Nucleic Acids Research, 1992, Vol. 20, No. 24 ***6501–6507***

Modular organization of related Archaeal plasmids encoding different restriction – modification systems in *Methanobacterium thermoformicicum*

Jörk Nölling, Freek J.M.van Eeden, Rik I.L.Eggen and Willem M.de Vos*
Bacterial Genetics Group, Department of Microbiology, Wageningen Agricultural University, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands

Received October 8, 1992; Revised and Accepted November 23, 1992 EMBL accession nos X68366, X68367

ABSTRACT

Nucleotide sequence comparison of the related 13513-bp plasmid pFV1 and the 11014-bp plasmid pFZ1 from the thermophilic archaeon *Methanobacterium thermoformicicum* THF and Z-245, respectively, revealed a homologous, approximately 8.2 kb backbone structure that is interrupted by plasmid-specific elements. Various highly conserved palindromic structures and an ORF that could code for a NTP-binding protein were identified within the backbone structure and may be involved in plasmid maintenance and replication. Each plasmid contains at comparable locations a module which specifies components of different restriction – modification (R/M) systems. The R/M module of pFV1 contained, in addition to the genes of the GGCC-recognizing R/M system *Mth*TI, an ORF which may be involved in repair of G-T mismatches generated by deamination of m^5C at high temperatures.

INTRODUCTION

Complete nucleotide sequence analysis of genomes may provide insight into their organization and evolution. In contrast to their bacterial counterparts, only limited information exists on extrachromosomal elements from members of the domain *Archaea*, a group of prokaryotic organisms that differs fundamentally from *Bacteria* (1, 2). Nucleotide sequences of only a few extrachromosomal elements isolated from these organisms have been analyzed, including the 15.5-kb genome of the *Sulfolobus shibatae* virus SSV1 (3), three small plasmids, pHSB1 (1.7 kb), pHGN1 (1.8 kb) and pGRB1 (1.8 kb), isolated from extreme halophilic bacteria (4, 5, 6), and the 4.4-kb pME2001 from the methanogen *Methanobacterium thermoautotrophicum* strain Marburg (7).

Recently, a family of three related plasmids was identified in the thermophilic archaeon *Methanobacterium thermoformicicum* (8) comprising the nearly identical 11-kb pFZ1 and pFZ2, harbored by strains Z-245 and FTF, respectively, and the closely related 13.5 kb pFV1 of strain THF. In addition to the high interplasmid similarity, certain plasmids regions were identified with homology to chromosomal DNA from different thermophilic *Methanobacterium* strains that lack plasmid DNA (9). The plasmids may confer a selective advantage to their hosts since it has been shown that restriction and modification (R/M) systems were encoded by pFV1 (10) and pFZ1 and pFZ2 (11).

This report describes the complete nucleotide sequences of pFV1 and pFZ1 and their comparison. Several highly conserved regions were identified that may be involved in replication and maintenance of either plasmid. In addition, a putative DNA-repair enzyme encoded by pFV1 was identified.

MATERIAL AND METHODS

Bacterial strains, plasmids and growth conditions

M.thermoformicicum strains THF and Z-245 were cultivated as described previously on H_2/CO_2 as sole carbon and energy source (8). *Escherichia coli* strains DH5α (Life Technologies Inc., Gaithersburg, Md.) and TG1 (12) were used for propagation of plasmids pUC18 or pUC19 and phages M13mp18 or M13mp19 (13), respectively, and strain Q359 (14) served as host for propagation of bacteriophage λ GEM-12 (Promega Biotec, Madison, Wis.). All *E.coli* strains were grown at 37°C in Luria broth-based media (15). If appropriate, ampicillin (50 μg/ml), isopropyl-β-D-thiogalactopyranoside (1 mM), or 5-bromo-4-chloro-3-indolyl-β-D-galactoside (0.004%) was added.

Chemicals and enzymes

[α-^{32}P]dATP (110 TBq/mmol) was obtained from Amersham International (Amersham, United Kingdom). All enzymes used for manipulation of nucleic acids were purchased from Life Technologies, Inc. or Pharmacia LKB Biotechnologies (Uppsala, Sweden). Oligonucleotides were synthesized with a Biosearch Cyclone DNA synthesizer (New Brunswick Scientific Corp.) at the Netherlands Institute for Dairy Research (NIZO) or were purchased from Pharmacia (Uppsala, Sweden). Other chemicals were obtained in PA quality from Sigma Chemical Co. (St. Louis, Mo.), from Boehringer GmbH (Mannheim, FRG) or from Merck (Darmstadt, FRG).

* To whom correspondence should be addressed

DNA isolation and PCR amplification

Total DNA from methanogenic bacteria was isolated as described previously (8). A pFV1 fragment that spans the single *Xho*I site (position 0/13513) was amplified by PCR as described earlier (16) using total DNA from *M.thermoformicicum* strain THF and two oligonucleotides, 5'-CTGTAGTTCAGGATC-3' and 5'-GAGATACTGTTCAACTG-3', complementary to positions 377 to 363 and 12921 to 12937 of plasmid pFV1, respectively. The amplification product was treated with Klenow DNA polymerase for 30 min at room temperature in the presence of all four dNTPs (2 mM each), followed by an incubation for 20 min at 55°C in the presence of proteinase K (0.1 mg/ml) and sodium dodecyl sulfate (0.5%) essentially as described by Crowe and coworkers (17). Subsequently, the suspension was extracted with phenol/chloroform, the DNA was precipitated with ethanol, and dissolved in TE buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA). After digestion of the amplified DNA with *Hind*II and *Pst*I, the 0.4-kb fragment that contained the *Xho*I-site was recovered from agarose gels using the Geneclean kit (Bio 101, La Jolla, Calif.) and cloned into M13mp19.

DNA manipulations and sequence analysis

DNA manipulations were performed essentially as described by Sambrook and coworkers (15). For the determination of the nucleotide sequence, fragments of cloned pFZ1 (8) and pFV1 DNA (see below) were isolated from agarose gels and subcloned into M13mp18/19 vectors. Sequence reactions with vector- or insert specific primers were carried out by the dideoxy chain termination method (18) using either reagents from the Sequenase kit (United States Biochemical Corp., Cleveland, Ohio) according to the recommendations of the manufacturer or reagents for Taq-polymerase sequencing (19). The nucleotide sequence of plasmid pFV1 was determined on both strands apart from two small regions (positions 876–1182 and 2392–2786) that could not be sequenced in both directions due to a premature termination under all conditions. For plasmid pFZ1 the nucleotide sequence of overlapping fragments was determined on one strand and for 40% on both strands. Computer analysis of sequence data was done with the University of Wisconsin Genetics Computer Group package version 6.0 (20) and the CAOS/CAMM facilities at Nijmegen, The Netherlands.

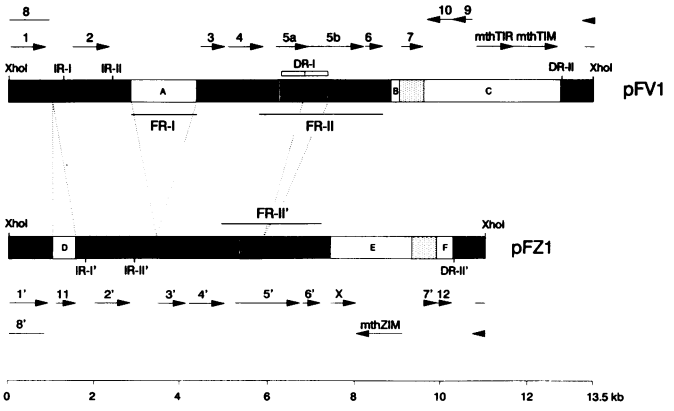

Figure 1. Comparison of the genetic organization of *Xho*I-linearized pFV1 and pFZ1 from *M.thermoformicicum* THF and Z-245, respectively. The degree of interplasmid homology is indicated by shading of the maps: the darkly shaded regions show high sequence similarity whereas the white regions A through F represent plasmid-specific sequences. Each ORF listed in Table 1 is shown by a line with an arrowhead indicating its direction. The positions of regions that contain direct (DR) or inverted (IR) repeats are indicated. Dotted lines show positions of fragment insertions. FR-I and FR-II represent regions with homology to chromosomal DNA from several thermophilic strains of the genus *Methanobacterium*.

RESULTS AND DISCUSSION

Cloning of plasmid pFV1

In a previous report we have shown that *M.thermoformicicum* strain THF harbors the low copy number plasmid pFV1 (8). To be able to analyze this plasmid in detail, covalently closed circular plasmid pFV1 DNA (8) was linearized with *Xho*I and ligated to *Xho*I-digested λ GEM-12. The ligation mix was packaged *in vitro* (15) and used to infect *E.coli* Q359. A 2.3-kb *Kpn*I-fragment of plasmid pFZ1 that strongly hybridized to plasmid pFV1 (8), was labelled by nick-translation and used as probe to screen the recombinant phages. A strongly hybridizing recombinant phage, designated λ GFV1, was identified and, after plaque purification, phage DNA was prepared (15). The integrity of the pFV1 insert of λ GFV1 was verified by restriction analysis and hybridization experiments (data not shown). Purified λ GFV1 DNA was used as source of pFV1 DNA in subsequent experiments.

Nucleotide sequence of plasmid pFV1 and pFZ1 and their comparison

The complete nucleotide sequences of the plasmids pFV1 (EMBL accession number X68366) and pFZ1 (EMBL accession number X68367) from *M.thermoformicicum* strains THF and Z-245, respectively, were determined. The single *Xho*I-site of each plasmids was taken as position 1. Since plasmid pFV1 was originally cloned as a *Xho*I fragment (see above), the uniqueness of this restriction site was confirmed by sequence analysis of a PCR amplified DNA fragment of pFV1 that overlaps the *Xho*I-site. Plasmid pFV1 had a size of 13513 bp and a calculated GC content of 41.8% which is significantly lower than that of its host strain THF (49.6%) (21).

The pFZ1 genome was 11014 bp in length. Alignment of the nucleotide sequences of plasmid pFZ1 and pFV1, schematically shown in Figure 1, revealed a high degree of similarity, confirming previous hybridization experiments (8), and a comparable sequence organization.

Several regions of considerable similarity were identified which are interrupted by plasmid-specific sequences. The homologous sequence blocks, the majority of which shares more than 91% identical nucleotides, add up to approximately 8.2 kb and constitute the backbone structure of both plasmids (Figure 1). As a consequence, these regions have comparable coding capacities and may specify proteins with more than 84% identical amino acids (Table 1).

Two plasmid regions, designated FR-I and FR-II (Figure 1), were found to have homologous counterparts in the chromosomal DNA of several thermophilic *M.thermoformicicum* strains that lack plasmid DNA (9). FR-I resembles an insertion sequence (9) and is specific for plasmid pFV1 (see below), whereas FR-II is part of both plasmids. FR-II of pFV1, however, contains a direct repeat of two identical 524-bp sequences (DR-I, position 6374 to 7421; Figure 1) that is not found in the corresponding region of plasmid pFZ1.

The most conspicuous difference between the plasmids pFV1 and pFZ1 is the presence of three unique sequences that could

be identified in each plasmid (element A–F in Figure 1). The location of two of those elements, A and D, in regions with high interplasmid similarity allowed to determine their exact size. The AT-rich element A (GC content 29.7%), which corresponds to FR-I (Figure 1), is located on plasmid pFV1 (position 2852 to 4352) and comprises 1501 bp. Element D is a 569-bp insertion of plasmid pFZ1 (position 917 to 1485; Figure 1) and contains open reading frame (ORF) 11 (Figure 1). Four other unique sequences, B and C of plasmid pFV1 and E and F of plasmid pFZ1, are located at similar positions relative to the plasmid backbone and flank a region of approximately 0.5 kb that exhibits only relatively low interplasmid similarity (69%; Figure 1). The two large plasmid-specific elements C and E of pFV1 and pFZ1, respectively, code for components of recently characterized R/M systems: the GGCC-recognizing *Mth*TI system of plasmid pFV1 comprising the methyltransferase (MTase) gene *mthTIM* and the endonuclease (ENase) gene *mthTIR* (10), and the CTAG-recognizing system *Mth*ZI of plasmid pFZ1 encoded by the MTase gene *mthZIM* and ORF X which may code for the corresponding ENase (11) (Figure 1). Moreover, two additional ORFs, ORF9 and 10, are located on the unique pFV1-fragment C in close proximity to the genes of the *Mth*TI system, one of which, ORF10, may encode a protein with considerable similarity to enzymes involved in DNA mismatch repair (see below).

Repeated sequences of plasmid pFV1 and pFZ1

The close relationship of pFV1 and pFZ1 (Figure 1) implies a similar replication mechanism for both replicons that should be specified by the conserved backbone structures. Since repeated sequences are known to be involved in the replication of bacterial plasmids (for reviews: 22, 23), we searched in the nucleotide sequence of pFV1 and pFZ1 for regions that contain direct (DR) or inverted repeats (IR).

Apart from DR-I of plasmid pFV1 described above, a region with several DRs was identified in the backbone structure of both plasmids (Figures 1). This 490-bp segment of plasmid pFV1 (DR-II; position 12681 to 13170) contains three perfect 51-bp and two nearly perfect 33-bp repeats. Comparison with the corresponding region of plasmid pFZ1 (DR-II′, position 10265 to 10671) showed, however, only limited similarity, and only the two 33-bp repeats were found to be almost completely conserved.

Inspection of the pFV1 and pFZ1 nucleotide sequences revealed two IRs, IR-I and IR-II, that are contained within highly conserved, non-coding regions of both plasmids (Figure 1). Both IR-I and IR-II exhibit a high GC-content of 52% GC and 62% GC, respectively, and have the capacity to form secondary structures. In IR-I (position 1102 to 1419 in pFV1; Figures 1 and 2), four potential hairpin structures (IR1-IR4, Figure 2) could be identified. Remarkably, the hairpin loops formed by IR1 and IR4 contain a partially overlapping pentamer with the sequence 5′-TCGCT which is repeated four times in the loop of IR1 and five times in the loop of IR4 (Figure 2).

The second region with palindromic sequences, IR-II, is located approximately 1 kb downstream of IR-I (ranging from position 2392 to 2593 in plasmid pFV1) and contains two palindromic sequences, IR5 and IR6 (Figure 2). Analogous to IR-I, two partially overlapping repeats of the pentamer TCGCT are present in both the hairpin loop and the stem/interior loop of IR6, whereas perfect repeats of this pentamer are also found in the region between IR5 and IR6 (Figures 1 and 2).

Both IR-I and IR-II represent one of the most conserved plasmid regions in pFV1 and pFZ1 that differ at only 11 positions out of 520 nucleotides (Figure 2). This high conservation of the IR-regions may indicate an essential function in plasmid maintenance or replication functions. Interestingly, in plasmids that replicate by a rolling circle mechanism via a single-stranded intermediate, (imperfect) palindromic structures are part of the so-called plus and minus origin, the former of which is used for initiation of replication by a plasmid-encoded Rep protein (22). Moreover, the hairpin loops of the minus origins were found to share the consensus sequence 5′-TAGCGT (24). This sequence resembles the complementary sequence of the repeated pentamer 5′-TCGCT which has been observed in hairpin loops of IR-I and IR-II of both plasmids pFV1 and pFZ1. However, we neither could detect single-stranded pFV1- or pFZ1 DNA in hybridization experiments (data not shown) nor found similarity of any of the plasmid-backbone encoded proteins with conserved sequence motifs of rolling circle replication initiator proteins (25). Therefore, the involvement of the IR-regions in replication remains to be confirmed.

Coding capacity and putative genes of plasmid pFV1 and pFZ1

Analysis of the nucleotide sequences of plasmid pFV1 and pFZ1 for coding regions revealed 13 and 12 ORFs, respectively, that could code for polypeptides composed of more than 100 amino

Table 1. Comparison of (putative) genes located on plasmids pFV1 and pFZ1

Gene	plasmid pFV1			Gene	plasmid pFZ1			% Identity
	Amino acids	Start codon	DNA sequence		Amino acids	Start codon	DNA sequence	
ORF1	364	ATG	13304–882	ORF1′	364	ATG	10805–882	99
ORF2	284	ATG	1511–2362	ORF2′	284	ATG	2101–2952	85
ORF3	197	ATG	4438–5028	ORF3′	217	ATG	3455–4105	98
ORF4	237	ATG	5152–5862	ORF4′	234	ATG	4225–4926	91
ORF5a	227	ATG	6229–6909	ORF5′	499	ATG	5295–6791	72
ORF5b	451	GTG	6912–8264	ORF5′	499	ATG	5295–6791	84
ORF6	122	ATG	8354–8719	ORF6′	122	ATG	6894–7259	98
ORF7	145	ATG	9186–9620	ORF7′	80	ATG	9647–9886	71
ORF8	361	ATG	822–13253	ORF8′	361	ATG	822–10754	88
ORF9	146	ATG	10733–10296	ORF11	148	ATG	1093–1536	–
ORF10	221	TTG	10300–9638	ORF12	117	GTG	9889–10239	–
mthTIR	281	ATG	10827–11669	ORFX	202	GTG	7539–8144	–
mthTIM	330	ATG	11669–12658	*mthZIM*	355	ATG	9203–8139	–

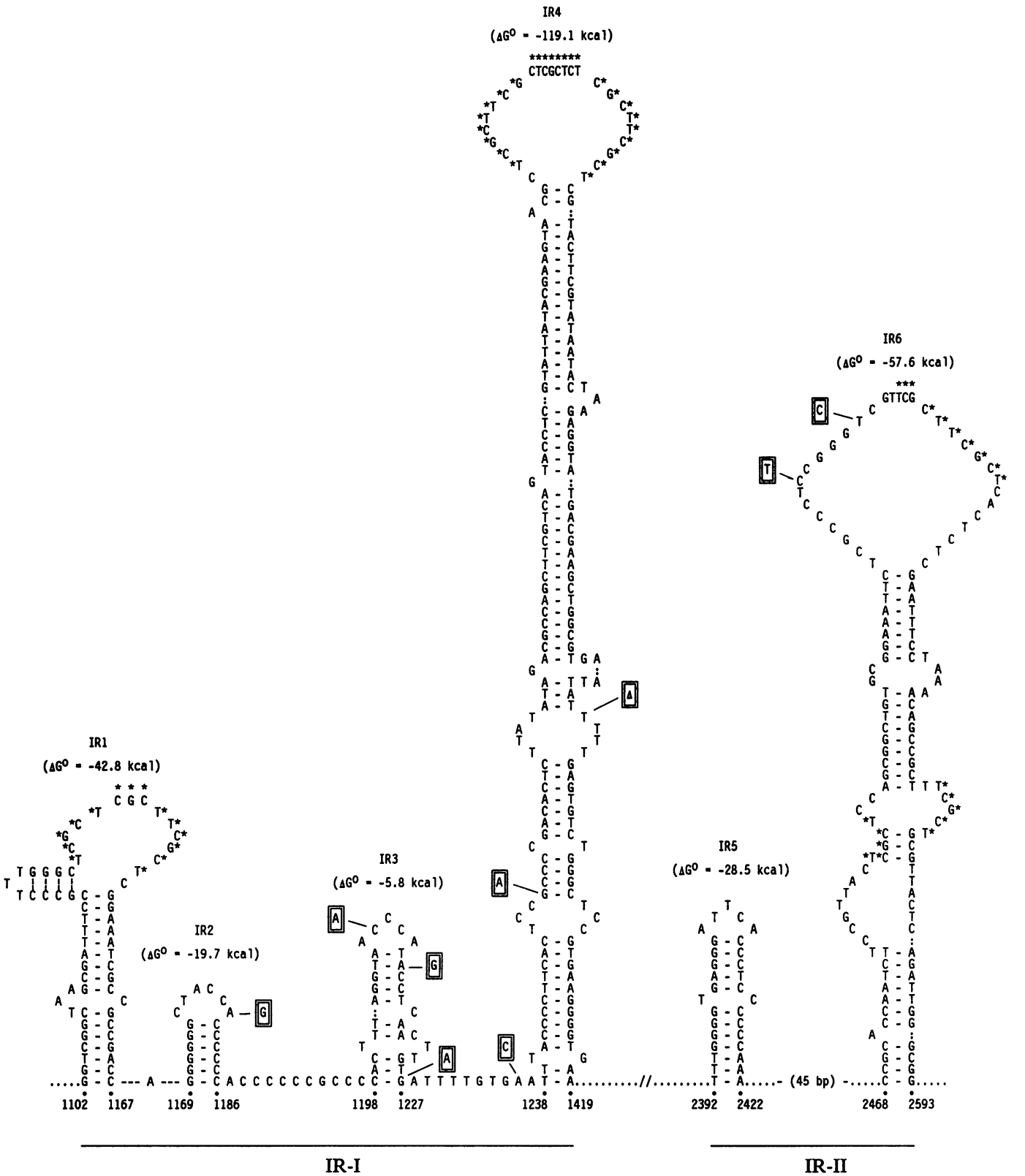

Figure 2. Potential secondary structures of regions IR-I and IR-II and their locations in pFV1. The free energy (ΔG°) values of the potential hairpin loops IR1 to IR6 were calculated using the PCGENE program RNAFOLD. Nucleotides different with the corresponding regions of pFZ1 are indicated. The repeated pentamers 5'-TCGCT are marked by asterisks.

acids (Figure 1; Table 1). Except for ORF8 of pFV1 and the homologous ORF8′ of pFZ1, all ORFs located in the backbone structure have the same orientation. Each plasmid specifies components of different R/M systems (10, 11) which are encoded by genes located on the plasmid specific elements C (pFV1) and E (pFZ1) (Figure 1). The deduced amino acid sequences of the other ORFs of pFV1 and pFZ1, listed in Table 1, were analyzed for similarity with sequences present in the SwissProt data base (release 22.0).

The homology search showed that ORF1 of plasmid pFV1 and its counterpart ORF1′ of plasmid pFZ1 contain a stretch of amino acids with similarity to the purine nucleotide binding site of ATP- or GTP-utilizing proteins (Table 2). This NTP-binding site is formed by a characteristic amino acid motif (NTP-motif) composed of two elements A and B (26, 27) and—analogous to several NTP-binding enzymes involved in DNA-metabolism (27–30)—is located within the N-terminal part of the ORF1 and ORF1′ proteins (Table 2). Together with the high conservation of the deduced ORF1 and ORF1′ proteins (>99% identical residues; Table 1), the presence and location of this NTP-binding motif strongly suggests that both proteins are involved in the replication of pFV1 and pFZ1. Transcription and translation of ORF1 and ORF1′ may be initiated at sequence motifs that were found preceding the coding regions. A sequence stretch with similarity to a typical archaeal promoter signal (31, 32) was identified ranging from position 13228 to 13258 in pFV1 and from position 10729 to 10759 in pFZ1. The first translation initiation codon (ATG) is preceded by a sequence with similarity to a consensus methanogenic ribosome binding site (33), ranging from position 13294 to 13300 in pFV1 and from position 10795 to 10801 in pFZ1. However, no sequences resembling the consensus archaeal terminator signal (33) have been recognized downstream of ORF1 and ORF1′.

Within the pFV1-specific DNA element C which contains the genes for the *Mth*TI R/M system, two other ORFs, designated ORF9 and ORF10, were identified that could code for proteins of 146 and 221 amino acids, respectively (Figure 1; Table 1). A database search revealed that the protein deduced from ORF10 shares significant sequence similarity with two *E.coli* proteins, MutY and endonuclease III (52.4% and 50.0% similarity and 28.8% and 28.4% identity), respectively, which are known to be involved in DNA mismatch repair. MutY and endonuclease III, composed of 350 (34) and 211 amino acids (35), respectively, both possess N-glycosylase activity as well as apurinic/apyrimidinic (AP) endonuclease activity (36–39). They differ, however, in their substrate specificity. While MutY is involved in the correction of A/G (35), A/C (37) and A/8-oxoG

Table 2. Alignment of the NTP-motif containing domain of ORF1 with NTP-binding proteins involved in DNA metabolism. Identical or similar residues are boxed.

Protein		Sequence A		Sequence B	Reference
RecD	166	SV I S G GP G T GKT TTVAKL L	-79-	V L V VDE ASMID	(28)
UvrD	24	LL V L A GA G S GKT RVLVHR I	-171-	N I L VDE FQNTN	(29)
RuvB	57	LL I F G PP G L GKT TLANIV A	-32-	V L F IDE IHRLS	(30)
ORF1	63	LL I V G PP G S GKT VTTKYV I	-54-	I I V MDE IDKTL	this study

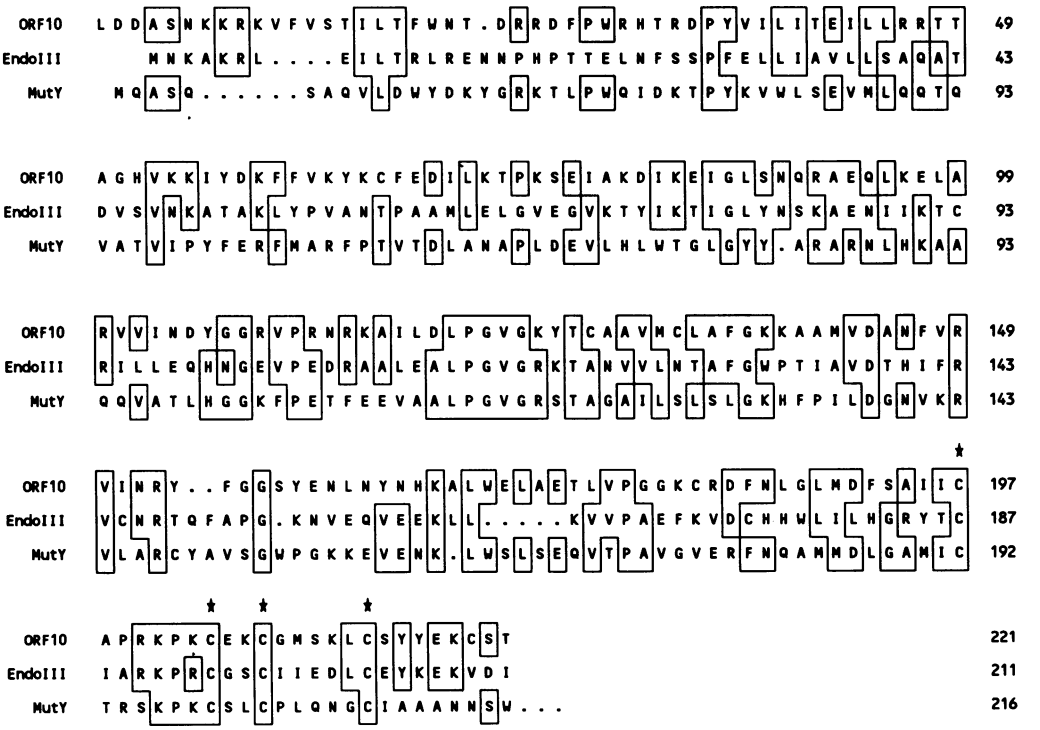

Figure 3. Amino acid sequence alignment of the putative ORF10 protein of pFV1 with endonuclease III and MutY of *E.coli*. Identical amino acids are boxed. Dots represent gaps that have been introduced to optimize the alignment. Positions marked with asterisks represent the four highly conserved cysteine residues. Numbers at the right refer to amino acid residues.

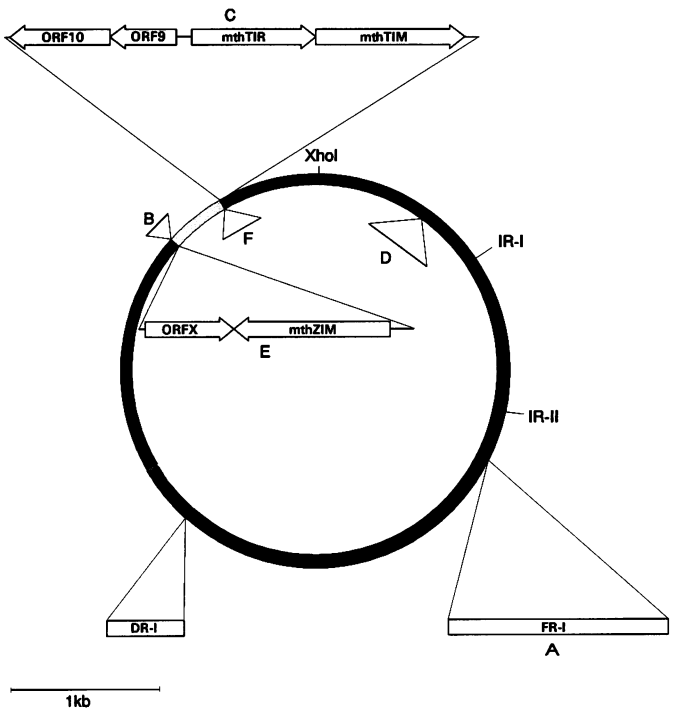

Figure 4. Schematic representation of the conserved 8.2 kb backbone structure obtained from nucleotide sequence alignment of pFV1 and pFZ1. Insertions indicated outside the circle represent pFV1-specific sequences, those inside are specific for pFZ1 (the bar represents a size of 1 kb).

mismatches (8-oxoG = 8-hydroxyguanine) (40), endonuclease III is active on mismatches generated by certain derivatives of thymidine and cytosine (38, 39). Alignment of the deduced amino acid sequences of the ORF10 protein, endonuclease III and MutY (Figure 3) revealed (i) a highly conserved motif LPGVG(R/K)XT (position 120 to 127 of the ORF10 protein; X=variable amino acid), and (ii) a stretch of four cysteine residues spaced by an identical number of amino acids ($CX_6CX_2CX_5C$), which is located at corresponding positions in the three polypeptides. Since endonuclease III was shown to be an iron-sulfur protein (41), those cysteine residues are believed to be involved in the $[4Fe\text{-}4S]^{2+}$ cluster binding (41). It therefore seems to be likely that MutY, as already suggested by Michaels and coworkers (34), and the putative ORF10 protein specified by plasmid pFV1 are iron-sulfur proteins with $[4Fe\text{-}4S]^{2+}$ clusters. The observed sequence similarities suggest that the ORF10 protein acts, like MutY and endonuclease III, as a DNA repair enzyme that recognizes incorrect base pairing.

The presence of such a repair system in the thermophilic *M.thermoformicicum* strain THF is plausible since the pFV1-encoded GGCC-recognizing methyltransferase M.*Mth*TI generates 5-methylcytosine (m^5C) (10) which is, especially under thermophilic conditions, subject to deamination (m^5C T) resulting in a G-T mismatch (42). To avoid the probably lethal effect of multiple transitions, strain THF should contain a mechanism that either prevents deamination of m^5C or is capable to correct the G-T mismatches. The pFV1-encoded ORF10 protein is a good candidate to act as a G-T mismatch repair enzyme in *M.thermoformicicum* THF. Moreover, the close proximity of ORF10 and the *Mth*TI system would support a functional relation of the endonuclease, the methyltransferase and the putative DNA mismatch repair enzyme. Such a clustering may also facilitate the simultaneous transfer of these genes between cells and ensures the coordinate activity of the encoded enzymes in the recipient which probably is necessary for its survival. Interestingly, a similar genetic linkage has been found for the functionally related methyltransferase and DNA mismatch repair enzyme in the *dcm-vsr* system of *E.coli* (42).

It seems paradoxical that *M.thermoformicicum* THF maintains a non-chromosomally encoded R/M system with a m^5C-producing MTase activity and, as a consequence, a DNA mismatch repair enzyme. A possible explanation is that, under thermophilic conditions, a combination of the enzymes specified by the GGCC-recognizing *Mth*TI R/M system and ORF10 may serve as a protection mechanism against infection by phages that exhibit the same methylation pattern as produced by the MTase M.*Mth*TI, i.e. $GG^{me}CC$. Such phages may have escaped the restriction barrier of strain THF or have acquired a protecting MTase gene. Since their methylated DNA is resistant against digestion by the host-encoded ENase, such phages normally would be able to infect the entire population. However, as a result of elevated deamination of m^5C at high temperatures without mispair correction, the genome of those phage particles accumulate, in contrast to the host genome, partially and fully deaminated $GG^{me}CC$-sites the latter of which would result in two juxtaposed G-T mismatches per site. One may envisage several scenarios once this phage DNA has entered the host cell. Firstly, the ORF10 protein could be unable to initiate mispair correction at juxtaposed G-T mismatches which may decrease the infectivity of the damaged phage DNA. The second scenario is based on the observation that the *mut*HLS-dependent mismatch repair systems of certain *E.coli* mutants produce double-strand breaks due to initiation of mismatch correction along with an endonucleolytic attack on either DNA strand (44, 45). Double-strand breaks in the deaminated phage DNA may be generated during initiation of mismatch correction by the putative ORF10 protein which—analogous to MutY and endonuclease III—may act as N-glycosylase and AP-endonuclease on the opposite strands of the juxtaposed G-T mismatch base pairs. This would generate a double-stranded break 3′ from the GG-dinucleotide of the GGCC site resulting in a similar effect as restriction of the unmethylated phage DNA by the ENase of the R/M system *Mth*TI.

CONCLUSION

Sequence analysis of the plasmids pFV1 and pFZ1 from the archaeon *M.thermoformicicum* strain THF and Z-245, respectively, has confirmed their close relationship. Both plasmids display a modular genome organization and consist of a highly conserved, approximately 8.2 kb backbone which is interspersed with plasmid-specific, accessory elements of variable size and function (Figure 4).

The high sequence similarity and comparable genetic maps of the plasmids pFV1 and pFZ1 indicate either the presence of a common ancestor or that one of the plasmids originated from the other. In either case, accessory elements have been acquired or lost.

The conserved backbone structure is expected to specify function necessary for plasmid replication and maintenance. These essential plasmid functions may include two regions, IR-I and IR-II (Figures 2 and 4), with the capacity to form highly conserved secondary structures, and, in close proximity to those

IR-regions, ORF1 that could encode a protein with NTP-binding activity (Figure 1, Table 2).

While the nature of the smaller elements specific for pFV1 and pFZ1 remains unknown, element A of plasmid pFV1, which corresponds to FR-I, has various features of an insertion sequence (9). In addition, each plasmid contains, analogous to the functional modules observed for lambdoid phages (46), at comparable locations an element which specifies components of a R/M system recognizing either the sequence GGCC (pFV1) (10) or CTAG (pFZ1) (11). Within the R/M module of plasmid pFV1, two additional ORFs were identified, one of which, ORF10, may code for a DNA-repair enzyme that is functionally related to the GGCC-recognizing R/M genes.

REFERENCES

1. Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Proc. Natl. Acad. Sci. USA **287**, 4576–4579.
2. Zillig, W. (1991) Current Opinion in Genetics and Development **1**, 544–551.
3. Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.-D. and Zillig, W. (1991) Virology **185**, 242–250.
4. Hackett, N.R., and DasSarma, S. (1989) Can. J. Microbiol. **35**, 86–91.
5. Hall, M.J. and Hackett, N.R. (1989) Nucleic Acids Res. **17**, 10501.
6. Hackett, N.R., Krebs, M.P., DasSarma, S., Goebel, W., RajBhandary, U.L. and Khorana, H.G. (1990) Nucleic Acids Res. **18**, 3408.
7. Bokranz, M., Klein, A. and Meile, L. (1990) Nucleic Acids Res. **18**, 363.
8. Nölling, J., Frijlink, M. and de Vos, W.M. (1991) J. Gen. Microbiol. **137**, 1981–1986.
9. Nölling, J., van Eeden, F.J.M., and de Vos, W.M. submitted for publication.
10. Nölling, J. and de Vos, W.M. (1992) J. Bacteriol. **174**, 5719–5726.
11. Nölling, J. and de Vos, W.M. (1992) Nucleic Acids Res. **20**, 5047–5052.
12. Gibson, T.J. (1983) Ph.D. thesis. University of Cambridge. Cambridge, United Kingdom.
13. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103–119.
14. Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Proc. Natl. Acad. Sci. USA **77**, 5172.
15. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. Nölling, J., Hahn, D., Ludwig, W., and de Vos, W.M. submitted for publication.
17. Crowe, J.S., Cooper, H.J., Smith, M.A., Sims, M.J., Parker, D. and Gewert, D. (1991) Nucleic Acids Res. **19**, 184.
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
19. Innis, M., (1988) Proc. Natl. Acad. Sci. USA **85**, 9436–9440.
20. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. **12**, 387–395.
21. Touzel, J.P., Conway de Macario, E., Nölling, J., de Vos, W.M., Zhilina, T., and Lysenko, A.M. (1992) Int. J. Syst. Bacteriol. **42**, 408–411.
22. Gruss, A. and Ehrlich, S.D. (1989) Microbiol. Reviews **53**, 231–241.
23. Kües, U. and Stahl U. (1989) Microbiol. Reviews **53**, 491–516.
24. del Solar, G.H., Puyet, A. and Espinosa, M. (1987) Nucleic Acids Res. **15**, 5561–5580.
25. Ilyina, T. and Koonin, E.V. (1992) Nucleic Acids Res. **20**, 3279–3285.
26. Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) EMBO J. **1**, 945–951.
27. Hodgman, T.C. (1988) Nature (London) **333**, 22–23.
28. Finch, P.W., Storey, A., Brown, K., Hickson, I.D. and Emmerson, P.T. (1988) Nucleic Acids Res. **14**, 8583–8594.
29. Finch, P.W. and Emmerson, P.T. (1984) Nucleic Acids Res. **12**, 5789–5799.
30. Benson, F.E., Illing, G.T., Sharples, G.J. and Lloyd, R.G. (1988) Nucleic Acids Res. **16**, 1541–1549.
31. Hausner, W., Frey, G. and Thomm, M. (1991) J. Mol. Biol. **222**, 495–508.
32. Hain, J., Reiter, W.-D., Hüdepohl, U. and Zillig, W. (1992) Nucleic Acids Res. **20**, 5423–5428.
33. Brown, J.W., Daniels, C.J. and Reeve, J.N. (1989) Crit. Rev. Microbiol. **16**, 287–337.
34. Michaels, M.L., Pham, L., Nghiem, Y., Cruz, C. and Miller, J.H. (1990) Nucleic Acids Res. **18**, 3841–3845.
35. Asahara, H., Wistort, P.M., Bank, J.F., Bakerian, R.H. and Cunningham, R.P. (1989) Biochemistry **28**, 4444–4449.
36. Au, K.G., Clark, S., Miller, J.H., and Modrich, P. (1989) Proc. Natl. Acad. Sci. USA **86**, 8877–8881.
37. Tsai-Wu, J.-J., Liu, H.-F., and Lu, A.-L. (1992) Proc. Natl. Acad. Sci. USA **89**, 8779–8783.
38. Katcher, H.L. and Wallace, S.S. (1983) Biochemistry **22**, 4071–4081.
39. Kim, J. and Linn, S. (1988) Nucleic Acids Res. **16**, 1135–1141.
40. Michaels, M.L., Cruz, C., Grollman, A.P. and Miller, J.H. (1992) Proc. Natl. Acad. Sci. USA **89**, 7022–7025.
41. Cunningham, R.P., Asahara, H. and Bank, J.F. (1989) Biochemistry **28**, 4450–4455.
42. Ehrlich, M., Gama-Sosa, M.A., Carreira, L.H., Ljungdahl, L.G., Kuo, K.C. and Gehrke, C.W. (1985) Nucleic Acids Res. **13**, 1399–1412.
43. Sohail, A., Lieb, M., Dar, M. and Bhagwat, A.S. (1990) J. Bacteriol. **172**, 4214–4221.
44. Glickman, B.W. and Radman, M. (1980) Proc. Natl. Acad. Sci. USA **77**, 1063–1067.
45. Wang, T.V. and Smith, K.C. (1986) J. Bacteriol. **165**, 1023–1025.
46. Susskind, M.M. and Botstein, D. (1978) Microbiol. Rev. **42**, 385–413.