Novel non-specific DNA adenine methyltransferases

Marek Drozdz[1], Andrzej Piekarowicz[1], Janusz M. Bujnicki[2,3] and Monika Radlinska[1,*]

[1]Department of Virology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, [2]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ul. ks. Trojdena 4, 02-109 Warsaw and [3]Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, ul. Umultowska 89, 61-614 Poznan, Poland

ABSTRACT

The *mom* gene of bacteriophage Mu encodes an enzyme that converts adenine to N^6-(1-acetamido)-adenine in the phage DNA and thereby protects the viral genome from cleavage by a wide variety of restriction endonucleases. Mu-like prophage sequences present in *Haemophilus influenzae* Rd (FluMu), *Neisseria meningitidis* type A strain Z2491 (Pnme1) and *H. influenzae* biotype aegyptius ATCC 11116 do not possess a Mom-encoding gene. Instead, at the position occupied by *mom* in Mu they carry an unrelated gene that encodes a protein with homology to DNA adenine N^6-methyltransferases (*hin1523*, *nma1821*, *hia5*, respectively). Products of the *hin1523*, *hia5* and *nma1821* genes modify adenine residues to N^6-methyladenine, both *in vitro* and *in vivo*. All of these enzymes catalyzed extensive DNA methylation; most notably the Hia5 protein caused the methylation of 61% of the adenines in λ DNA. Kinetic analysis of oligonucleotide methylation suggests that all adenine residues in DNA, with the possible exception of poly(A)-tracts, constitute substrates for the Hia5 and Hin1523 enzymes. Their potential 'sequence specificity' could be summarized as AB or BA (where B = C, G or T). Plasmid DNA isolated from *Escherichia coli* cells overexpressing these novel DNA methyltransferases was resistant to cleavage by many restriction enzymes sensitive to adenine methylation.

INTRODUCTION

Bacterial restriction–modification (R-M) systems function as a genetic immune system that cleaves foreign DNA, e.g. phage or plasmid DNA entering the host cell (1). Typically, R-M systems comprise enzymes responsible for two opposing activities: a DNA methyltransferase (MTase) and the restriction endonuclease (REase), which recognize the same sequence. A MTase introduces a specific methylation that protects the DNA against the REase cleavage.

Phage and plasmids employ various strategies to avoid restriction, such as modification of the phage genome, transient occlusion of restriction sites, subversion of host R-M activities and direct inhibition of restriction enzymes (2). Several phage genomes encode enzymes that modify nucleosides in DNA leading to generalized protection of phage DNA within bacterial hosts that carry R-M systems (2–5).

The bacteriophage Mu *mom* gene encodes a protein responsible for the dA′x DNA modification (6). The Mom protein modifies ∼15% of DNA adenine residues in loosely defined target sequences 5′-(C or G)-A-(C or G)-N-(C or T)-3′ (7). Mass spectrometry analyses have suggested that the modified deoxyribonucleoside dA′x corresponds to α-N-(9-β-D-2′-deoxyribofuranosylpurin-6-yl)-glycinamide (8). This unusual modification of DNA is not required for Mu lytic or lysogenic growth and is generally dispensable for phage growth (9). However, the bacteriophage Mu dA′x DNA modification protects the viral genome against cleavage by a wide variety of REases (10). Thus, it serves as a protective measure against nucleolytic attack when the Mu genome infects a cell possessing a DNA 'host specificity' different from that of the bacterium in which the phage replicated.

The expression of *mom* is harmful to the host, so it is strictly controlled and is a late function in the Mu growth cycle, when the host cell is already destined for death (11). The *mom* gene is subject to a series of unusual regulatory controls including the action of the phage-encoded Com protein ('zinc finger-like' translational regulator) (12). The *com* and *mom* genes constitute a single operon located at the right end of the Mu phage genome and the shared *com–mom* gene promoter is positively regulated by the zinc-binding protein Com (13,14).

*To whom correspondence should be addressed. Tel: +48 22 55 41 419; Fax: +48 22 55 41 402; Email: m.radlinska@biol.uw.edu.pl

Other Mu-like phages, such as SP18 (15), often encode homologs of Mom and its regulatory proteins at corresponding positions in their genome, although there are exceptions to this rule. For example, the transposable Mu-like phage B3 of *Pseudomonas aeruginosa* encodes a Com homolog (ORF47), but instead of a homolog of Mom it has a DNA adenine MTase (16). The activity of this ORF48 protein has yet to be demonstrated experimentally.

The determination of the genome sequences for *Haemophilus influenzae* Rd and *Neisseria meningitidis* type A strain Z2491 in 2002 led to the identification of the Mu-like prophages FluMu and Pnm1, respectively (17). A genomic island containing genes related to phage Mu was also discovered in the *H. influenzae* biogroup aegyptius Brazilian purpuric fever (BPF) strain F3031 by McGillivary *et al.* (18). Homologs of the *com* gene were readily identified in all of these prophages. However, at the position corresponding to the *mom* gene within the operons containing the *com*-like genes, the DNA sequences and the encoded proteins showed no similarity to the *mom* gene or the Mom protein, respectively. For example, downstream of the FluMu HI1522.1 gene (whose product is 44% identical to Mu Com), in the position analogous to *mom*, is the HI1523 gene, whose product appears to be unrelated to Mom. The gene *nma1821* encoded by prophage Pnm1 (17), and ORF44 from the genomic island of *H. influenzae* biogroup aegyptius BPF strain F3031 (18) are also located in the same *locus* as Mu *mom*, and their products show 47 and 70% identity to the HI1523 protein, respectively. It has been reported that the FluMu HI1523 protein weakly resembles the *Streptococcus sanguis* DNA m^6A MTase M.StsI (17), but this assertion was based on similarity that is statistically insignificant (a region of 56 residues with only 33% amino acid identity within a protein of 281 residues).

Further, bioinformatic analyses performed by us identified sequence similarity between HI1523 and NMA1821, and a family of DNA adenine-N^6 (m^6A) MTases. All known DNA m^6A MTases belong to the Rossmann-fold MTase superfamily that is structurally and evolutionarily unrelated to the Mom protein that is a member of the acyltransferase superfamily (19). Therefore, we speculated that the HI1523 gene product and its homologs from other prophages modify adenine bases in the DNA, perhaps utilizing a mechanism different from that of the Mom protein. The aim of this study was to test this hypothesis by functionally characterizing these putative MTases.

MATERIALS AND METHODS

Enzymes, chemicals and oligonucleotides

REases, T4 DNA ligase, DNA polymerases, DNaseI, calf intestine alkaline phosphatase (CIP), DNA size standards, unlabeled N^6-methyladenine (m^6A), N^4-methylcytosine (m^4C), C^5-methylcytosine (m^5C) and *S*-adenosyl-L-methionine (AdoMet) were obtained from MBI Fermentas (Vilnius, Lithuania); M.HaeIII, CviQI, TfiI

and NlaIII REases were from New England Biolabs (Ipswich, MA, USA). Snake venom phosphodiesterase I, P1 nuclease and deoxynucleosides were from Sigma. TLC PEI Cellulose F glass plates were from Merck. Chromatography resin was from Sigma (Ni-nitrilotriacetic acid-agarose, Ni-NTA). Synthetic deoxyoligonucleotides (sequences listed in Supplementary Tables S1 and S2) were synthesized by the Institute of Biochemistry and Biophysics, Polish Academy of Sciences (Warsaw, Poland) and Sigma-Aldrich (Poland). [methyl-^{3}H]-AdoMet (2.68 TBq/mmol) was from NEN Life Science Products (Boston, USA). Kits for plasmid preparation, genomic DNA isolation and DNA purification were from A&A Biotechnology (Gdynia, Poland). All enzymes were used following the manufacturer's recommendations.

Bacterial strains, plasmids, media and growth conditions

Escherichia coli strain Top10 (Invitrogen) F⁻ *mcr*A Δ(*mrr-hsd*MRS-*mcr*BC) φ80*lac*Z ΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(*ara*A-*leu*)7697 *gal*U *gal*K *rsp*L *end*A1 *nup*G was used for cloning experiments and strain ER2566 (New England Biolabs, Ipswich, MA, USA) F⁻ λ⁻ *fhuA2* (*lon*) *ompT lacZ::T7 geneI gal sulA11 Δ(mcrC-mrr)114:: IS10 R(mcr-73::mini-Tn10)2 R(zgb-210::Tn10)1* (Tet^s) *endA1 (Dcm)* was used for recombinant protein overproduction. *Escherichia coli* strains were cultured under standard conditions in Luria-Bertani (LB) medium (20). When required, media were supplemented with antibiotics at the following final concentrations: ampicillin (Ap)— $100 \mu g\, ml^{-1}$; kanamycin (Kn)—$50 \mu g\, ml^{-1}$; tetracycline (Tc)—$10 \mu g\, ml^{-1}$. To repress T7 RNA polymerase expression in ER2566 strains before induction, glucose was added to cultures at a final concentration of 1.0%.

The *H. influenzae* biotype aegyptius strain ATCC 11116 (kind gift from Promega to A.P.) and *H. influenzae* Rd (21) were grown at 37°C in 5% CO_2 on brain heart infusion (BHI) medium supplemented with both nicotinamide-adenine dinucleotide and hemin at $2 \mu g\, ml^{-1}$ (sBHI) or on sBHI medium supplemented with mitomycin C at $35\, ng\, ml^{-1}$.

The *hia5* gene from *H. influenzae* biotype aegyptius ATCC 11116 (a homolog of ORF44 from *H. influenzae* biogroup aegyptius BPF strain F3031, AAV37176), the *hin1523* gene from *H. influenzae* Rd (NP_439673) and the *nma1821* gene (YP_002343106) from *N. meningitidis* type A strain Z2491 were cloned from the respective genomic DNAs. The following plasmid vectors were used in this work: pBluescript KS (Ap^r, Stratagene, USA) in cloning experiments, and pET28a and pET30a (Kn^r, Invitrogen, USA) in recombinant protein expression experiments. For the determination of DNA MTase specificity, phage λ DNA (Fermentas, Lithuania) and plasmid pAltOKI (Tc^r) were used. pAltOKI is a derivative of the vector pAlterEx2 (Promega, Madison, USA) obtained by insertion of a synthetic deoxyoligonucleotide duplex (oligonucleotides Pst1TGCA and Pvu1AT, Supplementary Table S1) between the PstI and PvuI sites.

All routine DNA manipulations were carried out according to standard protocols (20). DNA sequencing

was performed at the Institute of Biochemistry and Biophysics, Polish Academy of Sciences (Warsaw, Poland).

Cloning and sequence analysis of the *hia5* gene

The genomic sequence of *H. influenzae* biotype aegyptius ATCC 11116 had not been determined. Assuming similarity between *H. influenzae* biotype aegyptius strains F3031 (18) and ATCC 11116, we designed primers for PCR (8Hf and 8Hr, Supplementary Table S1) based on the sequences flanking ORF44 (GenBank Acc. No. AY647244). The amplified PCR product was cloned into the SmaI site of the vector pBluescriptKS to produce plasmid pHia5KS. The cloned DNA fragment was sequenced and its length was found to be 1332 bp. This sequence has been submitted to the GenBank database under accession number JF268249.

For recombinant expression, the *hia5* gene was amplified by PCR from plasmid pHia5KS DNA using primers 8HfNde and 8HXho (Supplementary Table S1). The obtained fragment (0.85 kb) was double digested with NdeI and XhoI, gel purified, and cloned into NdeI/XhoI-digested expression vector pET30a to add a C-terminal His_6-tag to the encoded protein. To confirm that no mutations had been introduced during the cloning process, the insert of the resulting plasmid pHia5ET was sequenced.

Cloning *nma1821* (ORF NMA1821) and *hin1523* (ORF HI1523)

The procedure for cloning the *hia5* gene was also employed to clone the *nma1821* gene from *N. meningitidis* type A strain Z2491 (primers NmeNde and NmeXho) into expression vector pET30a, and the *hin1523* gene from *H. influenzae* Rd (primers HindNco and HindXho, Supplementary Table S1) into expression vector pET28a. Therefore, C-terminal His_6-tags were added to the encoded Nme1821 and Hin1523 proteins. Subsequently, the efficiency of recombinant Hin1523 protein isolation was found to be lower than for the other two recombinant proteins. Therefore, we subcloned the *hin1523* gene on an NcoI/XhoI fragment into expression vector pET30a to add His_6-tag at the N-terminus. Thus, the final recombinant Hin1523 protein carried His_6-tags on both termini. The resulting plasmids pNmeET and pHinET were verified by sequencing.

Bioinformatic analyses

Sequence searches against the non-redundant protein sequence (nr) database to identify evolutionarily related proteins were performed using PSI-BLAST (22). Expectation (E) values of $<10^{-3}$ were considered to be indication of homology. A multiple sequence alignment of selected homologs of Hia5, Nme1821 and Hin1523 proteins was made using PROMALS (23). Additional similarity searches and the alignment of the Hia5 family with sequences of proteins with known structure were done using the GeneSilico metaserver (http://genesilico .pl/meta2) (24).

Site-directed mutagenesis

The substitution of codons for aspartate with those for alanine within the cloned genes was carried out by PCR according to the ExSite protocol (Stratagene, USA). Oligonucleotide primers (Supplementary Table S1) were used to introduce the desired changes into the coding sequences of recombinant plasmids. The obtained clones were named pHia5D194AET, pNme1821D191AET and pHin1523D194AET.

Expression in *E. coli* and protein purification

All His_6-tagged recombinant enzymes were expressed in *E. coli* strain ER2566. An overnight culture was diluted ($^1/_{100}$) into 500 ml of fresh LB medium supplemented with Kn and incubated at 37°C with shaking to an OD_{600} of 1.0. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and the incubation continued at 20°C for another 4 h. The culture was then centrifuged to harvest the cells and the bacterial pellet resuspended in buffer A (50 mM HEPES buffer, pH 8.0; 300 mM NaCl) supplemented with 1 mM PMSF, 10 mM 2-mercaptoethanol, 10% glycerol and 0.5% Triton X-100. After sonication to lyse the cells, the debris was removed by centrifugation at 40 000*g* for 1 h and the supernatant applied to a 6-ml Ni-NTA Agarose (Sigma) column pre-equilibrated with buffer A containing 20 mM imidazole. The column was washed with buffer A supplemented with 50 mM imidazole and then 70 mM imidazole. The recombinant protein was eluted from the column with buffer A containing 150 mM imidazole. The purity of the isolated protein was determined by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE).

DNA methylation assay

The DNA modification activities of the purified recombinant enzymes were assayed in a 30-μl reaction with 1 μg of λ DNA dam− dcm− at 37°C for 1 h in buffer M (10 mM Tris-HCl, pH 7.5, 10 mM EDTA, 5 mM 2-mercaptoethanol, 1 mg ml^{-1} bovine serum albumin (BSA)) and 2 μM [^{3}H]AdoMet. The reaction was terminated by the addition of 0.5% SDS and the mixture was spotted onto 20 mm × 20 mm DE81 filter paper discs (Whatman, Brentford, UK). The filters were air-dried and then washed three times (10 min each) in a large volume of 50 mM KH_2PO_4, once in pure water, and once in 70% ethanol. Incorporation of the ^{3}H-labeled methyl group was measured using a liquid scintillation counter (Wallace, Pharmacia).

For kinetic analysis, oligonucleotide duplexes were used as substrates in methylation assays to assess the sequence specificity of the enzymes. The complementary oligonucleotides (45 μM) were mixed, heated at 90°C for 5 min and then slowly cooled to 25°C to allow duplex formation (Supplementary Table S2). The reaction mixtures contained 1 μM oligo duplex, 6 μM [^{3}H]AdoMet, 100 μM unlabeled AdoMet and 106 nM recombinant protein in buffer M. Therefore, the AdoMet concentration exceeds more than 2-fold of the concentration of potential substrate adenine residues

in the TA50 duplex (10 times in the CA10 and GA10 duplexes, 5 times in the AA21 duplex, 106 times in Cm2A1, 35 times in Cm2A3). Reactions were incubated at 37°C, aliquots removed at specified time periods, mixed with 0.5% SDS and processed as described above. Each methylation experiment was carried out at least in triplicate. Initial velocities of reactions were calculated from the slope of reaction progress curves.

Thin-layer chromatography to identify the methylated base

The identity of the base methylated by the studied MTases was determined as described previously (25) using λ DNA dam− dcm− and [methyl-^{3}H]AdoMet.

REase cleavage of DNA methylated by the Hia5, Nme1821 and Hin1523 enzymes

The DNAs of plasmids pHia5ET, pNmeET and pHinET isolated from IPTG-induced and uninduced bacterial cultures were used as substrates for cleavage by selected restriction enzymes. Typically, 1 μg of DNA was digested with 10 U of REase for 4 h in a 20 μl reaction in appropriate buffer conditions, as suggested by the manufacturer. For enzymes with only one recognition site in the plasmid DNA, additional digestion with an enzyme not sensitive to adenine methylation was carried out to linearize the DNA. Sequences recognized by REases HindIII, PstI, EcoRI and Eco32I were absent from pHia5ET, so in these cases, plasmid pAltOKI was introduced into strain ER2566 (pHia5ET), and the plasmid preparation isolated from the double transformants was examined as described above. Digestion products were separated by electrophoresis on 0.8% agarose gels, which were then stained with ethidium bromide and photographed on a UV transilluminator.

The effect of *in vitro* enzyme activity toward GATC sites was determined using a protection assay in which λ DNA dam− dcm− was incubated with the recombinant proteins and then the DNA was challenged with REases. A typical assay reaction was a 100 μl mixture containing 80 μM AdoMet, 10 μg DNA and purified recombinant protein (∼0.9 μg) in buffer M. The reaction mixture was incubated at 37°C for 60 min, extracted with phenol/chloroform and the DNA was precipitated with ethanol. The DNA was then cleaved separately with REases MboI, DpnI and Bsp143I in the appropriate buffers at 37°C for 4 h. In control reactions λ DNA dam− dcm− or λ DNA dam+ dcm+ were used to assess the activity of each enzyme. Plasmid or λ DNA without added endonuclease was used as negative controls.

HPLC DNA methylation assay

The substrate λ DNA dam− dcm− (10 μg) was methylated as described above for 3 h. Following phenol/chloroform extraction and DNA precipitation, samples were incubated with 1 U DNaseI in 10 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 50 mM NaCl and 0.1 mg/ml BSA, for 1 h at 37°C in a total volume of 30 μl. The DNaseI was then heat-inactivated and the mixtures were treated with P1 nuclease (2.5 U) and CIP (1 U) overnight at 37°C. A volume of 20 μl of each mixture

was applied to a dC18 Atlantis column (3.0 × 150 mm, 5 μM particle size, Waters Associates Inc.) equilibrated with buffer E (50 mM KH$_2$PO$_4$ adjusted to pH 5.5 with KOH). High performance liquid chromatography (HPLC) runs were performed at a flow rate of 0.9 ml min^{-1} at ambient temperature using an isocratic mixture of buffer E and methanol (85 : 15). The column eluate was monitored by measuring absorbance at 260 nm. The obtained chromatograms were analyzed with the Empower^{TM} program (Waters Associates Inc.). All samples were analyzed in quadruplicate. The same procedure was used for quantification of the base composition of chromosomal DNA isolated from the *H. influenzae* biotype aegyptius ATCC 11116 strain (5 μg). As controls, λ DNA dam− dcm− and λ DNA dam+ dcm+ were analyzed in the same way.

In order to calibrate the retention times of different nucleosides, six standards containing 400 pmol of the nucleosides dC, dT, dA, and dG, dmA and dmC (obtained by treating dm^{5}CTP and dm^{6}ATP with alkaline phosphatase under the same conditions as the DNA samples) were also analyzed (10 replicates of each). The average HPLC peak areas of the samples were compared with those of the standards and the mole percents calculated.

Determination of the genomic DNA methylation status of *H. influenzae* biotype aegyptius ATCC 11116 and *H. influenzae* Rd by restriction digestion

Genomic DNA was isolated from cells from the cultures of both *H. influenzae* strains that were either untreated or treated with mitomycin C, using a Genome-DNA kit (A&A Biotechnology). DNA (1 μg) was incubated with 10 U of the selected REases at 37°C for 4 h. The extent of cleavage was assessed by agarose gel electrophoresis.

RESULTS

Identification of Mu-like sequences in *H. influenzae* biotype aegyptius strain ATCC 11116

The 1332 bp sequenced fragment of *H. influenzae* biotype aegytptius strain ATCC11116 (for cloning see 'Materials and Methods' section) showed 86% identity to the DNA sequence between nucleotides 34 230 and 35 558 of the genomic island identified in the *H. influenzae* biogroup aegyptius BPF strain F3031 (Acc no. AY647244) (Supplementary Figure S1). Inspection of the sequence of this 1332-bp fragment revealed two open reading frames (ORFs), named *comHia* and *hia5*, located between nucleotides 232–348 and 413–1255, respectively (Supplementary Figure S2). The *comHia* translation product was 98% identical to the protein encoded by ORF43 of *H. influenzae* biotype aegytptius strain F3031 (located on the genomic island) (18) and also shared sequence similarity with Mu Com (49% identical over its entire length of 39 residues). Moreover, the putative ComHia protein contained four cysteines (positions 9, 12, 29 and 32) in analogous positions to Mu Com (Supplementary Figure S3), where these residues (Cys6, 9, 26 and 29) are required to form the zinc finger (12). This similarity suggested that ComHia performs the
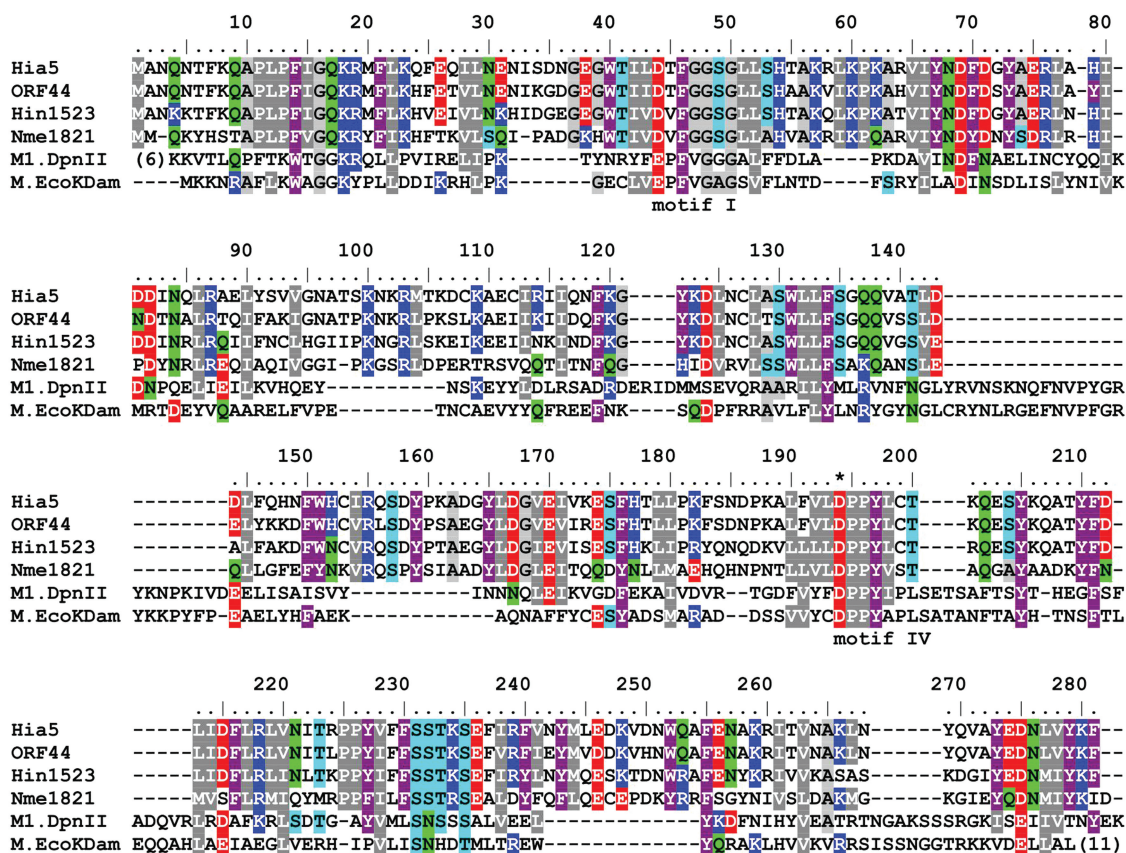
Figure 1. Sequence alignment of the Hia5, ORF44, Hin1523, Nme1821 proteins and two DNA MTases with known crystal structures: M1.DpnII (2dpm) and M.EcoKDam (2g1p). The number of terminal residues omitted for clarity is shown in parentheses (). Conserved motifs are indicated beneath the alignment and conserved residues are highlighted by shading (aliphatic in grey, aromatic in purple, positively charged in blue, negatively charged in red, with amide groups in green, with hydroxyl groups in cyan). The catalytic aspartate residue substituted by alanine is indicated by an asterisk.

same function as Mu Com. The predicted protein product of the *hia5* gene was 281 amino acids long and shared 80% identity with ORF44 of the *H. influenzae* biotype aegyptius strain F3031 genomic island, 69% with Hin1523 and 43% with Nme1821 (Figure 1).

Protein expression and purification

Recombinant Hia5, Nme1821and Hin1523 proteins were expressed in *E. coli* ER2566 cells carrying the plasmid constructs pHia5ET, pNmeET and pHinET, respectively. These His-tagged proteins were purified using metal affinity chromatography ('Materials and Methods' section). The apparent molecular masses of the recombinant proteins (34 kDa for Hia5, 33.6 kDa for Nme1821 and 38 kDa for Hin1523; Supplementary Figure S4) were in good agreement with those predicted from the nucleotide sequences.

Analysis of the amino acid sequences of Hia5, Nme1821 and Hin1523 proteins

The deduced amino acid sequences of the proteins Hia5, Nme1821 and Hin1523 were used as queries in PSI-BLAST searches of the National Center for Biotechnology Information (NCBI) nr database. The highest similarity scores over the full length of the sequences were with a large number of uncharacterized proteins annotated as putative MTases, followed by *bona fide* adenine MTases with lower, but still significant scores, regardless of the relatively low-level sequence identity (e.g. MboIA with the e-value 4e-7 and 15% identity, reported in a third iteration when using Hia5 as the query). Interestingly, the ORF48 of *P. aeruginosa* phage B3, which encodes a DNA adenine MTase homolog (16), was not detected among the 1000 most similar sequences. The homology between the Hia5 family and Dam-related MTases was further supported by the protein fold-recognition methods run via the GeneSilico metaserver (24). For example, using the HHsearch algorithm (26), *E. coli* Dam (M.EcoKDam), *Streptococcus pneumoniae* DpnM (M1.DpnII), and T4 Dam (M.EcoT4Dam) were identified with 100% probability as the closest homologs of Hia5 with known structure, and all consensus predictors indicated this prediction as the most probable (data not shown).

In agreement with their overall sequence similarity to known DNA adenine MTases, the Hia5, Nme1821 and Hin1523 sequences were found to contain the characteristic motif DPPY, also called motif IV (Figure 1), which is typical for MTases that modify amino groups in various substrates. The importance of this motif in the catalytic

activity of DNA m^{6}A MTases has been confirmed experimentally (27). In addition to the conserved DPPY motif, these enzymes possessed a well-conserved FxGxG motif, also called motif I, which is involved in AdoMet binding (Figure 1). However, the region corresponding to the target recognition domain in Hia5 (residues ~80–170) appeared highly divergent compared with those of known DNA MTase structures, which hampered the modeling of a Hia5–DNA complex and detailed prediction of the protein–DNA interactions.

Identification of the base methylated by the Hia5, Nme1821 and Hin1523 enzymes

In *in vitro* assays, recombinant Hia5, Nme1821 and Hin1523 were shown to incorporate the [^{3}H] methyl group from AdoMet into the substrate DNA. However, variants in which the DPPY motif had been altered to APPY by site-directed mutagenesis were inactive (Table 1). This result confirmed that these proteins exhibit DNA MTase activity. Of the three enzymes tested, Hin1523 displayed the lowest activity. The chemical identity of the methylated base produced by the DNA MTase activity of these enzymes was identified using thin-layer chromatography. Tritium label was present only in the position corresponding to the m^{6}dA standard (Table 2), which implied that all of the analyzed MTases modify adenine at the N6 position to form m^{6}A.

Table 1. Efficiency of *in vitro* methylation of phage λ DNA by M.HaeIII, Hia5, Nme1821 and Hin1523 enzymes and their altered catalytic motif IV (APPY) variants

Sample	Methylation level (dpm)
Negative control: no enzyme	130
Negative control: enzyme heated to 80°C for 15 min before reaction	150
M.HaeIII	19 000
Hia5	18 000
Hia5D194A	226
Nme1821	17 000
Nme1821D191A	231
Hin1523	11 000
Hin1523D194A	195

Table 2. Identification, by thin-layer chromatography, of the base methylated by the Hia5, Hin1523 and Nme1821 enzymes

Enzyme	^{3}H incorporation (dpm)			
	Solvent G[a]		Solvent D[b]	
	m^{6}dA	m^{4}dC + m^{5}dC	m^{6}dA + m^{4}dC	m^{5}dC
Hia5	5820	27	5700	22
Nme1821	5670	20	5595	20
Hin1523	5033	17	5001	25

[a]m^{4}dC and m^{5}dC comigrate in solvent G.
[b]m^{6}dA and m^{4}dC comigrate in solvent D.

Assessment of the sequence specificity of the Hia5, Nme1821 and Hin1523 enzymes using the endonuclease protection assay

To define the sequence specificity of the three novel MTases, we used a panel of adenine methylation-sensitive endonucleases in a REase digestion assay. It was anticipated that the presence of methylated adenine, introduced by the studied proteins, within a REase recognition site, would interfere with cleavage and therefore change the pattern of cleaved fragments.

First, we tested plasmids isolated from the *E. coli* ER2566 strain carrying the gene encoding the Hia5 protein for susceptibility to cleavage by enzymes sensitive or insensitive to methylation of adenine residues. Plasmid DNAs were prepared from the bacterial host, either induced for 4 h with 1 mM IPTG (pHia5ETi DNA) or not induced (pHia5ET DNA), and then subjected to digestion by REases. According to REBASE (28), 30 of the REases tested are blocked by adenine methylation at their cleavage site. Another 18 REases, insensitive to m^{6}A, were used as controls to confirm the susceptibility of the substrate DNAs to digestion. The results of digestions by different restriction enzymes as an indicator of modification introduced by the Hia5 enzyme are presented in Table 3 (Supplementary Figure S5).

Seven REases that do not have adenine in their target sequence, and eight others with only cytosine or guanine residues in the defined part of degenerate sequences (e.g. Bme1390I CCNGG or BglI GGCN$_5$GCC) cleaved all DNA substrates tested. Likewise, three other REases (Bsp68I, Bsp143I and PvuI) that are known to be insensitive to adenine methylation in their cognate sequences, produced the expected restriction pattern. In contrast, all 30 REases with previously identified sensitivity to the presence m^{6}A in their recognition sequences failed to cleave pHia5ETi DNA, or partial cleavage was observed with a large proportion of DNA fragments corresponding to linearized plasmid. The pHia5ET plasmid DNA isolated from non-induced cells was cleaved by all restriction enzymes tested.

The *E. coli* ER2566 strain contains EcoKDam MTase that modifies the adenine residue in the sequence GATC. To determine whether GATC sequences also represent a substrate for Hia5 modification, λ DNA dam− dcm− was methylated *in vitro* using Hia5. The status of this methylation was then tested by incubating the treated DNA with an excess of the following REases: DpnI (requires adenine methylation of GATC sites for cleavage), MboI (inhibited by m^{6}A methylation) and Bsp143I (cleaves both methylated and unmethylated GATC sites, but not methylated sites when the methylation is on the C). The extent of digestion of Hia5-modified λ DNA by these REases showed that virtually all GATC sites were methylated (Supplementary Figure S6).

Experiments analogous to those described above were performed for the Nme1821 and Hin1523 enzymes (Supplementary Tables S3 and S4). The cleavage pattern for Nme1821-methylated DNA was very similar to that for Hia5, whereas in the case of Hin1523, the pHinETi

Table 3. Summary of the sensitivity of different REases to DNA modifications introduced by the Hia5 enzyme

REase	Recognition site	Number of sites in pHia5ET	Sensitivity to m^6A	Cleavage
BglII	AGaTCT	1	S	N
Bsh1285I	CGRYCG R = A/G	5	S	N
Bsu15I	ATCGAT	2	S	N
CviQI	GTAC	3	S	N
DraI	TTTAAA	2	S	N
Eco32I	GATATC	0 (1)[b]	S	N
EcoRI	GAATTC	0 (1)[b]	S	N
HincII	GTYRAC	1 (GTTAAC)	S	N
HindIII	AaGCTT	0 (1)[b]	S	N
HinfI	GANTC	19	S	N
HphI	GGTGA/TCACC	16	S	N
KspAI	GTTAAC	1	S	N
MboI[a]	GATC	24	S	N
MluI	db ACGCGT	1	S	N
MlyI	GAGTC/GACTC	9	S	N
MunI	CAATTG	1	S	N
NlaIII	CATG	24	S	N
PaeI	GCATGC	1	S	N
PstI	CTGCAG	0 (2)[b]	S	N
PvuII	dbCAGCTG	3	S	N
SspI	aATATT	3	S	N
VspI	ATTAAT	5	S	N
XbaI	TCTaGA	1	S	N
XhoI	CTCGAG	1	S	N
TaiI	ACGT	16	S	N
TaqI	TCGA	16	S	N
TfiI	GAWTC	10	S	N
AluI	AGCT	24	S	P, L
TasI	AATT	28	S	P, L
Tru1I	TTAA	29	S	P, L
CaiI	CAGNNNCTG	1	U	N
CseI	GACGC	12	U	N
Eam1104I	GAAGAG	3	U	N
Eco147I	AGGCCT	1	U	N
Eco47III	AGCGCT	3	U	N
Eco91I	GGTNACC	1	U	N
LweI	GCATC	24	U	N
Mph1103I	ATGCAT	3	U	N
NdeI	CaTATG	1	U	N
PfeI	GAWTC	10	U	N
PscI	ACATGT	1	U	N
SchI	GAGTC/GACTC	9	U	N
XmiI	GTMKAC	1	U	N
XapI	RAATTY	6	U	N
AdeI	CACNNNGTG	1	U	N
BpiI	GAAGAC	5	U	N
Bsp119I	TTCGAA	1	U	N
BcnI	CCSGG	12	(−)	Y
BglI	GGCN$_5$GCC	1	(−)	Y
Bme1390I	CCNGG	21	(−)	Y
Bpu1102I	GCTNAGC	2	U	Y
Bsh1236I	CGCG	37	(−)	Y
Bsp143I[a]	GATC	24	Not	Y
Bsp120I	GGGCCC	1	(−)	Y
Bsp68I	TCGCGA	1	Not	Y
BspLI	GGNNCC	20	(−)	Y
BspOI	GCTAGC	0 (2)[b]	U	Y
Bst1107I	GTATAC	1	U	Y
BsuRI	GGCC	22	(−)	Y
Csp6I	GTAC	3	U	Y
DpnI[a]	Gm^6ATC	24	Not	Y
Eco130I	CCWWGG	1	(−)	Y
Eco88I	CYCGRG	2	(−)	Y
EcoO109I	RGGNCCY	3	(−)	Y
FspBI	CTAG	6	U	Y
Hin1I	GRCGYC	5	(−)	Y

(continued)

Table 3. Continued

REase	Recognition site	Number of sites in pHia5ET	Sensitivity to m^6A	Cleavage
Hin6I	GCGC	32	(−)	Y
HpyF10VI	GCN$_7$GC	41	(−)	Y
MbiI	GAGCGG/CCGCTC	4	U	Y
MspI	CCGG	29	(−)	Y
NheI	GCTAGC	0 (2)[b]	U	Y
PvuI	CGATCG	1	Not	Y
SmaI	CCCGGG	1	(−)	Y

[a]Tested in λ DNA dam- *in vitro* methylated with the Hia5 enzyme.
[b]number of recognition sites in plasmid pAltOKI.
Y: complete cleavage; N: no digestion; P: partial digestion, L: plasmid linearized; S: sensitive to m^6A; U: unknown sensitivity to m^6A; Not: not sensitive to m^6A; (−): absence of adenine residues in recognition sequence; **db**: methylation of both strands required to block cleavage; **A**: the enzyme will not cut if the **marked** adenine is methylated; a: the enzyme is not sensitive to methylation of the marked adenine; A: there are no data on the sensitivity of the enzyme to methylation of the unmarked adenine.

plasmid DNA was more sensitive to REase cleavage than the pNmeETi or pHia5ETi DNAs.

## Quantitative determination of m^6A in DNA methylated by Hia5 or Hin1523

The results described above indicated that m^6A was introduced in many nucleotide contexts. An HPLC DNA methylation assay was conducted to evaluate the extent of adenine modification resulting from the activity of the Hia5 and Hin1523 enzymes. The results of this analysis (Figure 2, for control DNAs see Supplementary Figure S7) clearly showed that the tested enzymes catalyzed extensive methylation of adenine residues in the substrate λ DNA. In contrast, identical analysis of λ DNA treated with the Hia5D194A variant did not reveal any dmA content, which confirmed the loss of DNA MTase activity in the Hia5D194A protein (Figure 2).

The HPLC assay results also allowed calculation of the quantity of adenine residues transformed into m^6A (Table 4) by the studied enzymes: Hia5 methylated >60% of adenine residues in λ DNA and Hin1523 methylated 29.6%.

As summarized in Table 4, the %GC content measured for λ DNA that had been *in vitro* methylated with the tested enzymes was in good agreement with reference data, i.e. 49.8% GC. The lower level of m^6A in λ DNA dam+ dam+, than expected from the number of dam target sites (0.46 versus 0.92%), probably reflected undermethylation of the substrate GATC sites by M.EcoKDam. It has previously been demonstrated that only ∼50% of Dam sites in λ DNA are methylated, presumably because the MTase does not have the opportunity to fully methylate the DNA before it is packaged into the phage head (29,30).

Sequence preferences of MTases

The observed high level of adenine methylation strongly suggested that the Hia5 and Hin1523 enzymes have
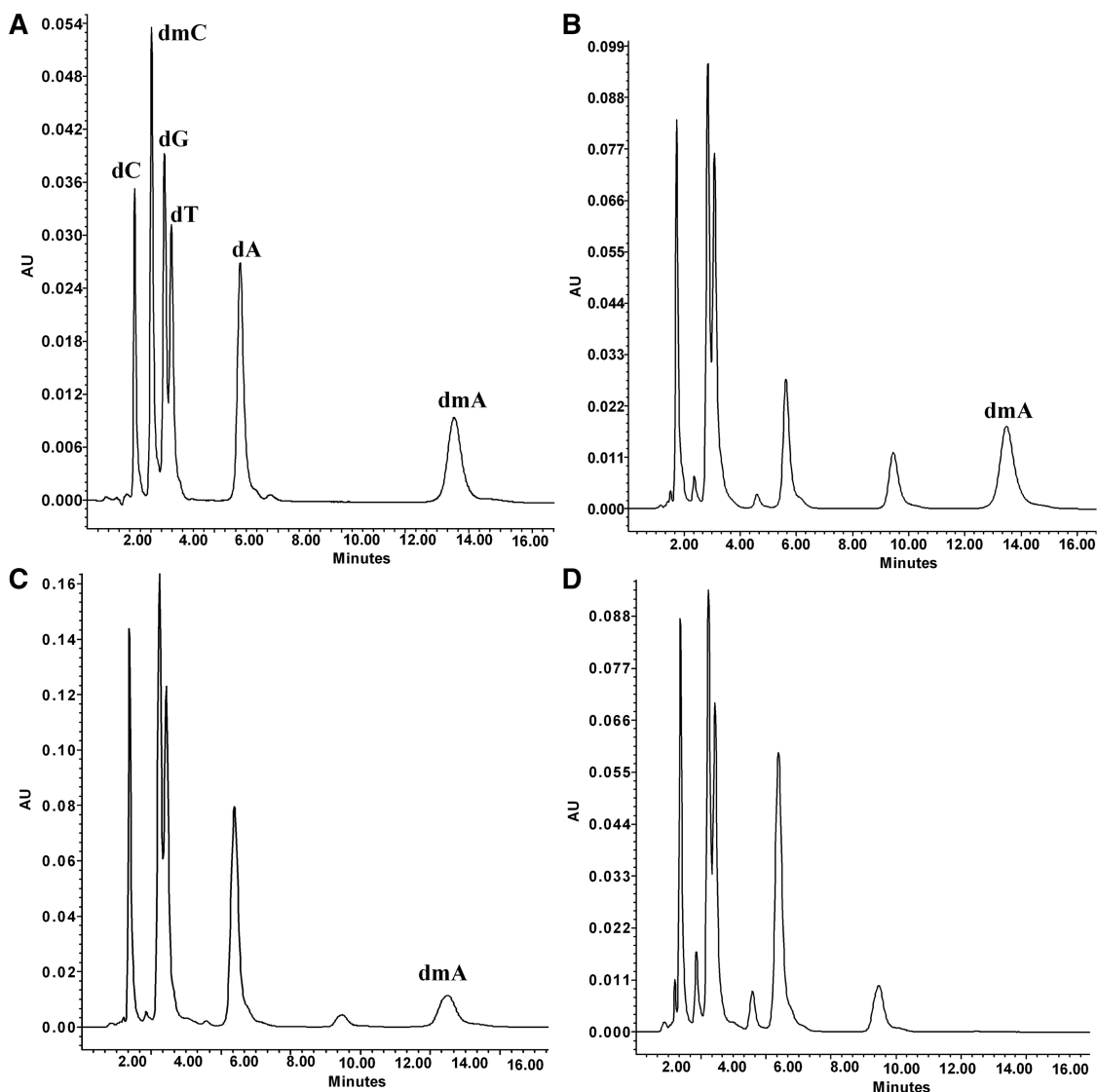
Figure 2. HPLC profiles of deoxynucleosides in the standard mixture (**A**), and the enzymatic digests of Hia5-methylated λ DNA (**B**), Hin1523-methylated λ DNA (**C**) and Hia5D194A-methylated λ DNA (**D**). In digests of Hia5- and Hin1523-methylated λ DNA, in addition to major deoxynucleoside components, a peak was eluted at 13.7 min, corresponding to that of dmA. In the Hia5 and Hin1523 HPLC profiles and also those of Hia5D194A and λ DNA dam+ dcm+ (Supplementary Figure S7), we detected an unknown peak eluting at 9.5 min, but were unable to assign it to any obvious chemical species.

Table 4. Percentages of the base content of λ DNA (control or methylated with Hia5, Hia5D194A or Hin1523 proteins) and total DNA of *H. influenzae* biotype aegyptius ATCC 11116 as determined by the HPLC DNA methylation assay

DNA	dC	dmC	dG	dT	dA	dmA	dmA/ΣA	%GC
λ dam− dcm−	23.60	0.00	26.10	25.30	25.00	0.00	0.0	49.70
λ dam+ dcm+	23.60	0.04	25.88	24.58	25.78	0.12	0.46	49.52
λ methylated by Hia5	23.90	0.00	25.40	24.00	10.20	16.50	61.79	49.30
λ methylated by Hia5D194A	25.00	0.00	25.10	24.60	25.20	0.00	0.00	50.20
λ methylated by Hin1523	23.40	0.00	24.90	24.70	19.00	8.00	29.63	48.30
Total DNA of *H. influenzae* biotype aegyptius ATCC 11116	21.00	3.70	19.20	26.80	28.80	0.50	1.70	43.90

ΣA = dA + dmA. Values presented are the means obtained using several independent DNA preparations.

minimal sequence specificity. Therefore, instead of using the classical approach to identify multiple methylated sequences followed by the search for consensus between them we have used a set of oligonucleotide duplexes as the substrates. Each of the duplexes contained repetitions of a dinucleotide in which an adenine was accompanied by another nucleotide (CA, GA, TA and AA). Examples of methylation reaction are shown in
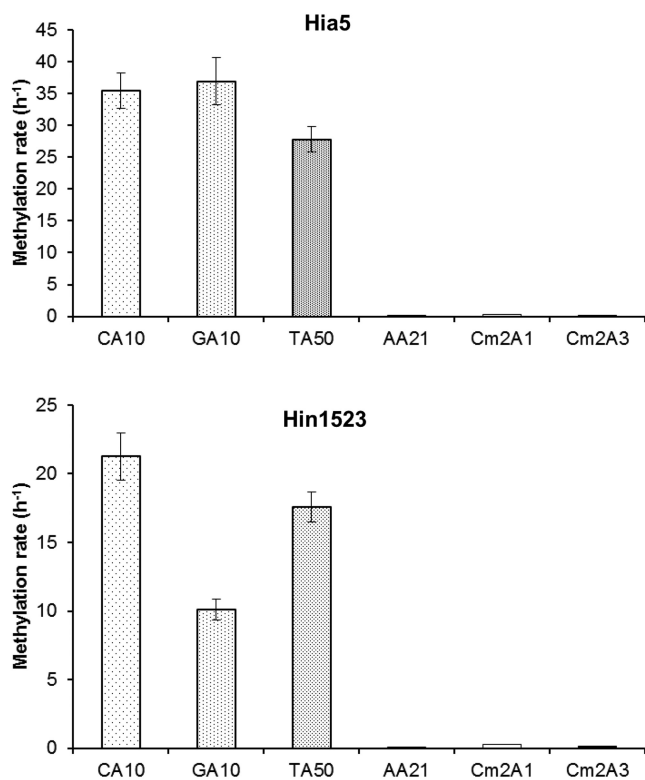
Figure 3. Comparison of initial velocities of methylation reactions of oligonucleotide duplexes (Supplementary Table S2) carried out by the Hia5 and Hin1523 enzymes.

Supplementary Figure S8. Increase of radiolabel incorporation was linear for ∼30 min.

Substrates with repetitions of the dinucleotides CA and GA (CA10, GA10, respectively) were methylated with comparable rates by Hia5, and TA was methylated slightly slower (Figure 3). The dinucleotide CA was methylated by the Hin1523 enzyme with the highest rate, TA was methylated slightly slower and GA two times slower than two others (Figure 3). The poly-dA/poly-dT duplex was not methylated by either the Hia5 or Hin1523 enzymes. To exclude the nature of the sequence (tract character) and potential duplex instability as the cause of this negative result, we tested substrates containing one or three adenines flanked by m^6As i.e. m^6A $\underline{A}$ m^6A (Cm2A1) and m^6A $\underline{AAA}$ m^6A (Cm2A3). These duplexes were also not methylated by Hia5 and Hin1523.

Hia5, Nme1821 and Hin1523 enzymes used as tools to test the sensitivity of REases to adenine methylation in their recognition sequences

Due to the extensive methylation of adenines introduced by the activity of Hia5, Nme1821 and Hin1523, these enzymes might be useful as tools in testing the sensitivity of REases to adenine methylation in their recognition sequences. We examined the activity of 23 commercially available REases, for which sensitivity to m^6A had not been established, following methylation of their target sequences by Hia5, Hin1523 or Nme1821 (Table 3 for Hia5, Supplementary Table S3 for Nme1821 and

Supplementary Table S4 for Hin1523). Sixteen of the tested REases were unable to cleave plasmid pHia5ETi DNA isolated from the strain overexpressing the Hia5 protein. These results were particularly valuable for nine REases that possess only one adenine residue in their recognition sequence (AdeI, CaiI, CseI, Eco147I, Eco47III, Eco91I, LweI, PfeI and SchI), which had to be the target for Hia5 methylation. It has previously been shown that NdeI is insensitive to methylation of the first adenine within its recognition sequence (C$\underline{A}$TATG), but its sensitivity to methylation of the second adenine was unknown. NdeI did not cut Hia5-methylated DNA, so it may be suggested that the Hia5 enzyme modified the second adenine within the CAT$\underline{A}$TG sequence and as a result, NdeI lost its ability to recognize this target. Seven REases with unknown sensitivity to m^6A that did not cleave DNA methylated by Hia5 contain more than one adenine residue in their recognition sequence and it is not certain which position interfered with cleavage when methylated. Another seven REases digested Hia5 methylated DNA to produce the expected restriction pattern. Since our results demonstrated that Hia5 can modify adenines in many nucleotide contexts, it may be concluded that these endonucleases are indeed insensitive to adenine methylation. Lastly, cleavage by CviQI was blocked when the adenine in its recognition site GTAC was methylated. However, REase Csp6I, an isoschizomer of CviQI, was not sensitive to adenine methylation in target sites.

m^6A status in *H. influenzae* genomic DNA

Genomic DNAs of *H. influenzae* biotype aegyptius ATCC 11116 and *H. influenzae* Rd isolated from regular cultures and cultures treated with mitomycin C were incubated with an excess of REases and then the extent of cleavage was assessed by agarose gel electrophoresis. Treatment of bacterial cultures with mitomycin C had no apparent effect on the restriction patterns of the genomic DNAs (data not shown). Each DNA was found to be resistant to digestion with MboI, while at the same time cleaved by DpnI, which was as expected, because these bacterial strains contain Dam activity and therefore, GATC sites were modified. Bsp143II did not digest the genomic DNA of *H. influenzae* biotype aegyptius ATCC 11116 on account of the presence of M.HaeII activity (31) and HindIII did not digest the genomic DNA of *H. influenzae* Rd due to M.HindIII activity in this strain (32). The genomic DNAs were sensitive to all other enzymes used in this experiment (e.g. HinfI, CviQI and Csp6I) (Supplementary Figure S9—data shown only for genomic DNA of *H. influenzae* biotype aegyptius ATCC 11116 isolated from a regular culture).

The results described above are in good agreement with the HPLC analysis of the m^6A content in the *H. influenzae* biotype aegyptius ATCC 11116 chromosomal DNA, which was calculated to be 1.7%. The m^6A detected in this experiment was probably the result of the Dam MTase activity (Table 4, Supplementary Figure S7). Taken together, these data indicate that the Hia5 and Hin1523 enzymes are normally inactive in the respective

Haemophilus strains, even in cultures treated with mitomycin C.

DISCUSSION

Since the annotation of the Mu genome in 2002 (17), Mu-like transposable phage family members have been identified during the sequencing of several bacterial genomes. Comparison of these Mu-like phages, prophages and phage-related elements indicates that their genome organizations are similar.

The 1332-bp fragment amplified from genomic DNA of the *H. influenzae* ATCC 11116 strain analyzed in this study was highly similar (86% identity) to a segment of the *H. influenzae* F3031 genomic island containing genes related to phage Mu (18). This high level of sequence similarity indicated that this ATCC 11116 genomic fragment represents part of a complete or defective Mu-like prophage present in this bacterial host.

Mu-like prophages often carry genes homologous to those of phage Mu and also other apparently unrelated genes that appear to perform analogous functions (17). Two of ORFs (*nma1821*, *hin1523*) characterized in this study are present at the right end of two Mu-like prophage genomes (17), and the third (*hia5*) is predicted to be a part of a Mu-like sequence. All three localize in the prophage genomes at positions syntenic to the *mom* gene of Mu. The putative protein products of the *hia5*, *hin1523* and *nma1821* genes also share sequence similarity, but are not homologous to the Mom protein of Mu—the modifying enzyme that converts adenine residues in DNA to acetamidoadenine. In this study, we cloned, expressed and characterized the *hia5*, *hin1523* and *nma1821* genes and identified their products as novel DNA MTases with extremely relaxed specificity.

According to the nature of the reactions they catalyze, DNA MTases can be divided into three separate groups, generating m^6A, m^4C or m^5C. All known DNA MTases have conserved motifs, of which I–VIII and X are common to most subfamilies (33). X-ray crystallographic analyses have revealed that the region containing the conserved motifs corresponds to the structurally and evolutionarily conserved catalytic domain, in which motifs X and I–III are involved in AdoMet binding and motifs IV–VIII form the active site. The common AdoMet binding region is conserved across the whole MTase superfamily, while the 'catalytic motifs', especially motif IV, exhibit subfamily-specific patterns (33,34).

The enzymes that were the subject of this study belong to a previously uncharacterized subfamily of DNA adenine MTases. Despite their relatively distant sequence similarity to characterized MTases, they possess the characteristic signature motifs, including the catalytic motif IV (DPPY), which is typical for MTases that modify amino groups in various substrates, including RNAs and proteins. We have shown that *in vitro* these enzymes catalyzed the transfer of methyl groups from an AdoMet donor to the substrate DNA, and that they modified only adenine residues, converting them into m^6A. To further confirm that the Hia5, Hin1523 and

Nme1821 proteins act as adenine MTases, we altered their predicted catalytic sites by site-directed mutagenesis. Substitution of aspartate in the DPPY motifs with alanine (point substitution D194A for Hia5 and Hin1523, and D191A for Nme1821) completely abolished the adenine MTase activity of these enzymes.

To define the target sequences of these novel MTases, we used an endonuclease protection assay. All 30 REases with previously established sensitivity to m^6A in their recognition sequences were unable to cleave DNAs isolated from strains expressing the tested genes or only a partial cleavage was observed. On the contrary, the control REases, which are insensitive to m^6A, were able to cleave the DNAs to produce the expected restriction patterns. These results were the first indication of the massive DNA methylation catalyzed by the Hia5, Hin1523 and Nme1821 enzymes. Corroborating evidence was obtained using an HPLC DNA methylation assay to evaluate the base composition of λ DNA methylated with Hia5 *in vitro*. This analysis revealed that as much as 61% of the adenine residues were converted to m^6A (29.6% by Hin1523). In other words more than every second adenine in λ DNA was modified by the Hia5 enzyme, which strongly suggested that a predicted sequence specificity cannot be longer than a dinucleotide. To directly test this assumption, we used substrates with repetitions of the dinucleotides CA, GA, TA and AA (poly-A tract). The first two were methylated by the Hia5 enzyme with comparable rates and TA was methylated slightly slower. This slight difference could be associated with the nature of the substrate. Adenines are present in both the upper and lower strands of the TA50 duplex, unlike in the CA10 and GA10 duplexes, where they are present only in one strand. It is possible that only one of the two diagonally arranged adenines in the TA duplex is methylated efficiently, while the other may be methylated at a lower rate.

Hin1523 also methylated the aforementioned duplexes, although the rate of CA10 methylation was the fastest. The TA50 duplex was modified by Hin1523 slightly slower similar to the Hia5 enzyme. Weak discrimination against the GA10 duplex observed for the Hin1523 does not exclude GA dinucleotides as substrates, since the GA10 duplex methylation rate was only two times lower than the rates measured for the CA10 and TA50 duplexes.

The incorporation signal obtained for the AA21 substrate (poly(A)/poly(T) tract) was at the background level. Adenines occurring in the Cm2A1 and Cm2A3 duplexes also appeared to be poor substrates for both the Hia5 and Hin1523 enzymes, suggesting that internal adenines within poly(A)-stretches are effectively not methylated.

Our results suggest that all adenine residues in double-stranded DNA, with the exception of poly(A)-tracts, constitute potential substrates for the Hia5 and Hin1523 enzymes, and that these enzymes are indeed sequence non-specific. Formally, their potential 'sequence specificity' could be summarized as AB or BA (where B = C, G or T). 71% of adenines in λ DNA occur in AB (BA) sequences. Sixty-one percent level of Hia5 λ DNA methylation in HPLC assay is close to this value.

To our knowledge, this is the first example of an exocyclic amino MTase with such extraordinary sequence promiscuity. High levels of adenine residue methylation *in vivo* and *in vitro* in a heterologous host (*E. coli*) have been described for M.CviQII that modifies the dinucleotide sequence A̲G (35), but this was demonstrated only by a restriction protection assay and the extent of M.CviQII modification has never been quantified.

REases are commonly used as molecular biology tools and their sensitivity to methylation has to be considered when digesting DNA, because cleavage may be blocked or impaired when a particular base in the recognition site is methylated. Our results suggest that the novel MTases characterized in this study can be very useful in verifying the adenine methylation sensitivity of many REases and in revealing such behavior in those endonucleases with unknown sensitivity to this DNA modification. Indeed, we demonstrated that 16 REases with unknown sensitivity to m^{6}A in their cognate sequences were in fact sensitive to this modification. The knowledge whether a particular REase is sensitive to a nucleotide modification in its recognition sequence may be utilized to analyze patterns of regional methylation or to establish the methylation status of genomic DNA.

In eukaryotic cells, nuclear DNA is subject to enzymatic methylation resulting in the formation of m^{5}C residues, mainly in CG and CNG sequences. In plants and animals, this DNA methylation is species-, tissue- and even organelle-specific. It changes with age and is regulated by hormones. On the other hand, genome methylation can control hormonal signaling (36).

It has long been known that m^{6}A is present in the DNA of algae and their viruses, fungi and protists. Furthermore, and in spite of the common opinion that m^{6}A is not found in the DNA of higher eukaryotes, this modified base has been detected in plastid, mitochondrial and nuclear plant DNA and in mosquito DNA (37). Accumulating evidence suggests that m^{6}A affects the regulation of gene expression in mammalian cell (38–40). The use of the novel MTases identified in this work might allow us to study the effects of methylation on eukaryotic DNA–protein interactions and, in consequence, its impact on transcription in mammalian or plant cells.

Sequence searches with the Hia5 sequence against a non-redundant database (NCBI) revealed many closely related proteins, suggesting that there may exist more DNA : m^{6}A MTases with similar properties, i.e. extremely relaxed substrate specificity. All identified Hia5 homologs were present in bacterial parasites, pathogens or commensal strains in mammals, including humans (data not shown). Thus, the contribution of Mu-like sequences to the virulence of their bacterial hosts should be considered.

Despite hosting Mu-like prophages encoding these novel MTases, we did not observe additional methylation of DNA during the normal growth of *H. influenzae* Rd (lysogenic for FluMu) and *H. influenzae* biotype aegyptius ATCC 11116, or after induction with mitomycin C, beyond the native methylation resulting from the activities of M.HinDam and M.HindIII, and M.HaeII, respectively. We were also unable to isolate FluMu phage particles (data not shown). Similarly, McGillivary *et al.* (18) failed to purify phage particles from the *H. influenzae* biotype aegyptius F3031. The expression of the *mom* gene is under strict control mediated by the Com protein (11). Possible Com (or Com homolog) binding sites and associated regulatory elements have been identified at the 5′-end of the *hin1523* gene (17). This suggests that the Mom protein of Mu and these novel MTases could be regulated by a similar mechanism and may explain why we were unable to detect *in vivo* MTase activity under laboratory conditions; the *hin1523* gene presumably remains repressed until the prophage enters the lytic cycle.

Saariaho *et al.* (41) revealed that FluMu encodes a functional transposase and contains critical transposase binding sites, but they were also unable to obtain FluMu virus plaques. Other attempts to induce Mu-like prophage excision and cell lysis using the DNA-damaging agents mitomycin C and UV irradiation have also been unsuccessful (42). Under certain conditions it is likely that the host may benefit from the presence of prophages (in particular those unable to enter a lytic cycle) by conferring immunity to superinfection by related phages (43).

Although the biological role of these newly characterized MTases for Mu-like phages and their hosts remains unknown, the discovery of a group of DNA:m^{6}A MTases with extremely relaxed substrate specificity has significance due to their potential as tools in molecular biology research, particularly in the study of the poorly characterized role of adenine methylation in eukaryotic DNA.

ACCESSION NUMBER

JF268249.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–4 and Supplementary Figures 1–9.

REFERENCES

1. Bertani,G. and Weigle,J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113–121.
2. Tock,M.R. and Dryden,D.T. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.

3. Wilkins,B.M. (2002) Plasmid promiscuity: meeting the challenge of DNA immigration control. *Environ. Microbiol.*, **4**, 495–500.

4. Bickle,T.A. and Kruger,D.H. (1993) Biology of DNA restriction. *Microbiol. Rev.*, **57**, 434–450.

5. Kruger,D.H. and Bickle,T.A. (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.*, **47**, 345–360.

6. Hattman,S. (1979) Unusual modification of bacteriophage Mu DNA. *J. Virol.*, **32**, 468–475.

7. Hattman,S. (1980) Specificity of the bacteriophage Mu mom+ -controlled DNA modification. *J. Virol.*, **34**, 277–279.

8. Swinton,D., Hattman,S., Crain,P.F., Cheng,C.S., Smith,D.L. and McCloskey,J.A. (1983) Purification and characterization of the unusual deoxynucleoside, alpha-N-(9-beta-D-2'-deoxyribofuranosylpurin-6-yl)glycinamide, specified by the phage Mu modification function. *Proc. Natl Acad. Sci. USA*, **80**, 7400–7404.

9. Toussaint,A. (1976) The DNA modification function of temperate phage Mu-1. *Virology*, **70**, 17–27.

10. Kahmann,R. (1983) Methylation regulates the expression of a DNA-modification function encoded by bacteriophage Mu. *Cold Spring Harb. Symp. Quant. Biol.*, **47(Pt 2)**, 639–646.

11. Hattman,S. (1999) Unusual transcriptional and translational regulation of the bacteriophage Mu mom operon. *Pharmacol. Ther.*, **84**, 367–388.

12. Witkowski,R.T., Hattman,S., Newman,L., Clark,K., Tierney,D.L., Penner-Hahn,J. and McLendon,G. (1995) The zinc coordination site of the bacteriophage Mu translational activator protein, Com. *J. Mol. Biol.*, **247**, 753–764.

13. Hattman,S., Ives,J., Wall,L. and Maric,S. (1987) The bacteriophage Mu com gene appears to specify a translation factor required for mom gene expression. *Gene*, **55**, 345–351.

14. Wulczyn,F.G., Bolker,M. and Kahmann,R. (1989) Translation of the bacteriophage Mu mom gene is positively regulated by the phage com gene product. *Cell*, **57**, 1201–1210.

15. Asadulghani,M., Ogura,Y., Ooka,T., Itoh,T., Sawaguchi,A., Iguchi,A., Nakayama,K. and Hayashi,T. (2009) The defective prophage pool of *Escherichia coli* O157: prophage-prophage interactions potentiate horizontal transfer of virulence determinants. *PLoS Pathog.*, **5**, e1000408.

16. Braid,M.D., Silhavy,J.L., Kitts,C.L., Cano,R.J. and Howe,M.M. (2004) Complete genomic sequence of bacteriophage B3, a Mu-like phage of *Pseudomonas aeruginosa*. *J. Bacteriol.*, **186**, 6560–6574.

17. Morgan,G.J., Hatfull,G.F., Casjens,S. and Hendrix,R.W. (2002) Bacteriophage Mu genome sequence: analysis and comparison with Mu-like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J. Mol. Biol.*, **317**, 337–359.

18. McGillivary,G., Tomaras,A.P., Rhodes,E.R. and Actis,L.A. (2005) Cloning and sequencing of a genomic island found in the Brazilian purpuric fever clone of *Haemophilus influenzae* biogroup aegyptius. *Infect. Immun.*, **73**, 1927–1938.

19. Kaminska,K.H. and Bujnicki,J.M. (2008) Bacteriophage Mu Mom protein responsible for DNA modification is a new member of the acyltransferase superfamily. *Cell Cycle*, **7**, 120–121.

20. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

21. Glover,S.W. and Piekarowicz,A. (1972) Host specificity of DNA in *Haemophilus influenzae*: restriction and modification in strain Rd. *Biochem. Biophys. Res. Commun.*, **46**, 1610–1617.

22. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

23. Pei,J. and Grishin,N.V. (2007) PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics*, **23**, 802–808.

24. Kurowski,M.A. and Bujnicki,J.M. (2003) GeneSilico protein structure prediction meta-server. *Nucleic Acids Res.*, **31**, 3305–3307.

25. Piekarowicz,A., Golaszewska,M., Sunday,A.O., Siwinska,M. and Stein,D.C. (1999) The HaeIV restriction modification system of *Haemophilus aegyptius* is encoded by a single polypeptide. *J. Mol. Biol.*, **293**, 1055–1065.

26. Soding,J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.

27. Bheemanaik,S., Reddy,Y.V. and Rao,D.N. (2006) Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochem. J.*, **399**, 177–190.

28. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.

29. Ringquist,S. and Smith,C.L. (1992) The *Escherichia coli* chromosome contains specific, unmethylated dam and dcm sites. *Proc. Natl Acad. Sci. USA*, **89**, 4539–4543.

30. Hattman,S. (1972) Plasmid-controlled variation in the content of methylated bases in bacteriophage lambda deoxyribonucleic acid. *J. Virol.*, **10**, 356–361.

31. Roberts,R.J., Breitmeyer,J.B., Tabachnik,N.F. and Myers,P.A. (1975) A second specific endonuclease from *Haemophilus aegyptius*. *J. Mol. Biol.*, **91**, 121–123.

32. Roy,P.H. and Smith,H.O. (1973) DNA methylases of *Hemophilus influenzae* Rd. II. Partial recognition site base sequences. *J. Mol. Biol.*, **81**, 445–459.

33. Malone,T., Blumenthal,R.M. and Cheng,X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.*, **253**, 618–632.

34. Klimasauskas,S., Timinskas,A., Menkevicius,S., Butkiene,D., Butkus,V. and Janulaitis,A. (1989) Sequence motifs characteristic of DNA [cytosine-N4] methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. *Nucleic Acids Res.*, **17**, 9823–9832.

35. Chan,S.H., Zhu,Z., Dunigan,D.D., Van Etten,J.L. and Xu,S.Y. (2006) Cloning of Nt.CviQII nicking endonuclease and its cognate methyltransferase: M.CviQII methylates AG sequences. *Protein Expr. Purif.*, **49**, 138–150.

36. Vanyushin,B.F. (2005) Enzymatic DNA methylation is an epigenetic control for genetic functions of the cell. *Biochemistry*, **70**, 488–499.

37. Ratel,D., Ravanat,J.L., Berger,F. and Wion,D. (2006) N6-methyladenine: the other methylated base of DNA. *Bioessays*, **28**, 309–315.

38. Tronche,F., Rollier,A., Bach,I., Weiss,M.C. and Yaniv,M. (1989) The rat albumin promoter: cooperation with upstream elements is required when binding of APF/HNF1 to the proximal element is partially impaired by mutation or bacterial methylation. *Mol. Cell Biol.*, **9**, 4759–4766.

39. Knebel,D. and Doerfler,W. (1986) N6-methyldeoxyadenosine residues at specific sites decrease the activity of the E1A promoter of adenovirus type 12 DNA. *J. Mol. Biol.*, **189**, 371–375.

40. Truss,M., Bartsch,J., Chalepakis,G. and Beato,M. (1992) Artificial steroid hormone response element generated by dam-methylation. *Nucleic Acids Res.*, **20**, 1483–1486.

41. Saariaho,A.H., Lamberg,A., Elo,S. and Savilahti,H. (2005) Functional comparison of the transposition core machineries of phage Mu and *Haemophilus influenzae* Mu-like prophage Hin-Mu reveals interchangeable components. *Virology*, **331**, 6–19.

42. Evans,T.J., Coulthurst,S.J., Komitopoulou,E. and Salmond,G.P. (2010) Two mobile *Pectobacterium atrosepticum* prophages modulate virulence. *FEMS Microbiol. Lett.*, **304**, 195–202.

43. Snyder,L. (1995) Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Mol. Microbiol.*, **15**, 415–420.