Novel non-specific DNA adenine methyltransferases

Marek Drozdz¹, Andrzej Piekarowicz¹, Janusz M. Bujnicki²,³ and Monika Radlinska¹,*

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

Received February 10, 2011; Revised October 20, 2011; Accepted October 23, 2011

ABSTRACT

The mom gene of bacteriophage Mu encodes an enzyme that converts adenine to N⁶-(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 (Pnm1) 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 N⁶-methyltransferases (hin1523, nma1821, hia5, respectively). Products of the hin1523, hia5 and nma1821 genes modify adenine residues to N⁶-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 2-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

© The Author(s) 2011. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
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 McGillivray 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 m6A MTase M.SstI (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-N6 (m6A) MTases. All known DNA m6A MTases belong to the Rossmann-fold MTase superfamily that is structurally and evolutionarily unrelated to the Mom protein that is a member of the acetyltransferase 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, DNaseII, calf intestine alkaline phosphatase (CIP), DNA size standards, unlabeled N6-methyladenine (m6A), N4-methylcytosine (m4C), C2'-methylcytosine (mC) 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, PI nuclease and deoxynucleosides were from Sigma. TLC PEI Cellulose F glass plates were from Merck. Chromatography resin was from Sigma (Ni-nitrolotriacetic 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-3H]-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- mcrA Δ(mrr-hsdMRS-mcrBC) φ80lacZ ΔM15 ΔlacX74 deoR recA1 araD139 Δ(ara-leu)7697 galU galK rpsL endA1 nupG was used for cloning experiments and strain ER2566 (New England Biolabs, Ipswich, MA, USA) F- λ- fluA2 (lon) ompT lacZ: T7 geneI gal sulA11 Δ(mcr-c-mrr)114:: IS10 R(mcr-73::mini-Tn10)2 R(zg-210::Tn10)1 (TetR) 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 µg ml\(^{-1}\); kanamycin (Kn)—50 µg ml\(^{-1}\); tetracycline (Tc)—10 µg 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 µg ml\(^{-1}\) (sBHI) or on sBHI medium supplemented with mitomycin C at 35 ng ml\(^{-1}\).

The hia\(_5\) gene from H. influenzae biotype aegyptius ATCC 11116 (a homolog of ORF44 from H. influenzae biogroup aegyptius BPF strain F3031, AA37176), 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\(^\ast\), Stratagene, USA) in cloning experiments, and pET28a and pET30a (Kan\(^\ast\), Invitrogen, USA) in recombinant protein expression experiments. For the determination of DNA MTase specificity, phase λ DNA (Fermentas, Lithuania) and plasmid pAltOKI (Tc\(^\ast\)) were used. pAltOKI is a derivative of the vector pALTEREx2 (Promega, Madison, USA) obtained by insertion of a synthetic deoxyoligonucleotide duplex (oligonucleotides PstI T1GCA and PvuI AT, Supplementary Table S1) between the PstI and PvuI sites.

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

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− dam− 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 [3H]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 KH2PO4, once in pure water, and once in 70% ethanol. Incorporation of the 3H-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 [3H]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-3H]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. DNA methylated as described above for 3 h. Following the DNA without added endonuclease (pHia5ET, pNmeET and pHinET) was incubated with the recombinant enzyme (1 U) overnight at 37°C. The obtained chromatograms were analyzed with the Empower™ 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 aegypti 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 dm3CTP and dm5ATP 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 aegypti 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 aegypti strain ATCC 11116

The 1332 bp sequenced fragment of H. influenzae biotype aegypti strain ATCC1116 (for cloning see ‘Materials and Methods’ section) showed 86% identity to the DNA sequence between nucleotides 34230 and 35558 of the genomic island identified in the H. influenzae biogroup aegypti 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 323–348 and 413–1255, respectively (Supplementary Figure S2). The comHia translation product was 98% identical to the protein encoded by ORF43 of H. influenzae biotype aegypti 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 (49% identical over its entire length of 39 residues). This similarity suggested that ComHia performs the
same function as Mu Com. The predicted protein product of the \textit{hia5} gene was 281 amino acids long and shared 80% identity with ORF44 of the \textit{H. influenzae} biotype aegyptius strain F3031 genomic island, 69% with Hin1523 and 43% with Nme1821 (Figure 1).

### Protein expression and purification

Recombinant Hia5, Nme1821 and Hin1523 proteins were expressed in \textit{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 \textit{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 \textit{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), \textit{E. coli} Dam (M.EcoKDam), \textit{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^6A 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 [^3H] 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^6dA standard (Table 2), which implied that all of the analyzed MTases modify adenine at the N6 position to form m^6A.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>[^3H] incorporation (dpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Solvent G</td>
</tr>
<tr>
<td></td>
<td>m^6dA</td>
</tr>
<tr>
<td>Hia5</td>
<td>5820</td>
</tr>
<tr>
<td>Nme1821</td>
<td>5670</td>
</tr>
<tr>
<td>Hin1523</td>
<td>5033</td>
</tr>
</tbody>
</table>

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

**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**

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

**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^6A, 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 BglII GGCN2GC) 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^6A 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^6A 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
plasmid DNA was more sensitive to REase cleavage than the pNmeETi or pHia5ETi DNAs.

Quantitative determination of m\textsuperscript{6}A in DNA methylated by Hia5 or Hin1523

The results described above indicated that m\textsuperscript{6}A 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 \(\lambda\) DNA. In contrast, identical analysis of DNA methylated with the Hia5 enzyme (Table 4) by the studied enzymes: Hia5 methylated only \(50\%\) of Dam sites in \(\lambda\) DNA, whereas Hin1523 methylated with the Hin1523 enzyme. The HPLC assay results also allowed calculation of the quantity of adenine residues transformed into m\textsuperscript{6}A (Table 4) by the studied enzymes: Hia5 methylated >60\% of adenine residues in \(\lambda\) DNA and Hin1523 methylated 29.6\%.

As summarized in Table 4, the %GC content measured for \(\lambda\) 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\textsuperscript{6}A in \(\lambda\) DNA dam\textsuperscript{+} dam\textsuperscript{+}, 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 \(\sim50\%\) of Dam sites in \(\lambda\) 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...
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 2.

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

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

ΣA = dA + dmA. Values presented are the means obtained using several independent DNA preparations.
Supplementary Figure S8. Increase of radiolabel incorporation was linear for \( /C_24 \) 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\(^6\)As i.e. m\(^6\)A A m\(^6\)A (Cm2A1) and m\(^6\)A AAA m\(^6\)A (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\(^6\)A 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 (CATATG), 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 CATATG sequence and as a result, NdeI lost its ability to recognize this target. Seven REases with unknown sensitivity to m\(^6\)A 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\(^6\)A 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\(^6\)A content in the H. influenzae biotype aegyptius ATCC 11116 chromosomal DNA, which was calculated to be 1.7%. The m\(^6\)A 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 island 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 syntetic 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 acetylated adenine. 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 m6A, m5C or m4C. 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 m6A. 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 m6A 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 m6A, 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 m6A (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)-tracts 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 AG (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^6^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.

ACKNOWLEDGEMENTS

We thank Krzysztof Skowronek for useful comments and critical reading of the manuscript. Adam Jagielski is gratefully acknowledged for his help with the HPLC DNA methylation assays.

FUNDING

Ministry of Science and Higher Education (2P04B00827 to M.R., NN301032634 to A.P.); 7th EU Framework Programme (HEALTH-PROT, GA No 229676 to J.M.B.). Funding for open access charge: Institute of Microbiology University of Warsaw.

Conflict of interest statement. None declared.

REFERENCES


