5208–5221 Nucleic Acids Research, 2009, Vol. 37, No. 15 Published online 3 July 2009
doi:10.1093/nar/gkp534

The MmeI family: type II restriction–modification enzymes that employ single-strand modification for host protection

Richard D. Morgan*, Elizabeth A. Dwinell, Tanya K. Bhatia, Elizabeth M. Lang and Yvette A. Luyten

New England Biolabs, Inc. 240 County Road Ipswich, MA 01938 USA

Received March 18, 2009; Revised and Accepted June 8, 2009

ABSTRACT

The type II restriction endonucleases form one of the largest families of biochemically-characterized proteins. These endonucleases typically share little sequence similarity, except among isoschizomers that recognize the same sequence. MmeI is an unusual type II restriction endonuclease that combines endonuclease and methyltransferase activities in a single polypeptide. MmeI cuts DNA 20 bases from its recognition sequence and modifies just one DNA strand for host protection. Using MmeI as query we have identified numerous putative genes highly similar to MmeI in database sequences. We have cloned and characterized 20 of these MmeI homologs. Each cuts DNA at the same distance as MmeI and each modifies a conserved adenine on only one DNA strand for host protection. However each enzyme recognizes a unique DNA sequence, suggesting these enzymes are undergoing rapid evolution of DNA specificity. The MmeI family thus provides a rich source of novel endonucleases while affording an opportunity to observe the evolution of DNA specificity. Because the MmeI family enzymes employ modification of only one DNA strand for host protection, unlike previously described type II systems, we propose that such single-strand modification systems be classified as a new subgroup, the type IIL enzymes, for Lone strand DNA modification.

INTRODUCTION

The type II restriction endonucleases and methyltransferases constitute a large and widely distributed family of enzymes (1). These enzymes are under strong selective pressure as a primary defense against parasitic DNA. They may also function to control the exchange of genetic material between 'self' and 'others' within microbial populations (2). As such they represent fertile ground for the study of protein evolution, particularly the evolution of those protein–DNA interactions that confer sequence specificity to these enzymes.

A restriction system must differentiate between 'self' and 'foreign' DNA in order to cut the identifiably foreign DNA but not host DNA. This discrimination is based upon modification of the DNA, usually in the form of methylation of a base in each strand within the discrete DNA sequence recognized by the endonuclease. Usually, hemi-methylated DNA is not cut by the type II endonucleases, ensuring that immediately following replication the hemi-methylated daughter chromosomes remain protected from cleavage by the endonuclease, allowing the DNA methyltransferase time to modify the newly replicated strand. Type II R–M systems most commonly have separate enzymes to accomplish host DNA modification and endonuclease cleavage of foreign DNA. However, some systems, such as some members of the type IIG subgroup, have one polypeptide that contains both DNA methyltransferase and endonuclease activity, partnered with a second, companion DNA methyltransferase (3,4). The type IIG enzymes generally recognize asymmetric sequences and cut away from the recognition sequence. The fusion proteins cut fully unmodified DNA, but can also modify a base in one DNA strand of unmodified DNA. Because the recognition site is not disrupted by the endonucleolytic cleavage, they can remain bound to and modify their recognition site even after cutting the DNA. However, previously described type IIG endonucleases, such as Eco57I, require a separate, companion DNA methyltransferase in order to achieve modification of both DNA strands, since the single-strand modification produced by the fusion protein is insufficient for host protection (5).

Since most bacteria and archaea possess one or more restriction systems, the recent availability of complete genome sequences has provided a wealth of putative restriction systems. These are typically identified *in silico*

*To whom correspondence should be addressed. Tel: 978 927 5054; Fax: 978 921 1350; Email: morgan@neb.com through amino-acid sequence similarity to conserved sequence motifs within the DNA methyltransferase. The type II DNA methyltransferases exhibit significant amino-acid sequence similarity in the regions involved in binding the methyl donor AdoMet and the catalysis of methyl transfer, which allows uncharacterized putative sequences to be reliably identified as DNA methyltransferases (6,7). However, the type II restriction endonucleases (REs) are generally not similar at the level of primary amino-acid sequence, even though many share a conserved structure (8). Type II REs typically exhibit a low level of sequence identity that falls in the so-called 'twilight zone' of sequence similarity, where genuine similarities between homologs disappear into the random noise of sequence comparison (9,10). Although significant similarity between putative and characterized endonucleases is occasionally observed, such enzymes are usually isoschizomers, i.e. they recognize the same DNA sequence and cut at the same position. For most type II REs, however, the lack of sequence conservation makes identification of these enzymes among the many putative sequences available in sequenced genomes a challenging task (11).

It was therefore surprising that after cloning the unusual type II endonuclease MmeI (12), we observed a number of highly similar putative sequences in available databases. Expression and characterization of these MmeI-like putative genes has allowed us to describe a new family of type II restriction enzymes. These enzymes employ the unusual strategy of using only single-strand methylation for host protection, a property not previously described for type II R–M systems. Modification of only one DNA strand raises the question of how host protection is maintained, since immediately following replication one of the two daughter chromosomes will be fully unmodified and would be expected to be vulnerable to restriction. However the MmeI-like systems described appear frequently in sequenced bacterial genomes, and it appears from bioinformatic analyses that such single-strand modification for host protection may be widespread in nature. The diversity of recognition sequences found in this closely related family of proteins affords a unique window into the evolution of specific protein–DNA-binding interactions.

MATERIALS AND METHODS

Restriction endonucleases, *S*-adenosyl-L-methionine (AdoMet), T4 DNA Ligase, DNA polymerases, DNA size standards and competent cells were from New England Biolabs (Ipswich, MA). Synthetic DNA oligonucleotides were from New England Biolabs, Organic Synthesis Division (Ipswich, MA) or Integrated DNA Technologies (Coralville, IA). [methyl-3H]-AdoMet was from GE Healthcare (Pittsburgh, PA). Gamma-^{33}P-ATP was from Perkin Elmer (Boston, MA). Plasmid preparation and DNA purification spin columns were from Qiagen (Valencia, CA) and Zymo Research (Orange, CA). Ultraclean^{TM} Soil DNA isolation kits were from MoBio Laboratories (Carlsbad, CA).

Endonuclease assays

Endonuclease activity was assayed by incubating various amounts of enzyme in reaction buffer (NEBuffer 4: 20 mM Tris–acetate, pH 7.9, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM DTT, supplemented with 100 µg/ml BSA and 80 µM AdoMet) containing 1 µg substrate DNA per 50 µl for one hour at 37°C. Reactions were terminated by addition of stop solution (50 mM EDTA, pH 8.0, 50% glycerol, 0.02% bromophenol blue), and reaction products were analyzed by electrophoresis in 1% LE agarose gels alongside DNA size standards lambda-HindIII and PhiX174-HaeIII, or lambda-BstEII and pBR322-MspI.

Identification of MmeI homologs in sequence databases

The MmeI protein sequence (GenBank accession no ACC85607) served as the query to identify significantly similar protein sequences using BLAST searches performed against available databases, such as the non-redundant protein sequences (nr) or the environmental samples (env-nr) amino-acid sequence databases at the National Center for Biotechnology Information (NCBI), the REBASE database of restriction systems (1), or TBLASTN searches against the Nucleotide collection (nr/nt) or environmental samples (env-nt) DNA databases at NCBI (13). Significant hits with an expectation value <*e*–20 were considered potential MmeI homologs. The putative homologs were evaluated to identify those sequences that contained the conserved PD-(D/E)xK endonuclease motif residues, aligned with the putative MmeI catalytic residues D_{70}, E_{80} and K_{82} in the amino terminal region, and that exhibited similarity throughout the entire length of MmeI. A number of the identified putative MmeI homologs were cloned and expressed in *Escherichia coli*.

Cloning and expression of MmeI homologs

The microbial strain, or the genomic DNA from the strain, encoding a MmeI homolog of interest was obtained from the source listed in Table 1. Where purified genomic DNA was not available, DNA was prepared from a freshly grown cell culture by standard techniques or directly from a lyophilized ATCC or DSM culture vial by chemical lysis and bead beating (MoBio Laboratories, CA). For environmental sequences, where neither genomic DNA nor the microbial strain was available, the putative endonuclease gene was obtained by *in vitro* synthesis.

Putative genes were PCR amplified from genomic DNA. The oligonucleotide sequences used for expression are listed in Supplementary Table T1. The amplified genes were cloned into one of several expression vectors, such as the high copy number pUC19 derivative pRRS (14) or the T7 expression based vectors pAII17 (15) or pSAPv6 (16), and transformed into an appropriate *E. coli* host, such as ER2683 (*F′proA*$^+$*B*$^+$ *lacI*^{q} *Δ(lacZM15)* (KanR) *miniTn10*/λ$^-$ *fhuA2 glnV44 e14*$^-$ *rfbD1? relA1? endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10 Δ(lacI-lacA)200*) for pRRS, or C2566, C3013 or ER3081 (F$^-$ λ- *fhuA2* *lacZ::T7 gene1 [lon] ompT gal attB::(pCD13-lysY,* $lacI^q$*) sulA11 R(mcr-73::miniTn10*–TetS*)2 [dcm] R(zgb-210::Tn10* –TetS*) endA1 Δ(mcrC-mrr)114::IS10*) for the T7 expression vectors. Transformed cells were grown and tested for endonuclease activity.

Protein purification: clones expressing endonuclease activity were grown, induced, harvested and disrupted by sonication. The expressed endonuclease was purified over a Heparin HiTrap column (GE Healthcare, Piscataway, NJ). The crude extract supernatant was applied to the column in buffer A (20 mM Tris–HCl, pH 7.5, 1 mM DTT, 0.1 mM EDTA) containing 50 mM KCl. The column was then washed with five-column volumes buffer A containing 50 mM KCl, then the enzyme was eluted with a 20–40 column volume linear gradient from 50 mM to 1M KCl in buffer A. The enzymes typically eluted from the heparin column between 0.3 and 0.4 M KCl. The fractions containing purified enzyme were used for subsequent characterization of the DNA recognition sequence and the position of DNA cleavage.

Correction of disrupted open reading frames

The putative ORFs reported in the database sequences for several of the identified MmeI homologs were less than the expected length when compared to the amino-acid sequence of MmeI or other active MmeI homologs; however *in silico* translation of all three frames of the DNA sequence adjacent to these incomplete ORFs revealed protein coding sequences similar to the entire MmeI protein sequence. The potential full-length ORF for three disrupted homologs; NmeAIII, DraRI and ApyPI, and the position and nature of the putative disruption to the ORF, was predicted by comparison of amino-acid translations of the DNA sequences with an amino-acid sequence alignment of the active MmeI family members. The putative ORFs were cloned from the start to stop positions and tested for expression; however none of the genes expressed active endonuclease. The disruption to the reading frame present in the database was confirmed by sequencing the clones, indicating the genes are in fact disrupted in the organism sequenced and not mere sequencing errors. In the case of a frame shift, the potential amino-acid sequence for all three frames in the region of the putative frame shift was compared to the aligned sequences of the active MmeI family members to identify the position where significant similarity to the aligned homologs changed from one reading frame to a different frame. The choice of amino-acid residue(s) to place at the correction point was guided by placing the amino-acid residue(s) observed in one or more of the most similar homolog sequences into the correction position. Corrections to the putative disruptions were then introduced into the cloned genes using mutagenic primers with the Phusion™ Site-Directed Mutagenesis Kit.

Determination of DNA recognition sequence

The recognition sequence for each enzyme was determined by mapping positions of cleavage in several DNAs, typically pUC19, pBR322, PhiX174 and pBC4, followed by analysis to identify sequences that occur at the mapped cutting positions but not elsewhere in the DNAs (17,18). The putative recognition sequences were confirmed by comparing the observed fragments obtained by cleavage of larger DNA substrates having multiple sites, such as lambda, phage T7 or phage T3 DNAs, to the predicted fragments generated by cutting at the putative recognition sequence.

Determination of endonuclease cutting position

The location of DNA cleavage relative to the recognition sequence for each enzyme was determined through dideoxy sequencing analysis of the terminal base sequence obtained from cleavage of a DNA substrate at each of several different recognition sites (19). Sequencing was performed on an ABI 3130xl Genetic Analyzer using the BigDye® Terminator v3.1 Cycle Sequencing Kit. DNAs having a recognition site for the endonuclease located between 100 and 600 bp 3′ to each of a pair of sequencing primers were chosen for analysis. The DNA was cut by the endonuclease, purified over a spin column and eluted and its concentration adjusted to 100 µg/ml. Two DNA sequencing reactions were performed for each recognition site; one with a top strand primer to locate the endonuclease's position of cutting relative to the recognition sequence on the bottom strand, and one with a bottom strand primer to locate the position of cutting on the top strand.

Determination of DNA methyltransferase product

The methylated base produced by the DNA methyltransferase activity of the MmeI homologs was tested for 13 enzymes using antibodies specific for *N*6-methyl adenine or *N*4-methyl cytosine. An unmodified substrate, T7 DNA, was *in vitro* modified by these enzymes in reaction buffer lacking magnesium. A reaction to which no enzyme was added served as a negative control, while reactions with MmeI and M.TaqI served as positive controls for *N*6-adenine methylation. A plasmid DNA expressing M.EsaBC4I, which modifies a cytosine in the sequence 5′-GGCC-3′ at the N4 position (1), served as the positive control for the *N*4-methyl cytosine antibody. The modified DNAs were purified over a spin column and 0.45 µg, 0.15 µg and 0.05 µg of the modified DNAs were spotted onto nitrocellulose filters. Antibodies specific for *N*6-methyl adenine or *N*4-methyl cytosine were incubated with the DNAs and detected as previously described (20).

Determination of genomic DNA methylation status for *Rhodopseudomonas palustris* BisB5 (RpaB5I) and *Pseudomonas species* OM2164 (PspOMII)

Genomic DNA from host cells expressing two of the characterized enzymes was examined to determine if any modification was present in either the top strand or the bottom strand of the enzyme's recognition sequence to protect against the endonuclease activity of the endonuclease. For PspOMII, 20 µg genomic *P. species* OM2164 DNA was digested with HincII and EcoO109I to produce a discrete 502 bp fragment that contained two PspOMII recognition sites oriented in the same direction. Fragments between ∼400 and 600 bp were excised from an agarose gel and purified over a spin column. For synthesis of unmodified top strand PspOMII sites, 1 pmol of a top strand primer (#7, Supplementary Table T1) was 5′ end labeled with ^{33}P-γ-ATP (NEG302H) using T4 polynucleotide kinase in a 30 µl reaction. The kinase was inactivated by heat treatment at 80°C for 20 min. One-half of the gel purified, HincII and Eco0109I cut *P. species* OM2164 genomic DNA was placed in 500 µl 1× Phusion HF reaction buffer containing 0.25 µM dNTPS and five units Phusion HotStart DNA polymerase. The reaction mix was denatured at 98°C 1 min, then held at 94°C while the 1 pmol of labeled primer was added, then the primer was annealed and extended for 5 min at 72°C. The DNA was purified over a spin column. The same procedure was performed for synthesis of unmodified bottom strand PspOMII sites using a complement strand primer (#8, Supplementary Table T1). Two 500 µl endonuclease reaction mixtures were formed, one for the top strand synthesis and one for the bottom strand synthesis. Each contained the labeled hybrid DNA in NEBuffer 4 supplemented with 80 µM AdoMet, BSA and 40 nM of a 31 bp dsDNA having a PspOMII recognition site (annealed oligonucleotides #11 and #12, Supplementary Table T1) and divided into five equal portions. The first received no enzyme (negative control), the second four units PspOMII, the third two units PspOMII, the fourth 10 units BanII (positive control), while the fifth aliquot was mixed with reaction mixture from the opposite strand and digested with eight units PspOMII (positive control for PspOMII activity). Following digestion the DNA was concentrated using a spin column. The products were resolved on a 6% acrylamide TBE gel alongside PhiX-HaeIII size standard that was previously end labeled with ^{33}P, the gel was dried onto a Hybond N+ nylon membrane and the products detected by autoradiography. The same experimental approach was performed for *R. palustris* BisB5 genomic DNA. The *R. palustris* BisB5 DNA was cut with BsrBI (at genome coordinates 3593567 and 3593936) to generate a 369 bp fragment containing one RpaB5I site. The primers used were #9 and #10, and the small dsDNA containing an RpaB5I site was formed from oligonucleotides #13 and #14 (Supplementary Table T1). For RpaB5I, 1× Phusion GC reaction buffer supplemented with 3% DMSO was substituted for 1× Phusion HF buffer.

Cleavage on single site substrates: activation by *in trans* DNA

Short dsDNAs having a recognition site for RpaB5I or NmeAIII were supplied *in trans* to reactions containing RpaB5I and the single site substrate pBR322, or NmeAIII and the three-site substrate pBR322. Oligonucleotides were synthesized in pairs and annealed to form a dsDNA that contained the RpaB5I site (oligonucleotides #13 and #14) or the NmeAIII recognition site (oligonucleotides #15 and #16, Supplementary Table T1). These recognition site containing DNAs extended 14 bases 3′ to the recognition site, which is six or seven bases short of the point of DNA cleavage. The NmeAIII site DNA, which lacked an RpaB5I site, was tested with RpaB5I as a negative control. One microgram (0.35 pmol, 7 nM RpaB5I sites) linear pBR322 DNA (PstI cut) was digested in a 50 µl reaction with 2 to 0.25 units RpaB5I in the absence of the small DNAs, in the presence of 2 pmol (40 nM) RpaB5I site DNA or in the presence of 2 pmol (40 nM) NmeAIII site DNA (no RpaB5I site) as a negative control. Similarly one microgram linear pBR322 DNA (PstI cut) was digested in 50 µl reactions with two units RpaB5I and a 2-fold dilution of the RpaB5I site DNA from 2 pmol (40 nM) to 0.0625 pmol (1.25 nM). One microgram (0.35 pmol, 21 nM NmeAIII sites) pBR322 DNA was digested in a 50 µl reaction with a 2-fold dilution of NmeAIII from 32 to 0.5 units in the absence of the small NmeAIII site DNA. Similarly one microgram pBR322 DNA was digested in 50 µl reactions with 16 units NmeAIII and a 2-fold dilution of the NmeAIII site DNA from 32 pmol (640 nM) to 1 pmol (20 nM) per reaction. The extent of cleavage was determined on a 1% agarose gel.

Bioinformatic analysis of MmeI homologs

Multiple sequence alignment (MSA) of MmeI family enzymes was performed using the PROMALS web server (21). Structure prediction for the MmeI protein was performed using the PHYRE web server (22,23). The phylogeny of the characterized enzymes was analyzed using distances calculated by PROMALS in performing the MSA. The genome context for the MmeI homologs was analyzed by examining the predicted genes located within 5 kb on either flank of the identified MmeI homologs.

RESULTS

Identification of putative MmeI-like restriction endonuclease genes

A BLASTP search using the protein sequence of MmeI as query against the non-redundant GenBank database returned more than 100 sequences that produced highly significant expectation values of $E < e$–20. The identified putative sequences were annotated as ‘hypothetical proteins’ or ‘putative DNA methyltransferases.’ While it might be expected that the DNA methyltransferase portion of the bi-functional MmeI protein would produce matches to DNA methyltransferase genes because these contain conserved sequence motifs, none of the top 100 hits included typical type II DNA methyltransferases. Many of the putative sequences identified, and especially the highest scoring sequences, were highly similar to MmeI throughout their entire protein sequence, including the endonuclease and DNA recognition domains. The identified protein sequences were aligned and two groups were identified. The first was similar to the entire MmeI protein and contained conserved amino-acid residues of the PD–ExK endonuclease family in their amino terminal domain that aligned with those of MmeI and with each other. Sequences in the second set did not contain the PD–ExK endonuclease motif and differed from MmeI in their first 100 amino-acid residues yet were highly similar to MmeI and the first set of putative genes throughout the rest of their sequences. No additional DNA methyltransferase genes were observed flanking either set of MmeI

Table 1. Characterized MmeI family enzymes

Name	Recognition (cleavage)	Source organism	Accession	gDNA source
ApyPI	ATCG**A**C(20/18)	*Arcanobacterium pyogenes*	FJ773371	Stephen Billington
AquII	GCCGN**A**C(20/18)	*Agmenellum quadruplicatum* PR-6	YP_001733624	ATCC 2726
AquIII	GAGG**A**G(20/18)	*Agmenellum quadruplicatum* PR-6	YP_001735369	ATCC 27264
AquIV	GRGGA**A**G(20/18)	*Agmenellum quadruplicatum* PR-6	YP_001735547	ATCC 27264
CdpI	GCGG**A**G(20/18)	*Corynebacterium diphtheriae*	NP_940094	ATCC 700971
CstMI	AAGG**A**G(20/18)	*Corynebacterium striatum* M82B	NP_862240	Andreas Tauch
DraRI	CAAGN**A**C(20/18)	*Deinococcus radiodurans* R1	FJ773373	ATCC 13939
DrdIV	TACG**A**C(20/18)	*Deinococcus radiodurans* NEB479	FJ768705	NEB479
EsaSSI	GACC**A**C(20/18)	Environmental sample Sargasso Sea	EAJ03172	gene synthesis
MaqI	CRTTG**A**C(20/18)	*Marinobacter aquaeolei* VT8	YP_956924	ATCC 700491
MmeI	TCCR**A**C(20/18)	*Methylophilus methylotrophus*	ACC85607	NEB1189
NhaXI	CAAGR**A**G(20/18)	*Nitrobacter hamburgensis* X14	YP_579008	Dan Arp
NlaCI	CATC**A**C(19/17)	*Neisseria lactamica* ST640 Sange	Sanger Center Server[a]	Ronald Chalmers
NmeAIII	GCCG**A**G(21/19)	*Neisseria meningitidis* Z2491	FJ773372	Mark Achtman
PlaDI	CATC**A**G(21/19)	*Parvibaculum lavamentivorans*	YP_001413872	David Schleheck
PspOMII	CGCCC**A**R(20/18)	*Pseudomonas species* OM2164	FJ768704	NEB1783
PspPRI	CCYC**A**G(21/19)	*Psychrobacter species* PRwf-1	YP_001274371	James Tiedje
RceI	CATCG**A**C(20/18)	*Rhodospirillum centenum* SW	YP_002299341	ATCC 51521
RpaB5I	CGRGG**A**C(20/18)	*Rhodopseudomonas palustris* BisB5	YP_570364	Caroline Harwood
SdeAI	CAGR**A**G(21/19)	*Sulfurimonas denitrificans*	YP_392994	ATCC 33889
SpoDI	GCGGR**A**G(20/18)	*Silicibacter pomeroyi* DSS-3	YP_167160	Mary Ann Moran
BsbI	CAAC**A**C(21/19)	*Bacillus species* NEB686	No sequence data	NEB686

The name, recognition sequence and position of DNA cleavage, the source organism, the GenBank accession number for the amino acid sequence (full length active form), and the source of the genomic DNA is listed for each enzyme. The penultimate adenine that is the target of methylation is in bold and underlined.

[a] The NlaCI sequence is not yet deposited in GenBank, but is accessible from the Sanger Center server: http://www.sanger.ac.uk/Projects/N_lactamica/.

homologs. No particular genes consistently flanked the homologs that contained the endonuclease domain motif. However, the set of homologs that lacked the endonuclease motif, such as YeeA (GenBank accession no NP_388558) from *Bacillus subtilis*, were flanked by two conserved genes; one identified as a putative DNA helicase, similar to YeeB of *B. subtilis* (GenBank accession no AAB66475), and a second identified as a conserved hypothetical protein similar to YeeC of *B. subtilis* (GenBank accession no AAB66476). Selected sequences were cloned into *E. coli* for expression and characterization.

Characterization of novel restriction endonucleases

We have identified, expressed and characterized 20 novel restriction endonucleases. For each new enzyme the name, recognition sequence and DNA cleavage position, source organism, protein accession number and source of genomic DNA is presented in Table 1. Of the 20 newly discovered enzymes, 19 have unique DNA specificities not previously known for type II restriction–modification (R–M) systems. The one exception, AquIII [5′-GAGGAG(20/18)], recognizes the same DNA sequence as BseRI [5′-GAGGAG(10/8)] but cuts the DNA at a different position, making it a neoschizomer to BseRI. While each enzyme recognizes a different sequence, some of the recognition sequences differ at only one base position, while others differ at every position except the penultimate adenine that is the target for the DNA methyltransferase activity of these proteins. The enzymes all required AdoMet for endonuclease activity. An example of recognition sequence determination is shown in Figure 1. The positions of the single RpaB5I cut in pBR322, and the four cuts in pBC4, were mapped by cutting the DNAs with RpaB5I and restriction enzymes having a single site in these DNAs. The sequence CGAGGAC or CGGGGA C was found to occur only at the mapped positions in these DNAs, and did not occur in two other DNAs, pUC19 and PhiX174, that were not cut by RpaB5I. The observed fragment sizes produced by RpaB5I digestion of larger DNA substrates such as lambda, T7 and T3 matched the predicted fragment sizes for cutting at CGRGGAC, confirming that this is the specific recognition sequence for RpaB5I.

The enzymes all cut DNA at essentially the same position relative to their recognition site, that is 20 (±1) bases 3′ to the recognition sequence in the top DNA strand that contains the adenine that is methylated, and two bases 3′ to this top strand cut in the bottom strand, to produce a two-base 3′ extension. The exact position of DNA scission can vary by one base at different sites for each enzyme, dependent upon the sequence that occurs between the recognition site and cleavage point. An example of run off sequencing to determine the cleavage position relative to the recognition sequence is shown for RpaB5I, which cuts at 5′-CGRGGAC(20/18)-3′ (Figure 2). For many of the enzymes variability of one base longer or shorter than the typical (20/18) reach was observed for cleavage at different sites, and some sites showed a mixture of cutting length products, for example at (21/19) and (20/18). The cleavage distance most frequently observed for each enzyme is reported in Table 1; however it should be understood that an enzyme reported as (21/19) may cut some sites at (20/18) and vice versa.

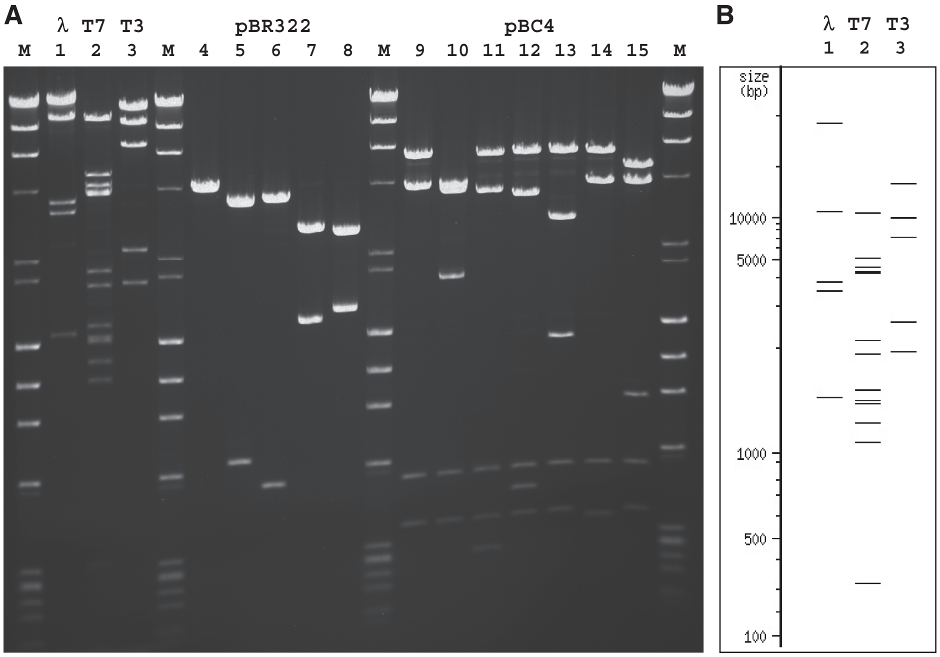

Figure 1. Determination of RpaB5I recognition site. (**A**) An agarose gel showing the products of RpaB5I digestion of various DNAs. The putative recognition sequence was derived from the positions of the cut sites and analysis of the pBR322 and pBC4 DNA sequences as described in 'Materials and Methods' section. Digestion of lambda, T7 and T3 phage DNAs served to verify the predicted specificity. Lane 1: lambda DNA, lane 2: T7 DNA, lane 3: T3 DNA. Lanes 4–8: pBR322 DNA cut by RpaB5I and: lane 4: RpaB5I only, lane 5: EcoRV, lane 6: BsmI, lane 7: NdeI, lane 8: PstI. Lanes 9–15: pBC4 DNA cut by RpaB5I and: lane 9: RpaB5I only, lane 10: NdeI, lane 11: AvrII, lane 12: PmeI, lane 13: AscI, lane 14: SpeI, lane 15: EcoRV. Lanes M are HindIII-lambda and HaeIII-PhiX174 DNA size standards. (**B**) Computer generated digestion patterns for cleavage at the predicted RpaB5I recognition sequence (5′-CGRGGAC-3′). Lane 1: lambda DNA, lane 2: T7 DNA, lane 3: T3 DNA.

Two members of the set of sequences that do not contain the PD–ExK endonuclease motif, YeeA of *B. subtilis* and MslORFHP of *Moraxella osloensis* NEB722, were similarly expressed but no endonuclease activity was observed.

Activation of inactive native genes

Three enzymes were activated from open reading frames that were found to be disrupted in the particular isolate used for genomic sequence determination. The DNA sequence reported in the database leading to the interruption in the coding frame was confirmed for all three cases. Successful prediction of where to introduce changes and what specific changes to make to correct the reading frames of these enzymes was possible due to the significant sequence conservation found among members of this protein family. Only a single base change was necessary to change early termination stop codons to coding codons for NmeAIII and DraRI. For NmeAIII the early termination TAG stop codon at amino-acid position 32 of the full length ORF was changed to TGG (tryptophan) using primers 1 and 2 (Supplementary Table T1). The early termination codon TAA at position 841 in DraRI was corrected to GAA using primers 3 and 4 (Supplementary Table T1). ApyPI contained a frame shift following R_{886} that was corrected by the addition of two bases, GC, to a run of three GC dinucleotides (GCGCGC changed to GC GCGCGC) using primers 5 and 6 (Supplementary Table T1). Individual transformants carrying the corrected genes were tested for endonuclease activity and found to express active endonuclease. *Deinococcus radiodurans* genomic DNA was tested and found to be cleaved *in vitro* by the activated DraRI endonuclease, indicating the DNA methyltransferase activity of DraRI is not active *in vivo* in the host *Deinococcus* strain.

DNA methyltransferase activity produces N6-adenine methylation

The conserved DNA methyltransferase motifs in the expressed MmeI-like enzymes are those of the amino DNA methyltransferases, and are most similar to the gamma class of *N*6-methyl adenine DNA methyltransferases. Fourteen enzymes were tested and confirmed to produce *N*6-methyl adenine by the use of antibodies specific for *N*6-methyl adenine (Figure 3A). None of these enzymes produced any detectable *N*4-cytosine methylation (Figure 3B). The antibody results confirm that the enzymes in this family modify adenine at the *N*6 position to form *N*6-methyl adenine (m6A).

Genome context of MmeI family homologs

None of the genes expressed have an additional DNA methyltransferase gene in close proximity to the single polypeptide coding for the fused endonuclease–DNA methyltransferase enzyme. Furthermore, no one conserved gene, putative or characterized, is observed to co-localize with the 20 characterized endonuclease genes in their genome context. The absence of a nearby

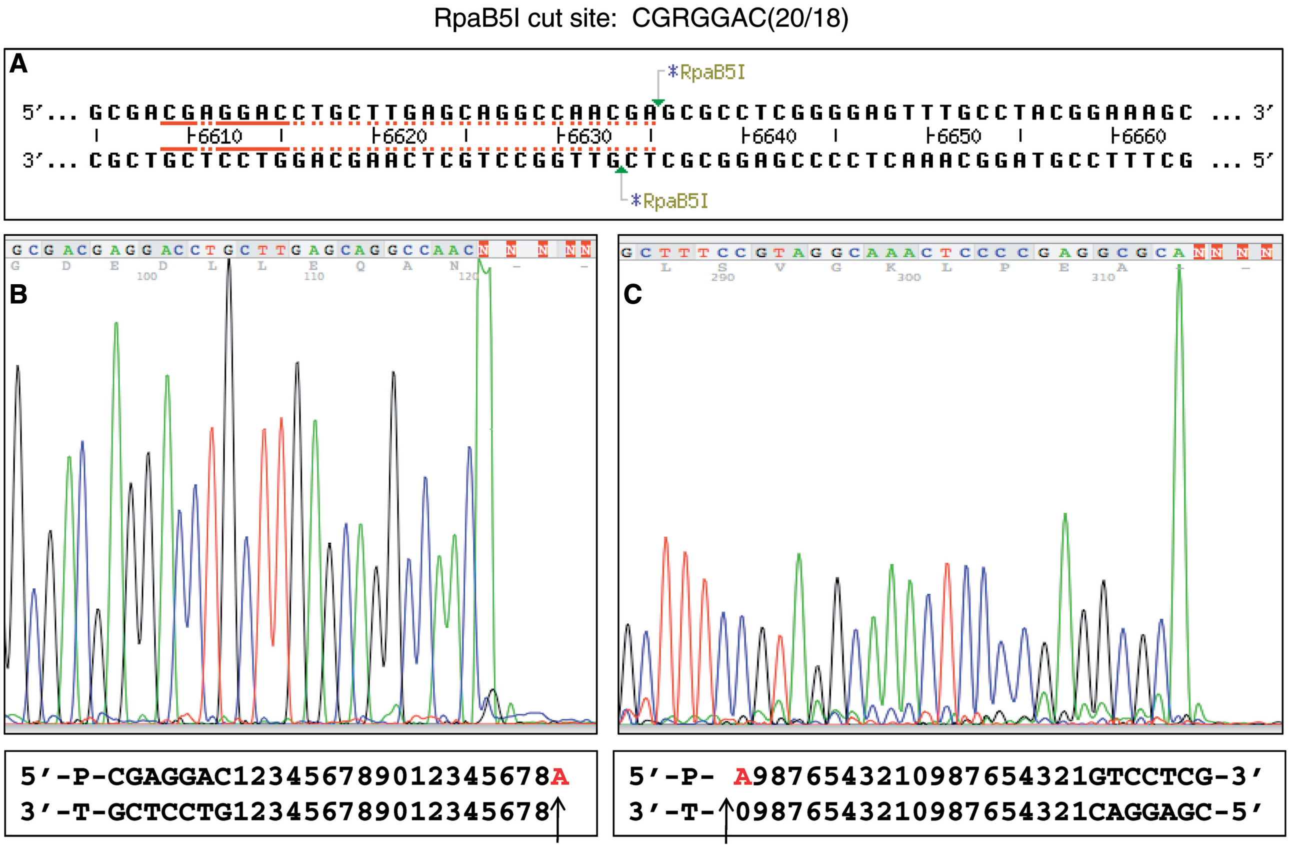

Figure 2. Determination of the position of DNA cleavage for RpaB5I. (**A**) The DNA sequence of both strands adjacent to the RpaB5I site at 6609 in pBC4 DNA. (**B**) Run-off dideoxy sequencing on the top strand that demonstrates cleavage in the bottom strand 18 bp 5′ to the RpaB5I site: 5′-/18GTCCTCG-3′. (**C**) Run-off dideoxy sequencing on the bottom strand that demonstrates cleavage in the top strand 20 bp 3’ to the RpaB5I site: 5′-CGAGGAC20/-3′. Note that the Taq polymerase used for sequencing adds an extra ‘A’ base onto the end of the DNA molecule.

methyltransferase gene is consistent with the observation that all the systems tested use only the single-strand DNA modification produced by the bi-functional enzyme for host protection. Indeed, 15 of the characterized enzymes do not have an adenine base in the complement strand that could serve as a target for m6A modification, while one member of this family, BsbI, has neither adenine nor cytosine bases in the complement strand.

Modification of host DNA occurs on only one DNA strand of the duplex recognition sequence

Host genomic DNA was examined for the presence or absence of modification able to protect against endonuclease cleavage in each DNA strand of the recognition sequence. DNA substrates that consisted of a hybrid of one strand of host genomic DNA, which will carry the respective DNA modification present in the host cell, and one newly synthesized and therefore un-modified DNA strand, were produced from a single round of primer extension on host genomic DNA for RpaB5I and PspOMII. Both enzymes cut the hybrid DNAs in which the bottom strand of the recognition sequence, 5′-GTCCYCG-3′ for RpaB5I and 5′-YTGGGCG-3′ for PspOMII, was derived from their host genomic DNA and the top strand was newly synthesized, indicating the host DNA has no modification present in the bottom strand of the recognition sequence to block cleavage by these enzymes (Figure 4, Supplementary Figure S2). DNA in which the top strand of the recognition sequence, 5′-CGRGGAC-3′ for RpaB5I and 5′-CGCCCAR-3′ for PspOMII was located in the host-derived genomic strand and the bottom strand was newly synthesized were not cut by these enzymes, indicating modification is present in the top strand to prevent cleavage. The same results were observed previously for MmeI (12). These results indicate that only the top strand adenine modification produced by the DNA methyltransferase activity of the bi-functional enzymes is present in their respective host DNA and able to block cleavage. No additional modification is present in the host DNA on the bottom strand of the enzyme’s recognition sequence to block endonuclease activity. This observation is consistent with the absence of a co-localized companion DNA methyltransferase in the genome sequence context of these enzymes and the lack of a conserved adenine base target for modification in the bottom strand of their recognition sequences. These results confirm that the entire modification used by the RpaB5I, PspOMII and MmeI restriction systems, and by inference all members of this family of R–M systems, is the methylation of only the one conserved top strand adenine produced by these single polypeptide, bi-functional enzymes themselves.

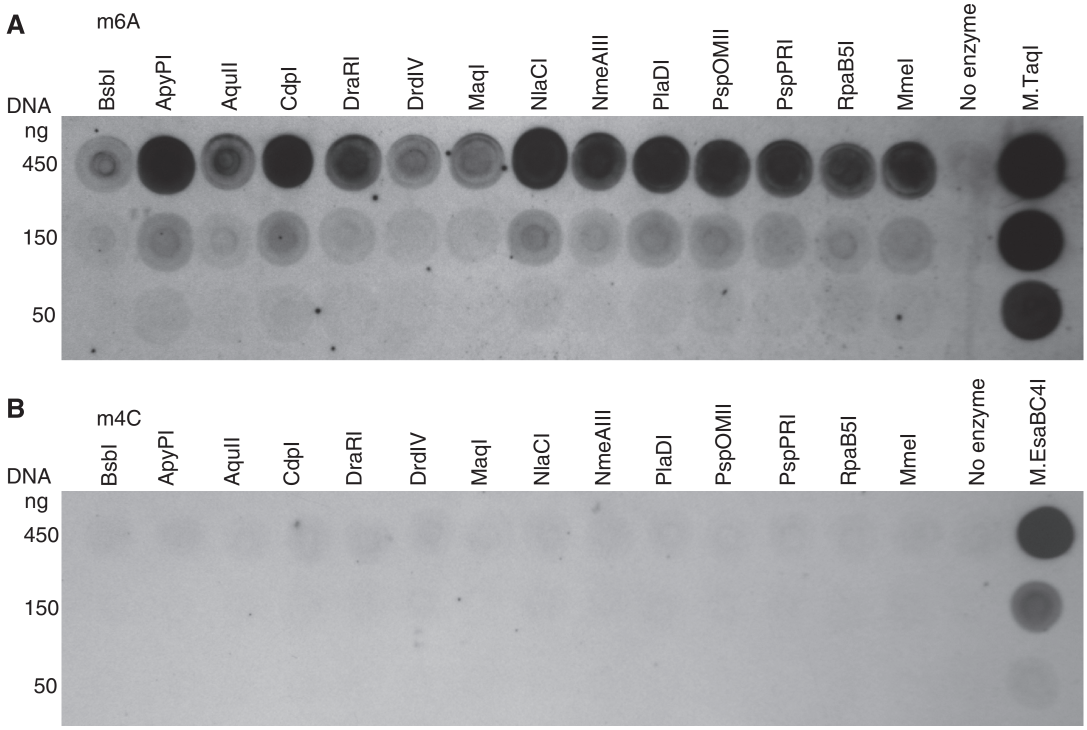

Figure 3. Detection of DNA methylation produced *in vitro* by MmeI family enzymes. Antibodies specific for either m6-methyl adenine or m4-methyl cytosine were incubated with T7 DNA that had been *in vitro* modified by the enzymes listed. M.TaqI modified T7 DNA served as positive control for m6-methyl adenine. M.EsaBC4I modified plasmid DNA served as positive control for m4-methyl cytosine. (**A**) N6-methyl adenine specific antibody. (**B**) *N*4-cytosine specific antibody.

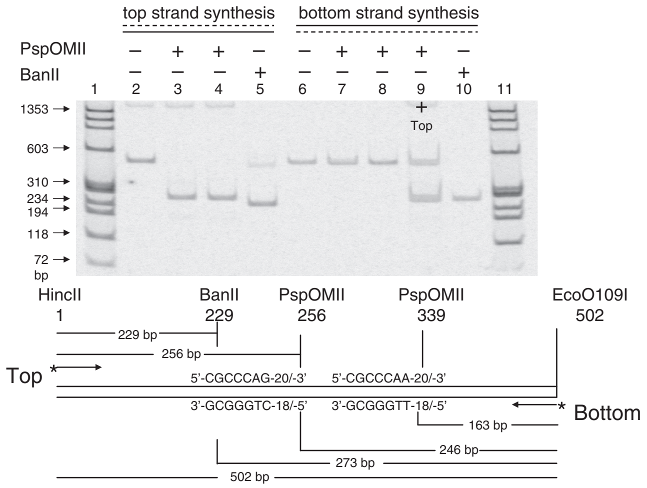

Figure 4. PspOMII digestion of DNAs containing one strand from native genomic *P. pseudomonas* DNA and one newly synthesized (unmodified) strand. Lanes 2–5 are newly synthesized top strand with genomic bottom strand, while lanes 6–10 are the newly synthesized bottom strand with genomic top strand. Lane 2: uncut, lane 3: four units PspOMII, lane 4: two units PspOMII, lane 5: 10 units BanII. Lane 6: uncut, lane 7: four units PspOMII, lane 8: two units PspOMII, lane 9: four units PspOMII digestion of mixed newly synthesized bottom strand and top strand DNA (as a positive control for PspOMII activity), lane 10: 10 units BanII. Lanes 1 and 11: PhiX174-HaeIII size standard. PspOMII cuts the DNA containing a genomic *P. species* OM2164 bottom strand and an unmodified top strand (lanes 3, 4 and 9), but not the DNA containing a genomic *P. species* OM2164 top strand and an unmodified bottom strand (lanes 7, 8 and 9). The native host DNA from *P. species* OM2164 is thus modified to prevent PspOMII cleavage only in the top DNA strand (5′-CGCCCAR-3′′) of the PspOMII recognition sequence.

MmeI family enzymes requires two sites for efficient DNA cleavage

Some type II endonucleases bind individual recognition sites and cleave their sites independently. Others require two or more sites for efficient cleavage, with the multiple sites either acting cooperatively to effect cleavage, or with one site binding to an effector position in the endonuclease to effect a conformational change required for DNA cleavage competence (4,24–26). The cleavage efficiency on DNA substrates containing single or multiple recognition sites was compared.

All the enzymes tested cleaved a single site DNA incompletely, achieving between 10 and 70% cleavage even with excess enzyme. For example, RpaB5I cuts its single site in pBR322 DNA only partially (Figure 5A). In contrast, the same single site DNA is nearly completely cleaved when a second recognition site is provided *in trans* by adding a synthetic DNA containing the RpaB5I recognition site (Figure 5B and C). The DNA bearing the recognition site need not be capable of being cleaved itself, as a DNA having only 14 bases 3′ to the recognition site facilitates cleavage of the single site plasmid as well as a DNA extending to or beyond the position of cleavage. Cleavage stimulation is dependent upon the presence of the enzyme's specific recognition sequence, as addition of a similar DNA lacking an RpaB5I site did not increase cleavage (Figure 5D). For RpaB5I the concentration of sites supplied *in trans* needed to stimulate cutting of the single site DNA was approximately equimolar (0.01 µM) with the concentration of recognition sites (0.007 µM)

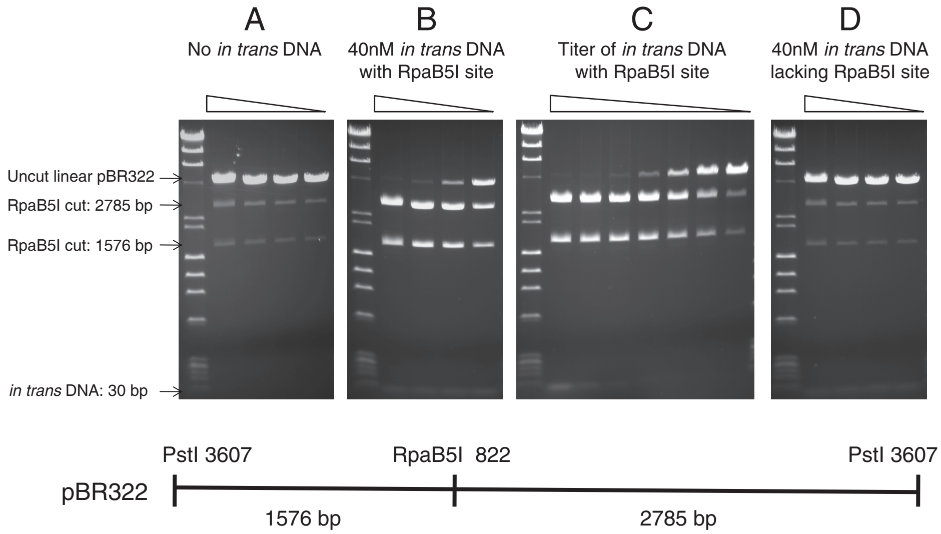

Figure 5. Cleavage of a single site substrate is incomplete but can be stimulated by *in trans* DNA containing a specific recognition site. (**A**) No *in trans* DNA. RpaB5I digestion in a 2-fold serial dilution from 2 units/μg DNA to 0.25 units/μg DNA on pBR322 DNA previously linearized by digestion with PstI. (**B**) Forty nanomolar *in trans* DNA containing an RpaB5I recognition site. The same reaction conditions as (A) supplemented with 40 nM of a 30 bp *in trans* DNA containing the RpaB5I recognition site. (**C**) 2-fold dilution series of the *in trans* DNA containing the RpaB5I recognition site, from 40 nM to 0.625 nM, in reactions containing two units RpaB5I and 1 μg PstI-linearized pBR322 per 50 μl reaction (7 nM RpaB5I sites). (**D**) The same reaction conditions as (A) supplemented with 40 nM of a 30 bp *in trans* DNA lacking the RpaB5I recognition site.

in the single site DNA (Figure 5C), in a reaction that contained two units of RpaB5I. These results are quite similar to those obtained for MmeI (12).

Several of the enzymes characterized, such as NmeAIII and PspOMII, cut even a multiple site substrate incompletely, producing a stable, partial digestion pattern even with excess enzyme. For example, NmeAIII cuts pBR322, which contains three sites, to a stable partial digestion pattern that does not change even with 32-fold excess enzyme (Figure 6A). NmeAIII cleavage of pBR322 is stimulated by the presence of its recognition site *in trans*, as observed for MmeI and RpaB5I; however in contrast to MmeI and RpaB5I, this stimulation requires an $\sim$10-fold excess of both the enzyme and the *in trans* DNA in order to drive the cleavage reaction on the pBR322 substrate to completion (Figure 6B). These results indicate that while all of the enzymes described require interaction between two specific recognition sites for cleavage, there are subtle differences in the endonuclease domains and their interactions that affect the extent of DNA scission produced.

Protein sequence features

The new enzymes described share many common features. They are single polypeptides that encode both the DNA methyltransferase activity required for host protection and the endonuclease activity for cleavage of identifiably foreign DNAs. The proteins are relatively large for type II restriction endonucleases, ranging in length from 908 amino acids (SdeAI) to 1184 amino acids (RpaB5I). The primary amino-acid sequences of the characterized enzymes are quite similar, with ApyPI and CstMI sharing 76% identity, and many of the enzymes exhibiting 40–50% identities. The amino-acid sequences align well, particularly when secondary structure predictions are included in the alignment algorithm (Supplementary Figure S1). The enzymes display a remarkable conservation of predicted secondary structure elements throughout the entire alignment, while also displaying the flexibility common to restriction enzymes for accommodating insertions of short sequence elements in individual enzymes between regions of conserved sequence and secondary structures.

The endonuclease domain is located at the amino terminus of these proteins and contains the conserved motifs of the PD-ExK endonuclease family. Secondary structure prediction indicates the endonuclease domain forms a structure containing four helices and five beta strands in the order α-β-β-β-α-α-β-β-α, suggesting these enzymes fall into the class III group of restriction endonucleases proposed by Niv, *et al.* (8).The aspartate of the PD–ExK motif is completely conserved and occurs at the start of beta strand 2 (D_{70} in MmeI). The E and K are also completely conserved and occur at the end of the third beta strand (E_{80} and K_{82} in MmeI). Mutations to these residues have recently been shown to abolish endonuclease activity (27). There is a highly conserved (17 of 20) glutamate residue at the c-terminal end of beta strand 1 (E_{51} in MmeI), though this is an aspartate in one case and a glutamine in two of the enzymes. A completely conserved glutamate also occurs just before the start of helix 2 (E_{25} in MmeI).

A second feature observed in the MSA is a region of predominantly helical nature located between the endonuclease domain and the methyltransferase domain, from approximately amino acids 151–300 in MmeI. This region is presumed to form the ‘arm’ that enables the enzyme to position the endonuclease domain two turns of the DNA helix, or 20 nt, away for DNA cleavage when the enzyme is bound at the recognition sequence. This region shares

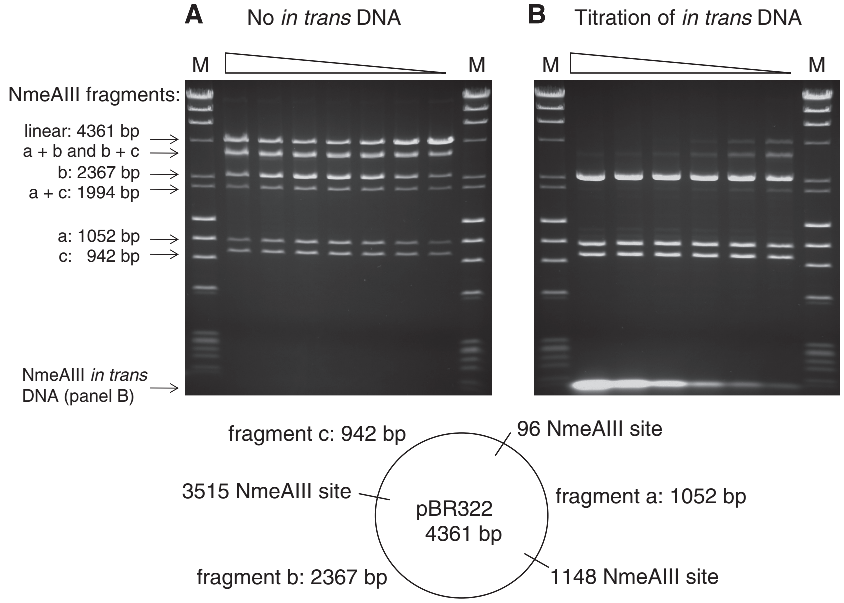

Figure 6. Cleavage of a multiple site substrate by NmeAIII. (**A**) 2-fold serial dilution series of NmeAIII digestion of pBR322 DNA (three NmeAIII sites), from 32 units per 50 µl reaction to 0.5 units per 50 µl reaction. (**B**) 2-fold dilution series of an *in trans* DNA containing the NmeAIII recognition site, from 640 nM to 20 nM, in reactions containing 16 units NmeAIII and 1 µg pBR322 per 50 µl reaction (21 nM NmeAIIII sites).

similarity to the amino terminal portion of the type I DNA methyltransferases, suggesting an evolutionary relationship between the MmeI family and type I DNA methyltransferases, as has recently been proposed (27).

The methyltransferase domain contains readily identifiable amino-acid sequence motifs of the amino DNA-methyltransferases. These motifs occur in the order found in the gamma class of *N*6-adenine methyltransferases: motif X, motif I to motif VIII. Structure prediction algorithms model the methyltransferase domain of these enzymes onto the structure of the gamma class m6A DNA methyltransferase M.TaqI (PDB: 1G38) with high accuracy probabilities, indicating that the methyltransferase domain of these enzymes is typical of the gamma m6A DNA methyltransferases. The methyltransferase domain corresponds to approximately amino acids 301–620 in MmeI.

In the type II gamma class m6A DNA methyltransferases, specific recognition is determined by the Target Recognition Domain (TRD) located C-terminal to the methyltransferase domain, as well as minor groove contacts located between methyltransferase motifs IV and V in the case of M.TaqI, while in the type I systems recognition is supplied by a separate specificity polypeptide. For the enzymes described the TRD appears to immediately follow the methyltransferase domain as in the type II gamma class m6A DNA methyltransferases, corresponding to approximately position 621–820 in MmeI. There is remarkable conservation in the predicted secondary structure elements within the TRD region, indicating that the enzymes are likely to contact the DNA using similar structural elements. The enzymes form two main branches in a phlyogenetic analysis, with the enzymes recognizing six-base sequences in one and those recognizing seven-base sequences in the other, with the one exception of DrdIV (Figure 7). The enzymes recognizing seven-base

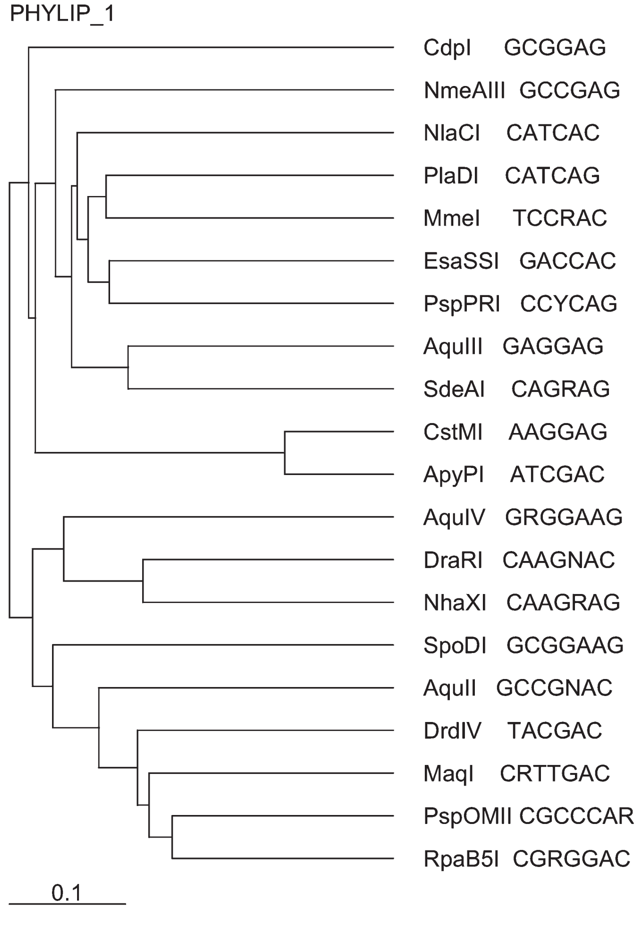

Figure 7. Phylogenetic tree of the MmeI family enzymes calculated from the distances generated by the PROMALS multiple sequence alignment.

sequences exhibit a small insertion of seven amino acids and a small deletion of four amino acids relative to the six-base enzymes within the putative TRD region.

The TRD appears to end at a conserved short sequence motif, 'FPFP', that is reminiscent of the PLPPL motif found in type I specificity subunits. The PLPPL motif occurs at the transition from one-half site TRD to the helical spacer arm that connects the two-half site TRD domains (28). Following this 'FPFP' motif there is a C-terminal region consisting of several predicted well-conserved helices of unknown function.

DISCUSSION

We have identified a new family of restriction endonucleases that have remarkably similar overall amino-acid sequences yet recognize different DNA sequences. The enzymes cut DNA at the same 20 (±1) base pair distance downstream from their recognition site, making them members of the type IIS subgroup of restriction enzymes. They possess both endonuclease and DNA methyltransferase activities in the same polypeptide and require AdoMet for endonuclease activity, making them members of the type IIG subgroup as well. The enzymes recognize 6 or 7 nt long contiguous sequences. The effective number of base pairs specifically recognized ranges from 5.5 bp to 7 bp, though there is a bias toward effective recognition of 6 bp, as among the enzymes recognizing bases at seven positions, two uniquely specify recognition of only 6 bp, six uniquely specify 6.5 bp, while only one uniquely specifies a unique base at all seven positions.

Modification of one DNA strand

Remarkably, enzymes in this family use modification on only one DNA strand for host protection. This allows a single DNA recognition domain to direct both host protective methylation and endonuclease activity. However, single-strand modification would be expected to pose a problem for the host, in that immediately following replication every recognition site will be completely unmodified on one daughter duplex DNA molecule, and thus presumably unprotected from the endonuclease activity of the enzyme. It is currently unclear just how the host overcomes this difficulty, though the requirement that enzyme bound at two specific unmodified sites interact for cleavage to occur may play a role. However, it is clear that the MmeI-like restriction systems described are viable and widespread in nature despite relying on only single-strand modification. While we believe these enzymes to be in fact restriction systems, a restriction phenotype has not been explicitly demonstrated in their native hosts, though cloned MmeI has been reported to restrict lambda phage in *E. coli* (27).

It appears likely that other type II restriction systems also use single-strand modification for host protection. For example, TaqII [5-GACCGA(11/9) or CACCCA (11/9) (29), Tth111II (CAARCA(11/9) (30) and TspGWI (ACGGA(11/9)] (31) all lack an adenine base target for modification in the bottom strand of their recognition sequences. Although the genome context of these three systems is not available, close homologs for which flanking genomic sequence is currently available, such as putative genes RPA3376 (accession no. NP_948715) or TK1158 (accession no YP_183571), are not associated with an additional flanking DNA methyltransferase gene. Bioinformatics analyses of available microbial genome sequences indicate a wide range of putative RM systems exists that potentially might use single-strand modification. Three rounds of an iterated Psi-BLAST search using MmeI as the starting query returns greater than 500 putative protein sequences, many of which, like the MmeI family, have readily identifiable adenine DNA methyltransferase motifs and are not flanked by a second DNA methyltransferase protein. A number of these have an N-terminal domain that contains a PD–ExK endonuclease motif. Others lack an endonuclease motif but are flanked by a putative DNA helicase protein. Still others are large single polypeptides that include a putative endonuclease domain and helicase domain along with the DNA methyltransferase and TRD. To date these systems have received little attention or biochemical characterization, yet their frequency would suggest they might have an important biological role.

Evolution of the MmeI family of R–M systems

The MmeI-like systems described may represent a kind of 'missing link' in the evolution of R–M systems. The remarkable similarity observed between the amino terminal portions of the type I DNA methyltransferases, which precedes the methyltransferase motifs, and the region in the MmeI family proteins located between the endonuclease domain and the start of the methyltransferase domain suggests a close evolutionary relationship. This region is absent in typical type II gamma class DNA methyltransferases, such as M.TaqI. In the MmeI family this region presumably forms the 'arm' that positions the endonuclease domain two turns of the helix away from the recognition sequence, while in the type I systems it presumably functions in protein–protein subunit interactions. The MmeI family enzymes thus resemble a type I DNA methyltransferase with two domain additions. At the amino terminus a PD–ExK endonuclease domain has fused with the methyltransferase, while at the C-terminus a specificity domain recognizing a contiguous DNA sequence replaces the separate specificity subunit found in type I systems. Some type IIG systems that, like their type I counterparts, recognize split sequences still have a separate specificity subunit, for example BcgI (32), while in others the specificity domain is fused with the endonuclease–methyltransferase polypeptide, for example CjeI or AloI (33).

The second group of MmeI homologs, that lack the endonuclease motif in their amino terminal domain, may represent a different link with the type I systems. These are invariably located next to two conserved putative genes in their various genome contexts, suggesting that these three putative proteins function together. One of these conserved putative genes is similar to the DEAD domain DNA helicase superfamily, CDD cl10452 (34). The second conserved putative protein contains conserved sequence motifs of the GIY-YIG endonuclease family (35). We speculate that in these systems the MmeI-like subunit provides DNA specificity and protective host modification, but in place of the PD-ExK endonuclease domain this group presumably substitutes a domain that interacts with the separate DNA helicase and endonuclease subunits to accomplish DNA scission. The architecture of separate subunits and the presence of a DNA helicase are similar to type I R–M systems; however, in contrast to type I systems, these systems are likely to use single-strand modification like MmeI.

The MmeI family enzymes require interaction between two molecules bound at specific recognition sites to achieve cutting. This indicates the endonuclease domain is released to interact with DNA only upon specific binding, suggesting there must be some intramolecular communication resulting from specific binding. In the type I DNA methyltransferase M.EcoKI, the N-terminal region has been implicated in reading the methylation status of the adenine at one half site and communicating this with the second M.EcoKI molecule positioned at the adenine on the opposite strand at the second half site, to direct the EcoKI system into either modification, if one DNA strand is methylated, or endonucleolytic digestion if neither adenine is modified (36). This type I N-terminal region is similar to the MmeI family region between the endonuclease domain and methyltransferase domain. The MmeI family enzymes have a completely conserved tryptophan residue (W_{287} in MmeI, Supplementary Figure S1) that aligns to M.EcoKI W_{115}, which has been shown to participate in the communication of the methylation status in M.EcoKI (36). The conservation observed between these groups of enzymes suggests that this region of the MmeI-like enzymes preceding the start of the methyltransferase motifs may be involved in communicating the methylation status of the recognition site to the endonuclease domain to activate this domain for cutting.

Distribution of the MmeI family of R–M systems

The identification of these novel type II restriction enzymes through the technique of ‘genome mining’ demonstrates the power of this sequence-based approach for enzyme discovery. That twenty enzymes having novel DNA specificity were isolated demonstrates many novel type II restriction endonuclease enzymes still await discovery.

Because type II restriction endonucleases generally exhibit significant protein sequence similarity only among enzymes that recognize the same sequence, we were surprised to find the MmeI family enzymes all recognize unique DNA sequences. In the context of their highly similar overall sequences, the diversity of DNA recognition observed indicates DNA specificity is evolving rapidly within this family. Because the entire restriction system of the MmeI family enzymes consists of a single bi-functional protein in which DNA cutting and protective host modification are directed from a common DNA recognition domain, any alteration in the single DNA recognition domain will simultaneously alter both restriction specificity for foreign DNA and methylation specificity to protect host DNA. This coordination of protective methylation and restriction activities from a common DNA recognition domain, as in type I systems, ensures host protection is available for any new restriction specificity generated, which may explain the great diversity of DNA specificity observed in this family. This capacity to evolve new recognition specificity may confer a selective advantage for organisms facing a rapidly evolving phage challenge. The enhanced ability to evolve new specificity may also explain the widespread occurrence of single-strand modifying R–M systems in Nature despite the potential deleterious effects from production of unmodified sites in one daughter strand following replication.

Three of the enzymes characterized have genes that are disrupted in the host organism. When expressed directly from the sequenced organism’s DNA these genes were inactive. However, because the members of this family share such a high degree of sequence similarity it was possible to predict the location of the lesions and correct the disruption to form functional genes. The ability to form active enzymes by a simple mutational event suggests these genes have not degenerated. This implies that in a natural population of microbes these enzymes may exist in both an active form and an inactive, but readily repaired, form. The active systems may confer a selective advantage in periodic times of challenge by phage or parasitic DNAs. The inactive form of these genes may represent drift in the absence of selection; however such inactive genes might also be particularly amenable to evolutionary changes within their DNA recognition domain. Because the inactive genes can be readily reactivated, their presence may confer a selective advantage to whichever host cells express active enzyme when the population experiences a challenge from parasitic DNA.

Proposed new sub-group: type IIL

The MmeI family enzymes we describe modify only one DNA strand to provide host protection. The absence of protective modification in the second strand has been explicitly demonstrated for three enzymes, and is consistent with both the observed absence of linkage to an additional DNA methyltransferase for modification of the second strand, and the absence of an adenine base to accept modification in the second strand of the majority of the enzymes. This family of enzymes thus has complete reliance on the single-strand modification provided by the fused endonucleases–DNA methyltransferase protein for host protection. This utilization of only single-strand modification for host protection is a newly described characteristic among type II R–M systems. Based on this feature we propose that such enzymes be classified in a new sub group within the type II restriction enzymes, the type IIL enzymes, where ‘L’ indicates Lone strand modification.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Romas Vaisvila for the *in vitro* synthesis of the EsaSSI gene. We thank Stephen Billington for *A. pyogenes* genomic DNA, Andreas Tauch for *C. striatum* genomic DNA, Ronald Chalmers for *N. lactamica* ST640 genomic DNA, David Schleheck for *P. lavamentivorans* genomic DNA, Dan Arp and Shawn Starkenburg for the *N. hamburgensis* X14 strain, James Tiedje for the *P. species* PRwf-1 strain, Caroline Harwood and Yasuhiro Oda for *R. palustris* BisB5 and BisB18 strains, Mark Achtman for *N. meningitidis* Z2491 genomic DNA and Mary Ann Moran for *S. pomeroyi* genomic DNA. We thank Chandra Pedamallu for assistance preparing figures. They also thank Richard Roberts, Geoffrey Wilson, Jim Samuelson, Iain Murray and William Jack for critical reading of the manuscript and discussions.

FUNDING

Funding for open access charge: New England Biolabs, Inc. 240 County Road Ipswich, MA 01938 USA.

Conflict of interest statement. None declared.

REFERENCES

1. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2007) REBASE–enzymes and genes for DNA restriction and modification. *Nucleic Acids Res.*, **35**, D269–D270.
2. Raleigh,E.A. and Brooks,J.E. (1998) Restriction modification systems: where they are and what they do. In De Bruijn,F.J., Lupski,J.R. and Weinstock,G.M. (eds), *Bacterial Genomes*. Chapman & Hall, New York, pp. 78–92.
3. Janulaitis,A., Petrusyte,M., Maneliene,Z., Klimasauskas,S. and Butkus,V. (1992) Purification and properties of the Eco57I restriction endonuclease and methylase-prototypes of a new class (type IV). *Nucleic Acids Res.*, **20**, 6043–6049.
4. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S.K., Dryden,D.T., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
5. Janulaitis,A., Vaisvila,R., Timinskas,A., Klimasauskas,S. and Butkus,V. (1992) Cloning and sequence analysis of the genes coding for Eco57I type IV restriction-modification enzymes. *Nucleic Acids Res.*, **20**, 6051–6056.
6. Posfai,J., Bhagwat,A.S., Posfai,G. and Roberts,R.J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.*, **17**, 2421–2435.
7. Malone,T., Blumenthal,R.M. and Cheng,X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.
8. Niv,M.Y., Ripoll,D.R., Vila,J.A., Liwo,A., Vanamee,E.S., Aggarwal,A.K., Weinstein,H. and Scheraga,H.A. (2007) Topology of Type II REases revisited; structural classes and the common conserved core. *Nucleic Acids Res.*, **35**, 2227–2237.
9. Bujnicki,J.M. (2003) Crystallographic and bioinformatic studies on restriction endonucleases: inference of evolutionary relationships in the 'midnight zone' of homology. *Curr. Protein Pept. Sci.*, **4**, 327–337.
10. Pawlak,S.D., Radlinska,M., Chmiel,A.A., Bujnicki,J.M. and Skowronek,K.J. (2005) Inference of relationships in the 'twilight zone' of homology using a combination of bioinformatics and site-directed mutagenesis: a case study of restriction endonucleases Bsp6I and PvuII. *Nucleic Acids Res.*, **33**, 661–671.
11. Kinch,L.N., Ginalski,K., Rychlewski,L. and Grishin,N.V. (2005) Identification of novel restriction endonuclease-like fold families among hypothetical proteins. *Nucleic Acids Res.*, **33**, 3598–3598.
12. Morgan,R.D., Bhatia,T.K., Lovasco,L. and Davis,T.B. (2008) MmeI: a minimal type II restriction-modification system that only modifies one DNA strand for host protection. *Nucleic Acids Res*, **36**, 6558–6570.
13. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
14. Skoglund,C.M., Smith,H.O. and Chandrasegaran,S. (1990) Construction of an efficient overproducer clone of HinfI restriction endonuclease using the polymerase chain reaction. *Gene*, **88**, 1–5.
15. Perler,F.B., Comb,D.G., Jack,W.E., Moran,L.S., Qiang,B., Kucera,R.B., Benner,J., Slatko,B.E., Nwankwo,D.O., Hempstead,S.K. *et al.* (1992) Intervening sequences in an Archaea DNA polymerase gene. *Proc. Natl Acad. Sci. USA*, **89**, 5577–5581.
16. Samuelson,J.C., Zhu,Z. and Xu,S.Y. (2004) The isolation of strand-specific nicking endonucleases from a randomized SapI expression library. *Nucleic Acids Res.*, **32**, 3661–3671.
17. Schildkraut,I. (1984) Screening for and characterizing restriction endonucleases. In Setlow,J.K. and Hollaender,A. (eds), *Genetic Engineering, Principles and Methods*. Vol. 6, Plenum Press, New York, pp. 117–140.
18. Xu,Q., Stickel,S., Roberts,R.J., Blaser,M.J. and Morgan,R.D. (2000) Purification of the novel endonuclease, Hpy188I, and cloning of its restriction-modification genes reveal evidence of its horizontal transfer to the *Helicobacter pylori* genome. *J. Biol. Chem.*, **275**, 17086–17093.
19. Brown,N.L., Hutchison,C.A. III and Smith,M. (1980) The specific non-symmetrical sequence recognized by restriction endonuclease MboII. *J. Mol. Biol.*, **140**, 143–148.
20. Kong,H., Lin,L.F., Porter,N., Stickel,S., Byrd,D., Posfai,J. and Roberts,R.J. (2000) Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res.*, **28**, 3216–3223.
21. Pei,J. and Grishin,N.V. (2007) PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*, **23**, 802–808.
22. Roberts,R.J. (1998) Restriction enzymes. In Hoelzel,A.R. (ed.), *Molecular Genetic Analysis of Populations: A Practical Approach*. Oxford, IRL Press, pp. 379–397.
23. Bennett-Lovsey,R.M., Herbert,A.D., Sternberg,M.J. and Kelley,L.A. (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*, **70**, 611–625.
24. Bath,A.J., Milsom,S.E., Gormley,N.A. and Halford,S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.*, **277**, 4024–4033.
25. Embleton,M.L., Siksnys,V. and Halford,S.E. (2001) DNA cleavage reactions by type II restriction enzymes that require two copies of their recognition sites. *J. Mol. Biol.*, **311**, 503–514.
26. Vanamee,E.S., Santagata,S. and Aggarwal,A.K. (2001) FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.*, **309**, 69–78.
27. Nakonieczna,J., Kaczorowski,T., Obarska-Kosinska,A. and Bujnicki,J.M. (2009) Functional analysis of MmeI from methanol utilizer *Methylophilus methylotrophus*, a subtype IIC restriction-modification enzyme related to type I enzymes. *Appl. Environ. Microbiol.*, **75**, 212–223.
28. Kim,J.S., DeGiovanni,A., Jancarik,J., Adams,P.D., Yokota,H., Kim,R. and Kim,S.H. (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc. Natl Acad. Sci. USA*, **102**, 3248–3253.
29. Barker,D., Hoff,M., Oliphant,A. and White,R. (1984) A second type II restriction endonuclease from *Thermus aquaticus* with an unusual sequence specificity. *Nucleic Acids Res.*, **12**, 5567–5581.
30. Shinomiya,T., Kobayashi,M. and Sato,S. (1980) A second site specific endonuclease from *Thermus thermophilus* 111, Tth111II. *Nucleic Acids Res.*, **8**, 3275–3285.

31. Skowron,P.M., Majewski,J., Zylicz-Stachula,A., Rutkowska,S.M., Jaworowska,I. and Harasimowicz-Slowinska,R.I. (2003) A new *Thermus sp.* class-IIS enzyme sub-family: isolation of a 'twin' endonuclease TspDTI with a novel specificity 5′-ATGAA(N11/9)-3′, related to TspGWI, TaqII and Tth111II. *Nucleic Acids Res.*, **31**, e74–e74.
32. Kong,H. (1998) Characterization of a new restriction-modification system, the BcgI system of *Bacillus coagulans*. *Ph.D. Thesis*. Boston University, pp. 1–130.
33. Cesnaviciene,E.E., Petrusyte,M.M., Kazlauskiene,R.R., Maneliene,Z., Timinskas,A., Lubys,A. and Janulaitis,A. (2001) Characterization of AloI, a restriction-modification system of a new type. *J. Mol. Biol.*, **314**, 205–216.
34. Marchler-Bauer,A., Anderson,J.B., Derbyshire,M.K., DeWeese-Scott,C., Gonzales,N.R., Gwadz,M., Hao,L., He,S., Hurwitz,D.I., Jackson,J.D. *et al.* (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.*, **35**, D237–D240.
35. Dunin-Horkawicz,S., Feder,M. and Bujnicki,J.M. (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**, 98–98.
36. Kelleher,J.E., Daniel,A.S. and Murray,N.E. (1991) Mutations that confer de novo activity upon a maintenance methyltransferase. *J. Mol. Biol.*, **221**, 431–440.