Identification of the CTAG-recognizing restriction-modification systems *MthZI* and *MthFI* from *Methanobacterium thermoformicum* and characterization of the plasmid-encoded *mthZIM* gene

Jörk Nölling and Willem M.de Vos*

Bacterial Genetics Group, Department of Microbiology, Wageningen Agricultural University, Hesselink van Suchtelenweg 4, 6703 CT Wageningen, The Netherlands

Received July 14, 1992; Revised and Accepted September 9, 1992

EMBL accession no. X67212

ABSTRACT

Two CTAG-recognizing restriction and modification (R/M) systems, designated *MthZI* and *MthFI*, were identified in the thermophilic archaeon *Methanobacterium thermoformicum* strains Z-245 and FTF, respectively. Further analysis revealed that the methyltransferase (MTase) genes are plasmid-located in both strains. The plasmid pFZ1-encoded *mthZIM* gene of strain Z-245 was further characterized by subcloning and expression studies in *Escherichia coli* followed by nucleotide sequence analysis. The *mthZIM* gene is 1065 bp in size and may code for a protein of 355 amino acids (M_r 42,476 Da). The deduced amino acid sequence of the *MthZI* enzyme shares substantial similarity with four distinct regions from several m⁴C- and m⁶A-MTases, and contains the TSPPY motif that is so far only found in m⁴C-MTases. Partially overlapping with the *mthZIM* gene and in reverse orientation, an additional ORF was identified with a size of 606 bp potentially coding for a protein of 202 amino acids (M_r 23,710 Da). This ORF is suggested to encode the corresponding endonuclease R.*MthZI*.

INTRODUCTION

Restriction and modification (R/M) systems are widespread among prokaryotes (1) and have been identified in about 25% of the presently examined species (2). Their major function is proposed to be the protection from contamination by DNA of foreign origin, particularly derived from bacteriophage infection (3). The most common R/M systems are those of the type II, consisting of two enzymatic activities, an endodeoxyribonuclease (ENase) and a DNA-methyltransferase (MTase), which both recognize an identical short DNA sequence (2). Genetic characterization of type II R/M systems have shown that the ENase and MTase genes most often are closely linked (2, 4).

Three types of MTases, forming different methylation products, can presently be distinguished: those generating

5-methylcytosine (m⁵C), N4-methylcytosine (m⁴C), and N6-methyladenine (m⁶A) (2). Based on their amino acid similarities, those MTases can be grouped into two families, the m⁵C- on the one and the m⁴C- and m⁶A-producing MTases on the other hand (5–9). In contrast to the MTases, ENases usually do not share significant similarities on protein sequence level, except for some isoschizomeric enzymes (10–12).

The thermophilic methanogen *Methanobacterium thermoformicum* belongs to one of the three fundamentally different groups of living organisms, the domain *Archaea*, distinguished from the *Bacteria* and the *Eucarya* by various unique characteristics (13, 14). Several strains of the species *M. thermoformicum* were found to harbor partially homologous plasmids (15). One of these plasmids, plasmid pFV1, carries the genes for the GGCC-specific *MthTI* R/M system of *M. thermoformicum* strain THF (12). In the present report, we describe the identification of two CTAG-recognizing R/M systems, designated *MthZI* and *MthFI*, in *M. thermoformicum* strains Z-245 and FTF, respectively, and the characterization of the plasmid-encoded *mthZIM* gene.

MATERIAL AND METHODS

Bacterial strains, vectors, and growth conditions

M. thermoformicum strains Z-245 (DSM 3720), FTF (DSM 3012), THF (DSM 3848), SF-4 (K. Yamamoto, Osaka City University, Japan), FF1 (15), FF3 (15) and CSM3 (16) were cultivated on H₂/CO₂ as the sole carbon and energy source as described previously (15). *Escherichia coli* strains DH5α (Life Technologies) and HB101 (17) were used for subcloning and propagation, respectively, of derivatives of plasmids pUC18 and pUC19 (18). *E. coli* TG1 (19) served as host strain for phages M13mp18 and M13mp19 (18). The *E. coli* strains were routinely grown in L-broth (20) at 37°C. 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 to the culture medium.

* To whom correspondence should be addressed

Figure 1. Restriction map and subclones of two adjacent *KpnI* fragments (6.1 kb) of plasmid pFZ1. The orientation and location of the *mthZIM* gene (M) and ORF X are indicated. The *in vivo* ENase (R) and MTase (M) activity of the subclones are indicated on the right side; (▼) denote the location of the *lacZ* promoter of the vector part.

DNA manipulations and sequence analysis

Restriction endonuclease *MaeI* was used as recommended by the supplier (Boehringer Mannheim, Germany). All other restriction enzymes, T4 DNA ligase and Klenow DNA fragment were purchased from Life Technologies Inc. (Gaithersburg, Md.). DNA manipulations in *E. coli* were performed following established procedures (20). Total DNA was isolated from *M. thermoformicum* as described previously (15). For nucleotide sequence determination, DNA fragments were isolated from agarose gels using a GeneClean kit as described by the supplier (Bio 101, La Jolla, Calif.) and subcloned into M13mp18/19. The nucleotide sequences of overlapping fragments of both strands were determined by the dideoxy chain termination method (21) with the Sequenase kit according to the recommendations of the supplier (United States Biochemical Corp. Cleveland, Ohio). [α - 32 P]dATP (110 TBq/mmol) was purchased from Amersham (UK). Computer analysis of the sequence data was done with the GCG package version 6.0 (22) and the CAOS/CAMM facilities at Nijmegen, The Netherlands.

MthZIM plasmid constructions

Plasmids pUZ5 and pUZ8 (Fig. 1) were derivatives of pUC19 which contain *KpnI*-fragments of 2.1 kb and 3.7 kb, respectively, derived from plasmid pFZ1 from *M. thermoformicum* strain Z-245 (15). Plasmid pUZ8ΔNsi (Fig. 1) was constructed by digestion of pUZ8 with *NsiI* followed by treatment for 30 min with Klenow DNA polymerase in the presence of all four dNTPs (2 mM each) to remove 3'-overhanging ends. After ligation the DNA was used to transform *E. coli* HB101 cells. The resulting frameshift mutation in plasmid pUZ8ΔNsi was confirmed by restriction and sequence analysis (data not shown).

Several attempts failed to directly clone the 3.6 kb *SstI*-*PstI* fragment from plasmid pFZ1 (Fig. 1) into pUC18/19. Therefore, a plasmid that carries this fragment containing the *mthZIM* gene

and open reading frame X (ORF X) was constructed on the basis of cloned pFZ1 DNA. For this purpose, pUZ5 and a derivative of pUZ8, which contained the 3.7 kb *KpnI*-insert in opposite orientation relative to the pUC19-located *lacZ*-promoter, were used. For some manipulations, sites present in the polylinker of the vector parts were employed. First, a 2.3 kb *SstI*-*HindIII* fragment of plasmid pUZ8 was cloned into pUC18 digested with *SstI* and *HindIII* to yield plasmid pUZ12 (Figure 1). A short fragment of the polylinker located at the 5'-site of the truncated ORF X was removed from plasmid pUZ12 by digestion with *SmaI* and *PstI*. Subsequently, a 1.2 kb *SmaI*-*PstI* fragment containing the 5'-part of the ORF X gene was isolated from plasmid pUZ5 and inserted into the digested pUZ12 resulting in plasmid pUZ15. Due to this cloning strategy, which was chosen to avoid a possible lethal effect of transforming *E. coli* with a plasmid carrying an intact ORF X, plasmid pUZ15 harbors an ORF X that contains a frameshift mutation caused by a 14 bp *KpnI*-*SmaI*-*KpnI* polylinker fragment. Digestion of plasmid pUZ15 with *KpnI* removed the 14 bp fragment and religation restored wild-type sequence in the pFZ1-derived insert DNA, resulting in plasmid pUZ17 (Figure 1). The integrity of the pFZ1 DNA at this location in pUZ17 was confirmed by sequence analysis (data not shown).

Preparation of cell-free extracts

Approximately 0.1 g frozen cells of *M. thermoformicum* strain Z-245 or FTF were resuspended in 0.5 ml lysis buffer (0.1 M NaCl, 0.01 M EDTA, 0.01 M β -mercaptoethanol and 0.05 M Tris-HCl pH 7.6) and disrupted by sonication. After centrifugation for 2 min at 15000×g, the supernatant, which contained about 3.5 to 5.0 μ g protein/ μ l, was collected, mixed with 1 vol glycerol and stored at -20°C before use. The extracts were found to be stable for at least 12 month without loss of the R/M activities.

Analysis of the *MthZI* and *MthFI* ENase and MTase activities

The *in vitro* activities of the *MthZI* and *MthFI* systems were assayed essentially as described previously (12) with extracts prepared from *M. thermoformicum* strains Z-245 and FTF. The buffer used under ENase conditions contained 0.1 M NaCl, 0.2 M KCl, 0.01 M MgCl₂, 0.001 M dithioerythritol and 0.05 M Tris-HCl pH 7.6. As methyl-group donor 0.8 mM S-adenosylmethione (AdoMet) was used under MTase conditions. All assays were performed at 55°C. When specified, the ENase assays were deproteinized prior to agarose gel electrophoresis by incubation with proteinase K (0.1 mg/ml; Boehringer GmbH, Mannheim, Germany) in the presence of 0.5% sodium dodecyl sulfate for 15 min, followed by extraction with phenol/chloroform. Subsequently, the DNA was precipitated with ethanol and resuspended in ENase buffer. As substrate for the R/M enzymes, phage λ DNA and plasmid pUZ3, consisting of a 6.5 kb *SmaI*-fragment of plasmid pFZ1 from *M. thermoformicum* strain Z-245 (15) cloned into pUC19, were used.

Primer extension

Primer extension was carried out as described previously (12). DNA from plasmid pUV1 that consists of a completely sequenced, 7.4 kb *XhoI*-*SstI* fragment of plasmid pFV1 from *M. thermoformicum* strain THF cloned into pUC19 (12, 16), was digested with *R.MthZI* or *R.MaeI* and used a template for extension of a primer (5'-CTGTAGTTCAGGATC) that binds to position 375 to 361 of the insert of pUV1.

RESULTS

Identification of the *MthZI* and *MthFI* R/M systems

Following the characterization of the plasmid-encoded restriction-modification system *MthTI* from *M. thermoformicum* strain THF (12) we also analyzed other thermophilic members of the genus *Methanobacterium* for the presence of ENase activity (Table 1). Extracts prepared from *M. thermoformicum* strains Z-245 and FTF, incubated with phage λ- or plasmid pUZ3 DNA under ENase conditions, were found to produce discrete DNA fragments indicating the presence of site-specific ENase activity (Figure 2A). Identical DNA patterns were derived from ENase activity of strains Z-245 and FTF, which was designated *R.MthZI* and *R.MthFI*, respectively. In addition, *in vitro* methylation assays showed that the cell-free extracts of both strains exhibit AdoMet-dependent MTase activity that generated at least partially modified pUZ3 DNA (Figure 2A, lanes 9–12). Since this DNA methylation affects the ENase activities of strain Z-245 and FTF, it follows that the DNA is modified by the corresponding MTases *M.MthZI* and *M.MthFI*, respectively. To verify the *in vivo* activity of both MTases, total DNA isolated from *M. thermoformicum* strain Z-245 or FTF was incubated with extracts prepared from either strain. Figure 2B illustrates that DNA from strains Z-245 and FTF was resistant to digestion with *R.MthZI* (lanes 2, 3) but not with *R.KpnI* which was used as a control (lanes 4, 5). Similar results were observed using *R.MthFI* (data not shown). The resistance of DNA from strain Z-245 and FTF to digestion with ENase from either strain confirmed the *in vivo* activity of *M.MthZI* and *M.MthFI*. In contrast, DNA isolated from strain THF, which carries the GGCC-recognizing *MthTI* system (12), was sensitive to *R.MthZI*

Table 1. Distribution of plasmid DNA and ENase activity in *M. thermoformicum*

Strain	Plasmid (kb)	Recognition sequence of ENase activity	Reference
Z-245	pFZ1 (11.0)	CTAG	This work
FTF	pFZ2 (11.0)	CTAG	This work
THF	pFV1 (13.5)	GGCC	(12)
FF1, FF3, CSM3, SF-4	—	—	This work

(Figure 2B, lane 1). These results indicate that *M. thermoformicum* strains Z-245 and FTF each harbor a R/M system, designated *MthZI* and *MthFI*, respectively, with identical sequence specificity.

Sequence specificity of *R.MthZI* and *R.MthFI*

In order to determine the recognition sequence of *R.MthZI* and *R.MthFI* we compared the restriction patterns produced by these ENases with those produced by several commercially available ENases. We identified the CTAG-recognizing ENase *R.MaeI* as an isoschizomer of *R.MthZI* and *R.MthFI* since it generates restriction patterns identical to that produced by the *M. thermoformicum* enzymes using phage λ- or plasmid pUZ3 DNA as substrate (Figure 2A, lanes 2–4 and 6–8). For some of the restriction fragments generated by digestion with extracts from the *M. thermoformicum* strains, slightly aberrant mobilities were observed if samples were not deproteinized before gel electrophoresis, i.e. large fragments show an increased and small fragments a decreased mobility (see Figure 2A, lanes 1 and 5 for results with Z-245 extract). This most likely is due to components in the extracts with similar properties as the histone-like DNA binding protein HMt from *M. thermoautotrophicum* strain ΔH (23).

Since both R/M systems *MthZI* and *MthFI* recognize the same target sequence, *MthZI* from strain Z-245 was used for a more detailed analysis. The exact cleavage position of *R.MthZI* was determined by primer extension experiments. The extension products of *R.MthZI* and *R.MaeI* digested DNA had identical sizes (Figure 3). Considering that the 5'-overhanging ends generated by both ENases were filled up by Klenow DNA polymerase and therefore the 3'-nucleotide of the extension products represent the 5'-nucleotide of the cleavage sites, comparison of the extension products with the parallel run sequence ladder revealed that both ENases cleave the target sequence to yield C↓TAG.

Identification of the *mthZIM* gene

M. thermoformicum strain Z-245 and FTF harbor identical or nearly identical plasmids, pFZ1 and pFZ2, respectively, which are partially homologous to plasmid pFV1 isolated from strain THF (15). Since the R/M system *MthTI* of strain THF is located on plasmid pFV1 (12) and no ENase activity was detected in extracts prepared from other *M. thermoformicum* strains that lack plasmid DNA (Table 1), we anticipated that the R/M system *MthZI* of strain Z-245 could be located on plasmid pFZ1. To study this possibility, cloned plasmid pFZ1 DNA was tested for *in vivo* modification by incubation with Z-245 extract and subsequent analysis on agarose gel. All plasmids containing pFZ1 DNA were sensitive to *R.MthZI* except one, pUZ8, which was only partially digested with *R.MthZI* despite the presence of

multiple *R.MthZI* sites in at least the vector part of pUZ8. The protection against digestion with *R.MthZI* was even higher when *E.coli* HB101 was used to propagate plasmid pUZ8 (Figure 2B, lane 7). In addition, we tested pUZ8 DNA for sensitivity to *R.XbaI* since its target sequence TCTAGA includes the CTAG site recognized by *R.MthZI*. Sequence analysis (see below) showed that pUZ8 contains two *R.XbaI* restriction sites, one in the insert and one in the vector part. Plasmid pUZ8 was—in contrast to other plasmids tested—partially resistant to digestion with *R.XbaI* (Figure 2B, lane 11). In contrast, plasmid pUZ8 Δ Nsi, which contains a frameshift mutation, was found to be sensitive to digestion with both *R.MthZI* and *R.XbaI* (Figure 2B, lanes 8, 12). These results indicate that pUZ8 encodes a MTase activity which modifies the *R.MthZI* and the *R.XbaI* recognition sites. Subcloning of pUZ8 DNA followed by *R.MthZI* digestion allowed the location of the *mthZIM* gene on a 2.3 kb *SstI-KpnI* fragment, contained in plasmid pUZ12 (Figure 1). Although the *in vivo* modification of DNA containing the *mthZIM* gene clearly showed that this gene is functionally expressed in *E.coli*, no activity of *M.MthZI* was observed in *in vitro* assays performed as described earlier (12).

Sequence analysis (see below) revealed that the insert of pUZ12 contained, in close proximity to the *mthZIM* gene, part of an open

reading frame (ORF X) which is a good candidate for the corresponding *mthZIR* gene. To exclude the possibility that the ENase gene is truncated in pUZ12 we constructed plasmid pUZ17 that contain the *mthZIM* gene and the entire ORF X (Figure 1). However, neither *in vitro* using cell-free extracts nor *in vivo* in restriction assays using bacteriophage λ , ENase activity was observed with *E.coli* HB101 harboring either plasmid pUZ17 or pUZ12 (data not shown).

Nucleotide sequence of the *mthZIM* gene

The complete nucleotide sequence of the 3561 bp *SstI-PstI* insert of plasmid pUZ17 has been determined. Analysis of the sequence revealed two large, converging and partially overlapping ORFs. The first ORF contains an ATG start codon at position 779 which is preceded by a potential ribosome binding site (RBS; position 770 through 775), and a TAG stop codon at position 1844. It could code for a polypeptide of 355 amino acids (M_r 42,476 Da). Since plasmids containing this ORF were resistant to digestion with *R.MthZI* but became sensitive to this ENase when a frameshift mutation was introduced (pUZ8 Δ Nsi), we conclude that it represents the *mthZIM* gene. The second ORF, designated ORF X, is located on the complementary strand and ranges from position 2488 to the termination codon TAA at position 1837.

Figure 2. *In vitro* (A) and *in vivo* (B) activities of the *MthZI* and *MthFI* R/M systems. (A) Lanes 1–8, ENase assays using extracts prepared from strain Z-245 (lanes 1, 5, not purified; lanes 2, 6, deproteinized), FTF (lanes 3, 7, deproteinized) and 5 Units *R.MaeI* (lanes 4, 8) with phage λ DNA (lanes 1–4) and pUZ3 DNA (lanes 5–8) as substrate; lanes 9–12, pUZ3 DNA incubated under MTase assay conditions with Z-245-extract (lanes 9, 11) and FTF-extract (lanes 10, 12), with (lanes 9, 10) and without AdoMet (lanes 11, 12) followed by digestion with *R.MthZI* (lanes 9, 11) and *R.MthFI* (lanes 10, 12). Lane 13, undigested plasmid pUZ3 DNA (1 μ g). Lane s, size marker phage λ DNA (0.7 μ g) digested with *R.HindIII*. (B) Lanes 1–5, about 0.5 μ g total DNA isolated from *M.thermoformicicum* strain THF (lane 1), Z-245 (lanes 2, 4) and FTF (lanes 3, 5) incubated under ENase conditions with extract prepared from strain Z-245 (lanes 1–3) and *R.KpnI* (lanes 4, 5). Lanes 7–14, about 1.0 μ g plasmid pUZ8 (lanes 7, 9, 11, 13) and pUZ8 Δ Nsi DNA (lanes 8, 10, 12, 14) incubated under ENase conditions with extract prepared from strain Z-245 (lanes 7, 8), *R.MaeI* (lanes 9, 10), *R.XbaI* (lanes 11, 12), and *R.KpnI* (lanes 13, 14). Lane 6, undigested plasmid pUZ8 DNA (1.0 μ g). Lane s, size marker phage λ DNA (0.7 μ g) digested with *R.HindIII*.

A potential start codon (GTG at position 2443) is preceded by a sequence that may serve as RBS (position 2454 through 2449). If the GTG is used, the ORF X encodes a protein of 202 amino acids with a calculated molecular mass of 23,710 Da.

The regions flanking the *methZIM* gene and ORF X were analyzed for archaeal transcription signals. Within an AT-rich 250 bp region preceding the *methZIM* gene, that contains several direct repeats, various DNA stretches with similarities to the essential TATA box of archaeal promoters could be identified (24, 25). Inspection of the upstream region of the ORF X revealed no obvious promoter-like sequences. Downstream of *methZIM* and ORF X, a conserved archaeal transcription termination signal consisting of a T-rich stretch (26) was identified.

Protein sequence comparisons

The deduced amino acid sequence of the *M.MthZI* protein was compared to that of known sequences of other MTases. No significant homology was observed with MTases that produce m⁵C. *M.MthZI* shares, however, similarities with distinct

regions from several m⁴C- and m⁶A-MTases. The MTases showing the highest similarity with *M.MthZI* include *M.BamH2* (27), *M.BamHI* (9), *M.SmaI* (28) and *M.Cfr9I* (7), all of which were proven or assumed to be m⁴C-MTases, and two representatives of the m⁶A-MTase family, *M.HinfI* (29) and *M.RsrI* (30). Four blocks with conserved amino acid motifs, similar to those reported by Brooks et al. (1991), could be identified (Figure 4). The longest conserved segment, block D, contains the FXGXG-motif (motif DII in Figure 4; X = variable amino acid) common to almost all MTases (6–8). Block B contains the (S/D)PP(Y/F) motif which is present in all m⁶A- and m⁴C-MTases (7, 9, 31). Block A is less conserved in *M.MthZI* but is, similar to other m⁶A- and m⁴C-MTases, located at the N-terminus and contains the characteristic DXXE motif.

In addition to the conserved amino acid motifs identified by Brooks et al. (1991), the protein sequence comparison data revealed two other amino acid regions, which were found to be conserved in all m⁴C-MTases (7, 27, 28, 32), except for *M.MvaI* (7), and in several m⁶A-MTases such as *M.HinfI* (29; Figure 4). Both regions are flanking the FXGXG motif (DII). One region, DI, contains the motif (P/S)XX(L/M)Y (Y = hydrophobic, non aromatic) and is located 22 amino acids N-terminal from the sequence FXGXG, whereas motif DIII, RXX(I/V)(G/S)XEX₁₁(R/K), is located 11 amino acids C-terminal from the FXGXG sequence. This latter motif contains three highly conserved charged residues.

Comparison of the amino acid sequence deduced from ORF X revealed no significant homology with other protein sequences found in the data base.

DISCUSSION

This report describes the identification of two novel R/M systems with the same sequence specificity, *MthZI* and *MthFI*, from the archaeon *M.thermoformicum*. In addition, the characterization of the first CTAG-recognizing methyltransferase, *M.MthZI*, is presented. Both *MthZI* and *MthFI* R/M systems are produced by two plasmid-harboring strains, Z-245 and FTF, respectively. Further analysis revealed that the *methZIM* gene is located on the 11 kb plasmid pFZ1 of strain Z-245 (15). Since the plasmids

Figure 3. Primer extension analysis of the *R.MthZI* cleavage-site. Plasmid pUV1 DNA, digested with *R.MaeI* (lane 1) and *R.MthZI* (lane 2), respectively, was used as template for extension of the primer. Arrows indicate the major extension product and its corresponding nucleotide derived from incubation with both ENases. A shorter extension product attributed to aspecific termination is also present. A DNA sequence ladder run in parallel is shown at the right.

A				B				C			
<i>M.MthZI</i>	KNSADMNELKDKS	22	INLVV	TSPPY	PMVE	36	EVDRV	TAPGGVVIINIGD	93		
<i>M.BamH2</i>	IHNH*CVQFMKEN	18	*D*T*	*****	DDLRL	23	*LY***KE****VWVV**		54		
<i>M.BamHI</i>	LYNG*CL**FKQV	133	VDTIF	AD***F	NNLD	151	*CI**LK***SLF*YINP		193		
<i>M.SmaI</i>	LDLF*QT*ECLE*	20	FDC**	*****	WGLR	57	D*R*TLKDD*TLWL****		101		
<i>M.PvuII</i>	MYIG*SL**LESF	42	*S***M	*****F	ALQR	60	V*NKKLK*DSFVDF*G		99		
<i>M.HinfI</i>	ILKG*CI*KLKTI	25	*D*IF	AD***	F*QT	43	*CK*ILKST*SIWVIGSF		98		

D							
DI		DII		DIII			
<i>M.MthZI</i>	PFELA	YRLINMYSIMGDWVLDP	FLGTG	TTMIAACAG	RNSIGYELDHNFKDLIESR	I	275
<i>M.BamH2</i>	*EK**	EDM*LSW*NE**I*F**	*M*S*	*AKM**LNN	*KY**T*ISKEYC*IANE*	L	244
<i>M.BamHI</i>	SVK*L	D*I*T*STNE**V****	*G*S*	*FAVSEML*	*KW**F**GNCEI*IKERLK	N	361
<i>M.SmaI</i>	*R*M*	RLCVLAG*RP*GK****	*F*S*	*GVVCQELD	*ECV*I**NEEYAS*AKE*	*	289
<i>M.PvuII</i>	*AK*P	EFF*R*LTEPD*L*V*I	*G*SN	*GLV*ERES	*KW*SF*MKPEYVAASAF*	F	308
<i>M.HinfI</i>	*ES*L	*KV*LS*SKPN*V****	*F***	*GAV*KAL*	*Y**I*REQYI*VA*K*	L	255

Figure 4. Local amino acid sequence alignment of *M.MthZI* and other DNA methyltransferases. Amino acid sequences of four conserved regions (A–D) from *M.MthZI*, *M.BamH2*, *M.BamHI*, *M.SmaI*, *M.PvuII* and *M.HinfI* are shown. Region D is subdivided into three conserved amino acid blocks (DI–DIII). Residues identical to those in *M.MthZI* are denoted by asterisks. Amino acid motifs mentioned in the text are blocked. Numbers at the end of each line indicate the position of the last amino acid in the region.

pFZ1 and pFZ2 of the strains Z-245 and FTF are highly similar (15), it is very likely that the *M. MthFI* enzyme is also encoded by plasmid pFZ2. Indeed, pFZ2 DNA showed strong hybridization with the *mthZIM* gene as probe (data not shown). In addition to the GGCC-specific, plasmid-located *MthTI* system of strain THF (12), the *MthZI* and *MthFI* systems represent another class of plasmid-encoded R/M systems in *M. thermoformicum*. Since no ENase or MTase activity could be detected in other *M. thermoformicum* isolates that appeared to be plasmid-free (Table 1), this would suggest that the occurrence of R/M systems in these methanogenic *Archaea* depends on the presence of extrachromosomal DNA elements.

Extracts prepared from *M. thermoformicum* strains Z-245 and FTF contained high ENase and low MTase activities. The cloned *mthZIM* gene could be functionally expressed in *E. coli* as shown by *in vivo* modification of vector DNA carrying the complete *mthZIM* gene. No activity was, however, detected in *in vitro* MTase assays using extracts prepared from different *E. coli* strains with the *mthZIM* gene in either orientation with respect to the *lacZ* promoter of the vector. Since the *in vitro* MTase conditions used worked adequately with extracts prepared from the methanogenic hosts, a low expression level of the *mthZIM* gene or instability of the *M. MthZI* protein in *E. coli* could explain the inability to detect *in vitro* activity.

Comparison of the deduced amino acid sequence of the *M. MthZI* protein revealed four regions with significant similarity to m⁴C- and m⁶A-MTases but not to m⁵C-MTases (Figure 4). Sequence comparisons did not reveal motifs that would allow a clear discrimination between m⁴C- and m⁶A-MTases. However, the deduced protein sequence of *M. MthZI* contains the TSPPY motif which so far has only been found in m⁴C producing MTases (7, 28, 29, 32).

The recognition sequence CTAG of *MthZI* and *MthFI* forms the core of the sequence TCTAGA, the target of the *XbaI* R/M system. We observed that DNA is protected to a similar extent against cleavage by either *R. XbaI* or *R. MthZI* when it is methylated by *M. MthZI*. If *M. MthZI* is a m⁴C-MTase, it would modify the *XbaI*-site to yield T^{m4}CTAGA. Alternatively, if *M. MthZI* generates m⁶A, it would modify the *XbaI*-site to yield TCT^{m6}AGA. Either modification pattern of the *XbaI*-site would be, in addition to the three already described (33, 34), a fourth one that prevent cleavage by *R. XbaI*.

Partially overlapping with the *mthZIM* gene and in opposite orientation, another ORF, ORF X, was identified. Although we have no direct evidence, ORF X could well represent the corresponding *mthZIR* gene. In favour of this possibility are the findings that (i) ENase and MTase genes most often are closely linked (2), (ii) the size of the protein encoded by ORF X is similar to that of other ENases, (iii) the *mthZIM* gene and ORF X constitute a 1.9 kb module in plasmid pFZ1 that is absent in plasmid pFV1, which in turn contains at this location a unique fragment encoding the *MthTI* R/M system of strain THF (12, 16), and (iv) no ENase activity was observed in other, closely related strains of *M. thermoformicum* that lack plasmid DNA (Table 1). The failure of detecting the expected ENase activity in *E. coli* might be ascribed to a low expression level of ORF X because of its unusual start codon GTG (35).

Among the large number of identified R/M systems (1) only four have previously been found to recognize the same sequence as *MthZI* and *MthFI* (1, 2, 36, 37) indicating that CTAG is a rarely occurring target. Two of those systems, *MaeI* and *MjaI*, were identified in the methanogens *Methanococcus aeolicus* (36) and *Methanococcus jannaschii* (1). This suggests that, together

with the here described *MthZI* and *MthFI*, R/M systems recognizing CTAG are more common in *Archaea* than in *Bacteria*.

ACKNOWLEDGEMENT

We thank Rik Eggen for helpful discussions and advise on the manuscript.

REFERENCES

- Roberts, R.J. (1990) *Nucleic Acids Res.* **18**, r2149–r2365.
- Wilson, G.G. and Murray, N.E. (1991) *Annu. Rev. Genet.* **25**, 585–627.
- Krüger, D.H. and Bickle, T.A. (1983) *Microbiol. Rev.* **47**, 345–360.
- Wilson, G.G. (1991) *Nucleic Acids Res.* **19**, 2539–2566.
- Lauster, R., Kriebardis, A. and Guschlbauer, W. (1987) *FEBS Lett.* **220**, 167–176.
- Lauster, R. (1989) *J. Mol. Biol.* **206**, 313–321.
- Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V. and Januliatas, A. (1989) *Nucleic Acids Res.* **17**, 9823–9832.
- Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) *Nucleic Acids Res.* **17**, 2421–2435.
- Brooks, J.E., Nathan, P.D., Landry, D., Sznyter, L.A., Waite-Rees, P., Ives, C.L., Moran, L.S., Slatko, B.E. and Benner, J.S. (1991) *Nucleic Acids Res.* **19**, 841–850.
- Stephenson, F.H., Ballard, B.T., Boyer, H.W., Rosenberg, J.M. and Greene, P.J. (1989) *Gene* **85**, 1–13.
- Kapfer, W., Walter, J. and Trautner, T.A. (1991) *Nucleic Acids Res.* **19**, 6457–6463.
- Nölling, J. and de Vos, W.M. (1992) *J. Bacteriol.* **174**, 5719–5726.
- Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) *Proc. Natl. Acad. Sci. USA* **287**, 4576–4579.
- Zillig, W. (1991) *Current Opinion in Genetics and Development* **1**, 544–551.
- Nölling, J., Frijlink, M. and de Vos, W.M. (1991) *J. Gen. Microbiol.* **137**, 1981–1986.
- Nölling, J. (1992) Ph.D. thesis, Wageningen Agricultural University, The Netherlands (in preparation).
- Boyer, H.W. and Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41**, 459–472.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* **33**, 103–119.
- Gibson, T.J. (1984) Ph.D. thesis. University of Cambridge. Cambridge, United Kingdom.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Musgrave, D.R., Sandman, K. and Reeve, J.N. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10397–10401.
- Reiter, W.D., Hüdepohl, U. and Zillig, W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9509–9513.
- Hausner, W., Frey, G. and Thomm, M. (1991) *J. Mol. Biol.* **222**, 495–508.
- Brown, J.W., Daniels, C.J. and Reeve, J.N. (1989) *CRC Crit. Rev. Microbiol.* **16**, 287–337.
- Connaughton, J.F., Kaloss, W.D., Vanek, P.G., Nardone, G.A. and Chirikjian, J.G. (1990) *Nucleic Acids Res.* **18**, 4002.
- Heidmann, S., Seifert, W., Kessler, C. and Domdey, H. (1989) *Nucleic Acids Res.* **17**, 9783–9796.
- Chandrasegaran, S., Lunnen, K.D., Smith, H.O. and Wilson, G.G. (1988) *Gene* **70**, 387–392.
- Kaszubska, W., Aiken, C., O'Connor, C.D. and Gumpert, R.I. (1989) *Nucleic Acids Res.* **17**, 10403–10425.
- Chatterjee, D.K., Hammond, A.W., Blakesley, R.W., Adams, S.M. and Gerard, G.F. (1991) *Nucleic Acids Res.* **19**, 6505–6509.
- Tao, T., Walter, J., Brennan, K.J., Cotterman, M.M. and Blumenthal, R.M. (1989) *Nucleic Acids Res.* **17**, 4161–4175.
- Van Cott, E.M. and Wilson, G.G. (1988) *Gene* **74**, 55–59.
- McClelland, M. and Nelson, M. (1992) *Nucleic Acids Res.* **20**, 2145–2157.
- Gold, L. and Stormo, G. (1987). In: *Escherichia coli and Salmonella typhimurium*; Neidhardt, C.(ed); American Society for Microbiology, Washington D.C.; Vol.2, 1302–1307.
- Schmid, K., Thomm, M., Laminet, A., Lane, F.G., Kessler, C., Stetter, K.O. and Schmitt, R. (1984) *Nucleic Acids Res.* **12**, 2619–2628.
- New England Biolabs, Inc. (1992) Product catalogue (2/92).