transmembrane spanning helices (TMH) were predicted with TMHMM and signal peptides were predicted by Signal-P (<http://www.cbs.dtu.dk/services>).

ORF	pTAU4	pORA1	pTIK4	Sequence identity/ similarity (%)	TMH motifs	Signal peptide	Annotation
1	ORF63	ORF44	ORF62	24-31/40-44		–	Regulatory, CopG
2		ORF50/ORFA				–	Regulatory, CopG-like
3		ORF866	ORF1053	26/48		–	Replication initiation, RepA
4	ORF759					–	Replication initiation, MCM
5	ORF157	ORF202	ORF166	22–48/44–51		–	Highly charged protein
6	ORF80		ORF72	35/56		–	Regulatory, PlrA
7		ORF98b				–	snRNP
8	ORF98a	ORF74	ORF206	20–25/40–53	3, 2, 2	+	Homology C-terminus ORF206
9			ORF181		3	+	Operon with ORF206
10		ORF324		22–38/43–53	5	–	<i>Sulfolobus</i> chromosome DUF974 domain
11			ORF627		1	+	Conjugative plasmid ORF

tween the pORA1 and pTIK proteins but as a group they are fairly distant from CopG family proteins of other known *Sulfolobus* plasmids where they are invariably conserved upstream from a RepA gene. The CopG homolog has been expressed from pRN1 and shown to bind to an inverted repeat sequence in its promoter region, thereby regulating expression of both the CopG homolog and the downstream RepA gene (Lipps et al. 2001a). However, no similar repeat sequences were detected upstream from the CopG–RepA operons of pORA1 and pTIK4. In pORA1, as well as in pXQ1, pSSVx and pST1, a smaller ORFA (a putative paralog of CopG) is located between the CopG and RepA ORFs (Garrett et al. 2004).

PlrA regulatory protein

The ORF80 and ORF72 of pTAU4 and pTIK4, respectively, encode the PlrA protein (Kletzin et al. 1999). This protein has been expressed and shown to bind site-specifically to double-stranded DNA, and carries a novel type of basic leucine zipper motif (Lipps et al. 2001b). The PlrA protein probably constitutes an important regulatory protein (Kletzin et al. 1999, Lipps et al. 2001b), an assumption that is consistent with its high sequence conservation in all known *Sulfolobus* plasmids (Garrett et al. 2004). Exceptionally, however, it is absent from pORA1.

ORF156, ORF202 and ORF165

The three homologs ORF156, ORF202 and ORF165 yield no sequence matches with other *Sulfolobus* plasmids and seem to be specific for the *S. neozealandicus* plasmids. Sequence alignments show that the proteins are highly charged, carrying 37–42% lysine, arginine, glutamic acid and aspartic acid, with lysine (12–16%) and glutamic acid (10–15%) predominating (Figure 4B). No matches were detected in public databases. The larger ORF202 of pORA1 contains a leucine zipper motif between positions 101 and 129 (Figure 4B).

ORF98a, ORF74 and ORF206

The ORF98a of pTAU4 and ORF74 of pORA1 both show significant sequence similarity to the C-terminal region of

ORF206 of pTIK4. All three ORFs are predicted to carry two to three transmembrane-helical motifs (TMH) and ORF206 of pTIK4 exhibits a signal peptide sequence (Table 1) and contains a repeat sequence (44 aa) not found in the other two ORFs. The ORF206 lies in a putative operon with ORF181 (Figure 1), which also carries three putative TMHs and exhibits a signal peptide sequence (Albers and Driessen, 2002). If the pTIK4 plasmid contributes to the swarming and chemotaxis of its host, TIK4/2, the latter two proteins would be possible contributors.

The following three ORFs, which are found exclusively in a single plasmid, gave significant matches with public sequence databases.

ORF98b of pORA1

The ORF98b is homologous to the protein component of a group of small nuclear ribonucleoproteins (snRNPs) involved in RNA maturation and processing. The ORF98b shows best sequence matches with ORFs found in *S. tokodaii* and *S. solfataricus* chromosomes (40/75% identity/similarity), but shows high identities/similarities (up to 36/58%) to U6 snRNA-associated Sm-like proteins of eukaryotes. The ORF lies in a region bordered by perfect 18-bp palindromic sequences (Figure 1), which could have been recombined into the plasmid, possibly together with genes of functional RNAs.

ORF324 of pORA1

No homologs of ORF324 were detected in other *Sulfolobus* plasmids, although they are present in *Sulfolobus* chromosomes where they are annotated as hypothetical proteins. There are five in *S. solfataricus* P2, four in *S. tokodaii* and four in *S. acidocaldarius*, all within the size range of 273–331 aa. They show 22–38% identity over most of the sequence with best matches in the C-terminal regions. They belong to a protein family (pfam) containing a conserved domain DUF973 (unknown function) that spans most of the protein. Each ORF exhibits four to six putative TMH motifs and approximately half of the ORFs carry a putative signal peptide. The ORF324 carries five putative TMHs but no signal peptide.

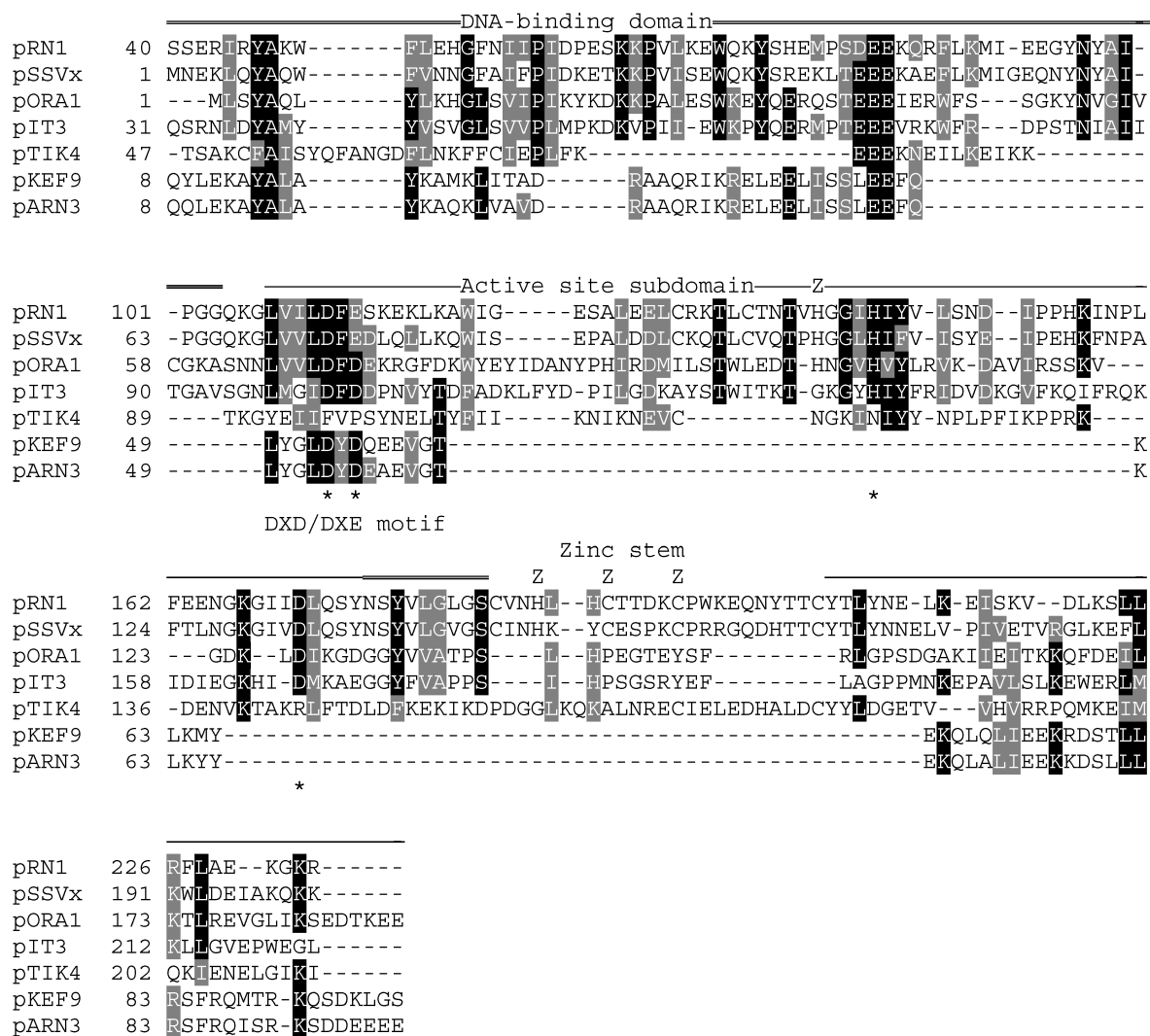


Figure 2. Alignments of the N-terminal regions of RepA homologs from the *Sulfolobus* cryptic plasmids pRN1, pSSVx, pORA1, pIT3 and pTIK4 and the conjugative plasmids pKEF9 and pARN3. The double line above the alignment denotes the DNA-binding domain and the single line delineates the active site subdomain. Functionally important amino acids in the latter are indicated by asterisks, and the DXD/DXE motif is shown (Lipps 2003, Lipps et al. 2004). Amino acids involved in the zinc-binding stem are denoted by Z. The sequence of putative RepA homolog of pIT3 is from GenBank accession no. AY591755 (H.-P. Klenk, e-Gene Biotechnologie GmbH, Feldafing, Germany, unpublished data). Alignments of identical and similar amino acids are indicated by black and gray boxes, respectively, obtained with Boxshade (<http://www.ch.embnet.org>).

ORF627 of pTIK4

The ORF627 is homologous to a protein in the size range of 660–780 aa that is found in all known self-transmissible plasmids of *Sulfolobus*. Together with seven to nine other proteins it is considered to generate the conjugative apparatus (Greve et al. 2004). Open reading frames showing sequence similarity to the N-terminal 35–350 aa are found in three *Sulfolobus* chromosomes with the following identities/similarities: *S. solfataricus* 22/36%; *S. tokodaii* 27/40%; and *S. acidocaldarius* 22/38%. In *S. tokodaii*, the ORF is located within an integrated conjugative plasmid, and in *S. solfataricus* the ORF is partitioned. As for many of these putatively secretable proteins, ORF627 is rich in asparagine and tyrosine and, as for the *Sulfolobus* chromosomal homologs, it carries a putative signal peptide sequence (Albers and Driessen 2002).

Repeat sequences

Each of the plasmids contains regions that are relatively rich in direct or inverted repeats, or both, and these are marked on the genome maps. Plasmid pORA1 carries five copies of the 18-bp palindromic repeat, GTGTTTAAATTTAAACAC, two pairs of which are separated by 20 and 40 nucleotides. These double repeats are separated by 1.2 kb, with a fifth sequence located between them (Figure 1). Several other direct and inverted repeats are located in this region and at a second site about 650 bp upstream from the CopG ORF. In pTAU4, two regions of about 200 bp are rich in direct and inverted repeats, one of which is located immediately downstream from the CopG–MCM operon. In pTIK4, direct and inverted repeats are concentrated in two regions of 300 and 1000 bp.

In all three plasmids, two repeat-rich regions are located

	Walker A	Walker B	C-motif
pTAU4 (411)	V L I I G D P G T A K S Q L L Q F V K N L 43	F L L I D E F D K I 32	T V I A V A N P
Sso (335)	I L I I G D P G T A K S Q M L Q F I S R V 43	I A V I D E I D K M 32	A V I A A G N P
Ape (338)	V L F V G D P G V A K S Q L L Q S T A Q V 43	I A V I D E F D K M 32	S L L A A G N P
Afu (209)	I L L V G D P G V A K S Q L L K Y V H R I 43	I A L V D E I D K M 32	A L L G A A N P
Sce-2 (538)	V L L L G D P G T A K S Q L L K Y V E K T 43	V C L I D E F D K M 32	S I I A A A N P
Sce-3 (404)	I L M V G D P S T A K S Q L L R F V L N T 43	V V C I D E F D K M 32	S V I A A A N P

Figure 3. Alignments of Walker A and Walker B motifs and the C-motif of the mini-chromosome maintenance (MCM) protein of pTAU4 with those encoded in the archaeal chromosomes of *S. solfataricus* (Sso), *A. permix* (Ape) and *Archaeoglobus fulgidus* (Afu), as well as the *Saccharomyces cerevisiae* MCM2 (Sce-2) and MCM3 (Sce-3) proteins. Identical and similar amino acids are indicated by black and gray boxes, respectively, obtained with Boxshade (see Figure 2).

roughly on opposite sides of the plasmids (Figure 1), suggesting that they could function as sites for initiation and termination of theta-like replication. However, no similar repeat sequences were shared between the three plasmids, and none of the repeat sequences match those found at the replication origins of *Sulfolobus* chromosomes (Robinson et al. 2004).

Conclusions

Each of the three *Sulfolobus* plasmids exhibits unusual properties when compared with one another and with other *Sulfolobus* plasmids. Most of these differences relate to the RepA/MCM proteins. At present, we know little about the mechanisms of DNA replication of *Sulfolobus* plasmids. It was proposed that pRN plasmids replicated by a rolling circle mechanism (Kletzin et al. 1999, Peng et al. 2000), but more recent work supports a theta mechanism (Lipps et al. 2003, 2004). The present work shows that the RepA protein of pORA1 is

clearly related to those of the pRN plasmids, but the pTIK4 protein carries a different DNA polymerase/primase N-terminal domain, and pTAU4 encodes an MCM protein which may be functionally homologous to the RepA proteins. This indicates, for the first time, that the replication apparatus of *Sulfolobus* plasmids can be quite diverse. Thus, each of the three plasmids may provide a simple model system for studying functional roles of important *Sulfolobus* proteins, including defining the role of an archaeal MCM protein in replication (pTAU4), investigating a novel polymerase-primase domain (pTIK4) and examining the function of the ubiquitous and highly conserved PlrA protein (pORA1).

Acknowledgments

We thank Patrick Forterre and Eugene Koonin for helpful discussions. This work was supported by a grant for an Archaea Center from the Danish Natural Science Research Council and by EU Grant No. QLK3-CT-2000-00649.

A

```

ORF63_pTAU4      1 MGETQMGRGNYDNRKRIVAKFELDEFELINLKKGNKTWSEFILELAG-YKKKENEMGRNSDGQR--
ORF44_pORA1      1 -----MRYP LVS-ITHTIPY--EYKRLVEKKGDKTWQVLLLEWA-----ETQGMKG-----
ORF62_pTIK4      1 ---MNPKEKESKIRIDIPVKKEEYLVLKVKGNKSWRQFILELAKLYEEMQNNKEKAHEMIAQA

```

B

```

ORF202_pORA1     1 MSTKKEEKIICPKCGQPCSWTVESEGSVFCVHYLGYEKTSDGKIKKKIKKHYMGRVOYVLRPGLQPLLR
ORF156_pTAU4     1 -----MADK--IAMQTKGKRGRKP--KYETKTCPCNCGLPYRST
ORF165_pTIK4     1 -----MKAMNSRRITKEDMREI LAKKSLPKKID EK--GRELYKPCRCSEFARYI

ORF202_pORA1     71 VKDELKHIEERIKIIEEETAKEKDEQKKEELLRKRELEKEKEGLERRVKEIETKNHILPADDKERFLK
ORF156_pTAU4     35 VEREVN-GHVVYVYAVHQYKDENGKWRKYECYLGPKQY-----IHAEKFNDFGLKGGHVE-----ERFVE
ORF165_pTIK4     48 DVDERD-GHFYIYAIHYNGTRGHGKPKLERHYLGAEY-----DYMERFNNINMQGLFNE-----KRHVE

ORF202_pORA1     141 YIKESLELMQLQDD--FDERKIMKMDIIESFMKNKEIMSEKTLAKVRLIENDV VNAIYSRQG
ORF156_pTAU4     94 YLER-IMRDQLTVDRRFDFAKGIQMLNTIFNLLNEKAETEDEKQLLLETLEGLKAKMEKGEKDS
ORF165_pTIK4     107 YIK--NAANQIDSRMTAYDFAETMDSLSKNMKSMIIDENDKK-ILEEKMKKIFQMLEK--KDH

```

Figure 4. Alignment of the amino acid sequences of (A) the CopG homologs from the three *S. neozealandicus* plasmids and (B) homologs ORF202, ORF156 and ORF165, which are specific for the *S. neozealandicus* plasmids. The ORF202 of pORA1 carries a putative N-terminal signal peptide and a leucine zipper motif marked by arrows. Identical and similar amino acids are indicated by black and gray boxes, respectively, obtained with Boxshade (see Figure 2).

References

- Albers, S.V. and A.M. Driessen. 2002. Signal peptides of secreted proteins of the archaeon *Sulfolobus solfataricus*: a genomic survey. *Arch. Microbiol.* 177:209–216.
- Altschul, S., T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 2:3389–3402.
- Arnold, H.P., Q. She, H. Phan, K. Stedman, D. Prangishvili, I. Holz, J.K. Kristjansson, R.A. Garrett and W. Zillig. 1999. The genetic element pSSVx of the extremely thermophilic crenarchaeon *Sulfolobus* is a hybrid between a plasmid and a virus. *Mol. Microbiol.* 34:217–226.
- Brüggemann, H., S. Bäumer, W.F. Fricke et al. 2003. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proc. Natl. Acad. Sci. USA* 100:1316–1321.
- Brügger, K., P. Redder and M. Skovgaard. 2003. MUTAGEN: multi-user tool for annotating genomes. *Bioinformatics* 19:2480–2481.
- Garrett, R.A., P. Redder, B. Greve, K. Brügger, L. Chen and Q. She. 2004. Archaeal plasmids. In *Plasmid Biology*. Eds. B.E. Funnell and G.J. Phillips. A.S.M. Press, Washington, DC, pp 377–392.
- Greve, B., S. Jensen, K. Brügger, W. Zillig and R.A. Garrett. 2004. Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. *Archaea* 1:231–239.
- Heger, A. and L. Holm. 2000. Rapid automatic detection and alignment of repeats in protein sequences. *Proteins* 41:224–237.
- Iyer, L.M., D.D. Leipe, E.V. Koonin and L. Aravind. 2004. Evolutionary history and higher order classification of AAA⁺ ATPases. *J. Struct. Biol.* 146:11–31.
- Keeling, P.J., H.-P. Klenk, R.K. Singh, M.E. Schenk, C.W. Sensen, W. Zillig and W.F. Doolittle. 1998. *Sulfolobus islandicus* plasmids pRN1 and pRN2 share distant but common evolutionary distance. *Extremophiles* 2:391–393.
- Kletzin, A., A. Lieke, T. Ulrich, R.L. Charlebois and C.W. Sensen. 1999. Molecular analysis of pDL10 from *Acidianus ambivalens* reveals a family of related plasmids from extremely thermophilic and acidophilic archaea. *Genetics* 152:1307–1314.
- Koonin, E.V. 1993. A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res.* 21:2541–2547.
- Lipps, G. 2003. The replication protein of the *Sulfolobus islandicus* plasmid pRN1. *Biochem. Soc. Trans.* 32:240–244.
- Lipps, G., M. Stegert and G. Krauss. 2001a. Thermostable and site-specific DNA binding of the gene product ORF56 from the *Sulfolobus islandicus* plasmid pRN1, a putative archaeal plasmid copy control protein. *Nucleic Acids Res.* 29:904–913.
- Lipps, G., P. Ibanez, T. Stroessenreuther, K. Hekimian and G. Krauss. 2001b. The protein ORF80 from the acidophilic and thermophilic archaeon *Sulfolobus islandicus* binds highly site-specifically to double-stranded DNA and represents a novel type of basic leucine zipper protein. *Nucleic Acids Res.* 29:4973–4982.
- Lipps, G., S. Röther, C. Hart and G. Krauss. 2003. A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. *EMBO J.* 22:2516–2525.
- Lipps, G., A.O. Weinzierl, G. von Scheven, C. Buchen and P. Cramer. 2004. Structure of a bifunctional DNA primase–polymerase. *Nat. Struct. Mol. Biol.* 11:157–62.
- Marck, C. 1988. ‘DNA Strider’: a ‘C’ program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* 16:1829–1836.
- Notredame, C., D. Higgins and J. Heringa. 2000. T-Coffee: a novel method for multiple sequence alignments. *J. Mol. Biol.* 302:205–217.
- Peng, X., I. Holz, W. Zillig, R.A. Garrett and Q. She. 2000. Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. *J. Mol. Biol.* 303:449–454.
- Prangishvili, D. and R.A. Garrett. 2004. Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Bioch. Soc. Trans.* 32:204–208.
- Prangishvili, D., S.-V. Albers, I. Holz et al. 1998. Conjugation in archaea: Frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* 40:190–202.
- Prangishvili, D., K.M. Stedman and W. Zillig. 2001. Viruses of the extremely thermophilic archaeon *Sulfolobus*. *Trends Microbiol.* 9:39–42.
- Reiter, W.D., P. Palm and W. Zillig. 1988. Transcription termination in the archaeobacterium *Sulfolobus*: signal structures and linkage to transcription initiation. *Nucleic Acids Res.* 16:2445–2459.
- Robinson, N.P., I. Dionne, M. Lundgren, V.L. Marsh, R. Bernander and S.D. Bell. 2004. Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* 116:25–38.
- She, Q., H. Phan, R.A. Garrett, S.V. Albers, K.M. Stedman and W. Zillig. 1998. Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon. *Extremophiles* 2:417–425.
- She, Q., K. Brügger and L. Chen. 2002. Archaeal integrative genetic elements and their impact on genome evolution. *Res. Microbiol.* 153:325–332.
- Stedman, K.M., Q. She, H. Phan, I. Holz, H. Singh, D. Prangishvili, R.A. Garrett and W. Zillig. 2000. The pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. *J. Bacteriol.* 182:7014–7020.
- Tolstrup, N., C.W. Sensen, R.A. Garrett and I.G. Clausen. 2000. Two different and highly organized mechanisms of translation initiation in the archaeon *Sulfolobus solfataricus*. *Extremophiles* 4:175–179.
- Torarinsson, E., H.-P. Klenk and R.A. Garrett. 2004. Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.* In press.
- Tye, B.K. 1999. MCM proteins in DNA replication. *Annu. Rev. Biochem.* 68:649–686.
- Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, H. Hain, M. Lanzendörfer and J.K. Kristjansson. 1994. Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. *Syst. Appl. Microbiol.* 16:609–628.
- Zillig, W., D. Prangishvili, C. Schleper, M. Elferink, I. Holz, S. Albers, D. Janekovic and D. Gotz. 1996. Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic archaea. *FEMS Microbiol. Rev.* 18:225–236.
- Zillig, W., H.P. Arnold, I. Holz et al. 1998. Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles* 2:131–140.