Nucleic Acids Research, 1998, Vol. 26, No. 17 **3961–3966**

Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC

Mai Xu, Michael P. Kladde, James L. Van Etten[1] and Robert T. Simpson*

Department of Biochemistry and Molecular Biology and the Center for Gene Regulation, 308 Althouse Laboratory, The Pennsylvania State University, University Park, PA 16802-4500, USA and [1]Department of Plant Pathology, University of Nebraska, Lincoln, NE 68583-0722, USA

Received May 20, 1998; Revised and Accepted July 14, 1998

DDBJ/EMBL/GenBank accession no. AF062394

ABSTRACT

A novel gene encoding a cytosine-5-DNA methyltransferase recognizing the dinucleotide GpC was cloned from *Chlorella* virus NYs-1 and expressed in both *Escherichia coli* and *Saccharomyces cerevisiae*. The gene was sequenced and a predicted polypeptide of 362 amino acids with a molecular weight of 41.903 kDa was identified. The protein contains several amino acid motifs with high similarity to those of other known 5-methylcytosine-forming methyltransferases. In addition, this enzyme, named M.*Cvi*PI, shares 66% identity and 76% similarity with M.*Cvi*JI, the only other cytosine-5-DNA methyltransferase cloned from a *Chlorella* virus. The short, frequently occurring recognition sequence of the new methyltransferase will be very useful for *in vivo* chromatin structure studies in both yeast and higher organisms.

INTRODUCTION

In vivo methylation of DNA has been used successfully in *Saccharomyces cerevisiae* to study protein–DNA interactions in the chromatin of living cells [1–3; for more information about the expression of foreign methyltransferases (MTases) in yeast see 4]. A high frequency of MTase targets is critical for high resolution mapping of chromatin structure. Among currently available MTase probes, the only *de novo* dinucleotide MTase is M.*Sss*I, which recognizes a CpG site (5). Due to under-representation of the CpG dinucleotide in the genome, the resolution of chromatin structure maps using this enzyme is ~35 bp on average in *S.cerevisiae* (6). With this moderate level of resolution, M.*Sss*I can possibly serve to detect the presence of a positioned nucleosome, 146 bp in yeast, without the need for introduction of additional CpG sites into native DNA sequences. However, this resolution is insufficient for mapping the interactions of non-histone, regulatory proteins, as the typical length of the target DNA sequence of yeast regulatory proteins is ~20–30 bp or less. For example, yeast TBP recognizes and binds to an 8 bp sequence (7), while the well-characterized transcriptional activator Gal4p binds to a 17 bp consensus sequence (8). Further, methylation of CpG islands has been implicated as an important controlling element for gene regulation in mammalian systems, which may limit the application of M.*Sss*I in higher organisms (9). To address both the limitation of resolution and the possible inability to utilize M.*Sss*I in higher organisms, cloning and expression of cytosine-5-DNA MTases (5-^{me}C MTase) with different specificities but similarly small recognition sites is essential.

A family of double-stranded DNA viruses that infect certain unicellular, eukaryotic, *Chlorella*-like green algae are reported to be a rich source of restriction/modification systems (10,11). Among the 37 viruses infecting *Chlorella* NC64A and the five viruses infecting *Chlorella* Pbi which have been partially characterized, 39 viral DNAs contain 5-methylcytosine, ranging in concentration from 0.1 to 47% of total cytosine (10,11). One cytosine MTase, M.*Cvi*JI, has been cloned from *Chlorella* virus IL-3A and shown to recognize RGC(T/C/G) (12). As determined by the resistance/sensitivity of the viral DNAs to over 70 methylation-sensitive restriction endonucleases, at least five independent 5-^{me}C modification systems are predicted to be encoded by some of the more highly modified viruses, including MTases thought to recognize CpC and RpCpY (10,11). Based on the composition of the yeast genome, on average, one CpC site per 13.9 bp and one RpCpY site per 10.7 bp can be expected in the genome. Achieving this level of resolution would allow mapping the interactions of most non-histone, regulatory proteins. The cloning of MTases from *Chlorella* viruses could greatly extend the resolution of chromatin mapping as well as allow extension of *in vivo* chromatin mapping to higher organisms.

Amino acid sequence comparison of most cloned cytosine MTases indicates the presence of conserved motifs in similar order (13). The crystal structures of *Hha*I and *Hae*III cytosine MTases provide physical support for this conservation by showing that the conserved motifs form the core structure of the protein and surround the active site cleft (14,15). The two most conserved motifs are motif I, which serves as the *S*-adenosyl methionine (Ado-Met) binding pocket, and motif IV, which contains the catalytic site (14). The presence of these highly conserved motifs provides a potential target for cloning new cytosine MTases. Indeed, a putative cytosine MTase has been isolated from *Arabidopsis* by sequence homology (16).

Here we report the cloning of a new dinucleotide 5-^{me}C MTase gene, called M.*Cvi*PI, from *Chlorella* virus NYs-1. We used the high conservation at motifs I and IV to design primers for PCR amplification of a fragment spanning these two regions. This PCR

*To whom correspondence should be addressed. Tel: +1 814 863 0276; Fax: +1 814 863 7024; Email: rts4@psu.edu

fragment was then used to probe an NYs-1 genomic library to finally clone the M.*Cvi*PI gene. A bacterially expressed protein was purified and demonstrated to methylate GpC *in vitro*. The M.*Cvi*PI gene was also successfully expressed in yeast, indicating that the enzyme can potentially be used to increase the resolution of *in vivo* chromatin mapping.

MATERIALS AND METHODS

Strains and plasmids

Escherichia coli strain DH5αmut10 (17), lacking the *mcr*ABC and *mrr* genes, and the vector pTZ18U were used for all the cloning and sequencing. For expression of M.*Cvi*PI in *E.coli*, the DH5αmut10 lysogen of bacteriophage DE3 was constructed carrying the T7 RNA polymerase gene under control of the *LacUV5* promoter (18). The plasmid pET20b+ is a T7 expression vector which adds a hexahistidine tag at the C-terminus of a cloned protein, while pLysS is a compatible plasmid providing a small amount of lysozyme (Novagen). The entire coding sequence of M.*Cvi*PI, with the SV40 nuclear localization signal (ATG CCA AAG AAG AAG AGA AAG GTT) appended at its N-terminus, was cloned into the yeast expression vector pMPK1 via the *Sfi*I and *Sph*I sites (2). The gene was subsequently integrated into the genomes of both YPH500ΔL (MATα *ade2-101° his3-Δ200 leu2-Δ1 lys2-Δ1 trp1-Δ63 ura3-52*) and YPH499ΔL (identical genotype but MAT**a**) cells to create yeast strains MXY107 and MXY108, respectively (2). These cell lines express the M.*Cvi*PI gene under the control of a *GAL1* promoter.

PCR amplification of a fragment with high homology to conserved sequences of 5-^{me}C MTases

Degenerate primers for PCR were MEC1 (5′-ccggatcCTNTTYGC-NGGNAT-3′), located in motif I, and MEC2 (5′-acctgcagRAAN-CCYTGRCANGGRAANCC-3′), corresponding to motif IV of the conserved amino acid sequence of 5-^{me}C MTase. The sequence was chosen based on either the consensus sequence or, where there was no consensus, the sequence of M.*Cvi*JI (12,13). Within the primer sequence, N represents a mixture of all four bases and lower case letters indicate sequence not existing in the MTase gene but introduced for the convenience of cloning (*Bam*HI and *Pst*I sites). Viral genomic DNA was amplified in 50 µl reactions which contained 40 pmol each primer, 10 pmol each dATP, dCTP, dGTP and dTTP, 1 µg DNA, 2.5 U *Taq* DNA polymerase (Fisher) in a buffer of 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 3 mM $MgCl_2$, 0.05% (v/v) NP-40 and 0.05% (v/v) Tween 20. PCR cycling parameters were as follows: preheating at 94°C for 3 min; five cycles of 94°C for 30 s, 42°C for 30 s and 72°C for 1 min; 20 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 1 min. Reaction products were separated on a 1.5% agarose gel and fragments ~200–300 bp in size were purified from the gel, digested with *Bam*HI and *Pst*I, and subsequently cloned into pTZ18U. About 20 clones were sequenced and the DNA sequences of insertion fragments were translated into peptide sequences to compare with the amino acid sequence of M.*Cvi*JI.

Construction and screening of a *Chlorella* virus NYs-1 genomic library

A genomic library of *Chlorella* virus NYs-1 was constructed by partial digestion of viral DNA with *Sau*3AI, gel electrophoretic separation to enrich for fragments in the 2–4 kb size range and then ligation of the size-selected DNA fragments into *Bam*HI-digested pTZ18U. The resultant plasmids were subsequently transformed into DH5αmut10 and clones with genomic DNA insertions were screened on plates containing X-Gal. The cloned, PCR-amplified fragment with homology to 5-^{me}C MTase was excised from the vector and random primer labeled in the presence of [α-^{32}P]dATP. About 1000–2000 colonies of the library were screened with this probe by *in situ* hybridization to obtain positive clones (19). Plasmid DNAs were purified from positive clones and the inserts sequenced to identify potential 5-^{me}C MTase genes.

Cell culture

Genes with high homology to known 5-^{me}C MTases were cloned in-frame in pET20b+ at the *Nde*I and *Eco*RI sites for expression in *E.coli*. The plasmid was then co-transformed into DH5αmut10(DE3) with pLysS. Expression was induced as follows. Cells were grown to an OD_{600} of ~0.4–0.6 and centrifuged prior to resuspension in fresh medium. IPTG was added to 0.4 mM for 2.5–4 h induction. Plasmids were then purified from induced cells and digested with a set of restriction enzymes (*Hin*dIII, *Hae*II, *Hae*III, *Hha*I and *Ava*II) to detect the presence of MTase activity. For the expression of 5-^{me}C MTases in yeast, a starter culture was grown overnight at 30°C in YPG medium (10 g yeast extract, 20 g peptone, 20 g galactose/l) to an OD_{600} of ~1. Cells were centrifuged and resuspended in fresh YPG medium for growth at 30°C for an additional 16 h. DNA from ~3 ml cells was rapidly isolated by the glass bead method for deamination as described below (20).

Determination of methylation specificity of M.*Cvi*PI in *E.coli* and yeast

The DNA sequence recognized by M.*Cvi*PI was investigated by bisulfite genomic sequencing. Either linearized plasmid (pET20b+ or pET-NYs-1-5) DNA purified from *E.coli* cells or genomic DNA purified from yeast cells was treated by the method initially developed by Frommer *et al.* (21) and further modified by Kladde *et al.* (2). Briefly, DNA is subjected to quantitative deamination with sodium metabisulfite, converting all deoxycytidine residues to deoxyuridine, while 5-^{me}C residues, created by the MTase *in vivo*, resist deamination. Subsequent PCR amplification of selected DNA regions yields a product that is directly sequenced to provide a positive display of 5-^{me}C. The primers used for the *E.coli* pET20b+ plasmid were as follows: pair I (Fig. 1a): CCdeamin1a, 5′-CCATT-CAACCCAACCACTACAC-3′ (the primer for sequencing), and CCdeamin1b, 5′-GGGTTTTGTGGTATTATTGTAGTATTGG-3′; pair II (Fig. 1b): PET3336, 5′-TACCTAACTCCCCATCATATA-AATAACTACA-3′, and PET3849, 5′-TTTTTAGAATGATTT-GGTTGAGTATTTATTAG-3′. The primers for the yeast *STE6* gene were as follows: STE6a1a, 5′-CTAATTATAATTCACAA-ATACACCTCAAAAA-3′ (the primer for sequencing, from 45331 to 45361 m.u. of chromosome XI), and STE6a2a, 5′-AAGTTAGGTTATTTTTGATGGTTTTATTG-3′ (from 45871 to 45843 m.u. of chromosome XI). After PCR amplification, products were analyzed directly by thermal cycle sequencing as described previously (2).

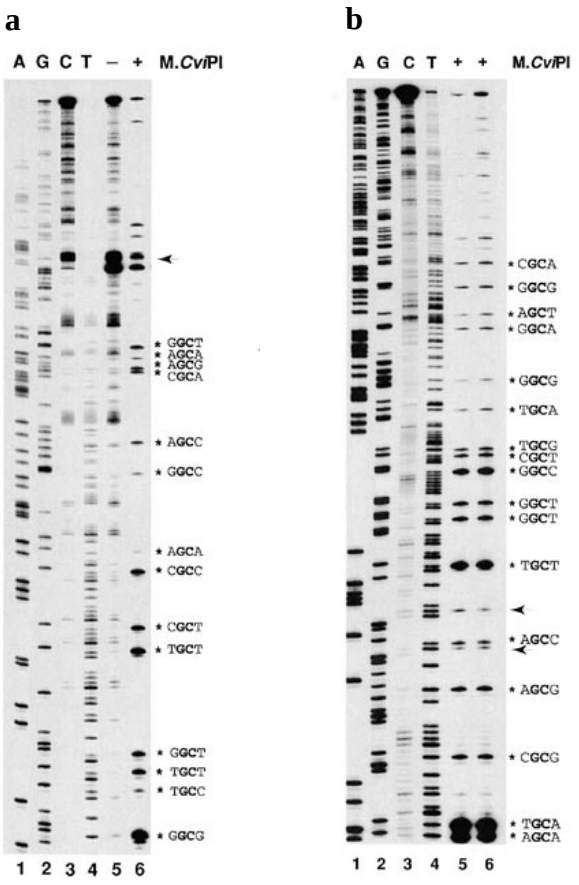

Figure 1. M.*Cvi*PI methylates GpC sites in *E.coli*. Plasmid pETNYs-1-5, containing the coding sequence of M.*Cvi*PI, or pET20b+, the parent vector, was purified from IPTG-induced *E.coli* cells. 5-^{me}C residues on the lower strand were identified in isolated DNA by deamination, subsequent PCR amplification and direct cycle sequencing of the purified PCR products. (**a** and **b**) Two different sequence regions that were investigated, which include GpC sites with all 16 possible combinations of flanking bases. In (a), lanes 1–4 are the sequencing lanes of the region investigated using the PCR product used in lane 5 as template. Lane 5 is a negative control where the same sequence region was examined in the pET20b+ vector only sample. The last lane is the methylation pattern of pETNYs-1-5 isolated from induced cells expressing M.*Cvi*PI. In (b), the first four lanes (lanes 1–4) are sequencing lanes while the last two lanes (lanes 5 and 6) represent the methylation pattern of pETNYs-1-5 isolated from two different induction experiments. In both parts, the arrow represents artefactual primer extension pauses which occurred in samples from both pET20b+ and pETNYs-1-5. The asterisks indicate every GpC site in the sequence that is resolvable on the gel. The flanking bases are listed 5′→3′ beside each methylation band. Due to the fact that methylation of the lower strand was analyzed, the sequence context can be ascertained by reading the gel from the top to the bottom.

Enzyme purification and MTase assay

A 1 l *E.coli* culture was induced as described above for enzyme purification. Cells were sonicated in 10 ml binding buffer (5 mM imidazole, 0.5 M NaCl and 20 mM Tris–HCl, pH 7.9). Cell debris was removed by centrifugation (20 min at 10 000 *g*) and the supernatant was applied to a Ni^{2+}–agarose column (1 × 1 cm). Following sequential washes with 10 ml binding buffer,10 ml wash buffer I (60 mM imidazole, 0.5 M NaCl and 20 mM Tris–HCl, pH 7.9) and 1.5 ml wash buffer II (100 mM imidazole, 0.5 M NaCl and 20 mM Tris–HCl, pH 7.9), the bound enzyme was eluted with 1.5 ml elution buffer (500 mM imidazole, 0.5 M NaCl and 20 mM Tris–HCl, pH 7.9). The eluate was applied to a phosphocellulose column (Whatman P11, 1 × 1 cm) equilibrated with 20 mM Tris (pH 8.0), 50 mM NaCl and 1 mM EDTA. Protein was eluted stepwise with 2 ml portions of the same buffer containing NaCl from 0.1 to 1.0 M in 0.1 M increments. Fractions of 0.5 ml were collected and the fractions (0.3–0.4 M NaCl) containing the predicted 41 kDa protein band were pooled and concentrated using a Centricon concentrator (Amicon). The final enzyme solution was kept at –80°C in a buffer of 20 mM Tris (pH 8.0), 1 mM EDTA, 2 mM DTT and 10% (v/v) glycerol.

Cytosine MTase activity was assayed in a 20 µl reaction containing 20 mM Tris (pH 8.0), 1 mM EDTA, 2 mM DTT, 0.32 nM *S*-adenosyl methionine (New England Biolabs), 1 µg pTZ18U plasmid DNA and 1 µl enzyme fraction. After incubation for 1 h at 37°C, the reaction was stopped by extraction with StrataClean resin (Stratagene). Following ethanol precipitation, the DNA was digested with *Hae*III and analyzed on 1% agarose gels. *Hae*III cleaves GGCC but not GG^{me}CC sequences (22,23). Therefore, lack of *Hae*III cleavage implies a MTase which methylates the internal cytosine in GGCC sites.

RESULTS AND DISCUSSION

Isolation and identification of the M.*Cvi*PI gene from *Chlorella* virus NYs-1

The *Chlorella* virus NYs-1 genome contains a very high level of 5-^{me}C (47.5%) (24). The resistance/sensitivity of its DNA to >70 methylation-sensitive restriction endonucleases indicated that the virus probably encodes several 5-^{me}C MTases (10,11). Sequence alignment of M.*Cvi*JI, the only cytosine MTase cloned from a *Chlorella* virus, with the conserved motifs of other 5-^{me}C MTases indicated that the two most conserved motifs of M.*Cvi*JI are motif I and motif IV. These motifs, which are usually ~200 bp apart (13), correspond to the Ado-Met binding site and the catalytic site in the crystal structure of the *Hha*I and *Hae*III MTases (14,15). A pair of degenerate primers based on the consensus amino acid sequence within the two motifs was used for PCR with NYs-1 DNA. These primers generated several bands ranging from 150 bp to 2 kb, with a prominent band around 200 bp (data not shown). The ~200 bp fragment was gel purified and cloned into pTZ18U following digestion with *Pst*I and *Bam*HI. Sequence analysis of ~20 different transformants identified six different fragments, each of which encoded part of an ORF with high homology to M.*Cvi*JI as well as other 5-^{me}C MTases. Southern blotting with NYs-1 genomic DNA using each of the six fragments as the probe confirmed that these sequences do occur in the viral genome.

To clone the full-length gene of the putative 5-^{me}C MTases, a genomic library of NYs-1 was constructed in pTZ18U, a non-expressing vector, to avoid possible toxicity resulting from a high level of methylation. One of the six cloned PCR fragments was used as a hybridization probe to isolate six positive clones which contained a viral genomic fragment with homology to the probe. Sequence analysis of each of the positive clones revealed that two of them contained a fragment encoding the same, full-length ORF with high homology to 5-^{me}C MTases.

To further characterize this potential MTase gene, the coding sequences were fused to a His_6 tag within the expression vector pTZ20b+, cloned into strain DH5αmut10(DE3) and tested for expression (Materials and Methods). One of the clones, designated pETNYs-1-5, expressed a MTase activity which resulted in

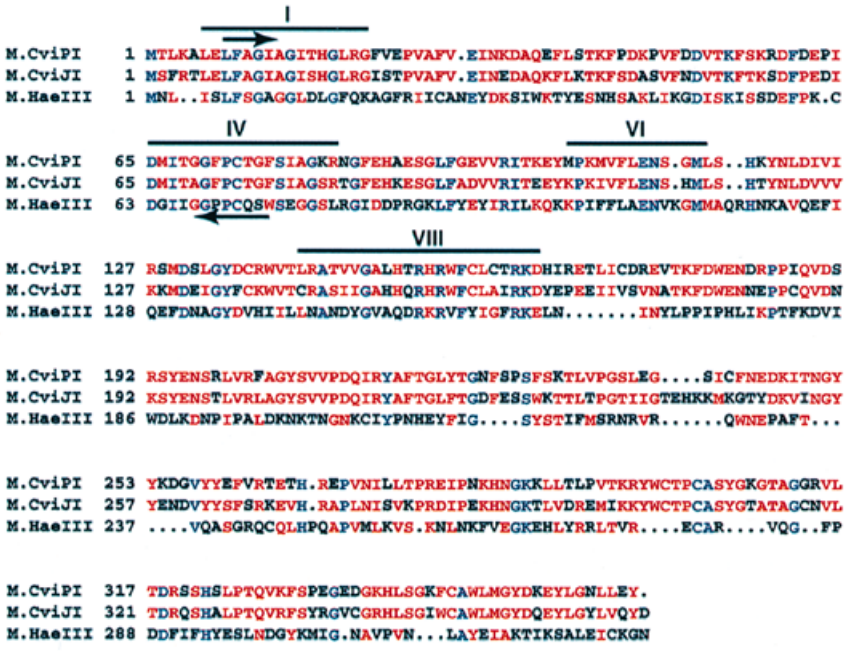

Figure 2. Protein sequence alignment of three cytosine MTases, M.*Cvi*PI, M.*Cvi*JI and M.*Hae*III, by CLUSTLW (31). The bars above the sequence represent the regions of conserved motifs, as indicated. Residues that are identical in all three proteins are indicated in blue and similar residues are shown in red. All the other residues remain as black letters. The regions utilized to create the two degenerate PCR primers are indicated by arrows (only the portion of the primer that is complementary to the MTase gene is indicated).

plasmid DNA that was resistant to digestion by *Hha*I, *Hae*II and *Hae*III (data not shown). NYs-1 viral DNA is also resistant to these same enzymes (data not shown).

The methylation specificity of clone pETNYs-1-5 was examined by bisulfite genomic sequencing. Briefly, pETNYs-1-5 purified from induced cells was deaminated and, following PCR amplification, its methylation pattern was determined by thermal cycle sequencing (Fig. 1). pETNYs-1-5 DNA was methylated at all GpC sites within the resolvable sequences under investigation, suggesting that the plasmid contains a gene encoding a MTase recognizing GpC sites. The pET20b+ vector purified from the same *E.coli* strain grown under the same induction conditions was devoid of methylation. Thus, the MTase activity is encoded by the viral genomic fragment inserted into the vector; the gene was named M.*Cvi*PI. In addition, the sequences investigated in Figure 1a and b include all GpC sites with all 16 possible combinations of flanking bases. Each of these sites was methylated, identifying the enzyme as a cytosine MTase recognizing just the dinucleotide GpC irrespective of flanking sequence context.

Sequence comparison of M.*Cvi*PI and other 5-^{me}C MTases

The amino acid sequence comparison of M.*Cvi*PI with the sequences of other 5-^{me}C MTases supports its identification as a 5-^{me}C MTase. Sequence alignment was performed for M.*Cvi*PI (GenBank accession no. AF062394), M.*Cvi*JI (GenBank accession no. M27265) and M.*Hae*III (GenBank accession no. AF051375), a bacterial MTase recognizing GpGpCpC sequences. The three enzymes have the dinucleotide GpC as the whole or a part of their recognition sequence. As shown in Figure 2, significant conservation exists for the sequences of all three proteins, although conservation between the enzymes from the two *Chlorella* viruses (66% amino acid identity) is higher than for either of these with the bacterial enzyme (~20% amino acid identity). In terms of the six most highly conserved motifs identified in other 5-^{me}C MTases (13), high conservation of motifs I and IV occurs for all three proteins. Motifs VI and VIII can be identified in all three, albeit with a lesser extent of amino acid conservation. No apparent conservation of motifs IX and X exists. According to the crystal structures of both M.*Hha*I and M.*Hae*III (14,15), motif I belongs to the structural segment that forms part of the Ado-Met binding site and motif IV contains the key catalytic cysteinyl residue. These two motifs are directly responsible for the methylation reaction. Motifs VI and VIII are also located around the active site and several interactions occur between them and the Pro-Cys catalytic region (14,15). In general, these four motifs comprise most of the structures that surround the active site cleft. In addition, Gln188 of M.*Cvi*JI, which when mutated leads to an inactive pseudogene in *Chlorella* virus, is conserved in M.*Cvi*PI (25). Motifs IX and X, on the other hand, are more likely to be involved in forming a structural framework for the functional domains. In motif X, the only real conservation includes several hydrophobic side chains involved in packing against α-helix A, an important component of the core structure of the protein. Motif IX has extensive interactions with the variable region and, therefore, may well be involved in sequence-specific recognition of DNA, a feature which should be variable among different 5-^{me}C MTases. In fact, M.*Hae*III interacts with its cognate DNA in a different way than does M.*Hha*I (15), consistent with the lack of significant sequence conservation between the two enzymes in this region. Both M.*Cvi*PI and M.*Cvi*JI were isolated from *Chlorella* viruses and are predicted to be very distant in evolutionary time from bacterial 5-^{me}C MTases. Given the evolutionary differences as well as different target sequences, it is not surprising to find high sequence conservation in the regions

involved directly in the mechanisms of cofactor binding and catalysis, with less sequence conservation in the other regions.

The high amino acid conservation between M.*Cvi*PI and M.*Cvi*JI, cytosine MTases isolated from two different viruses, NYs-1 and IL-3A, and their common GpC methylation sites is quite interesting in the context of mechanisms leading to the profusion of restriction/modification systems in *Chlorella* viruses. It will be interesting if additional enzymes can be isolated from some of the *Chlorella* viruses which recognize versions of GpC or RGC(T/C/G) sites. The high level of sequence conservation reinforces the possibility of cloning such additional 5-^{me}C MTases from *Chlorella* viruses by the sequence homology strategy.

Purification of M.*Cvi*PI protein

The pET20b+ plasmid, containing the M.*Cvi*PI gene, was introduced into *E.coli* strain DH5αmut10(DE3) for expression. Lack of the *mcr*ABC and *mrr* genes in this bacterial strain should decrease any possible toxicity of high levels of cytosine methylation (26). To further reduce the chance of possible deleterious effects of M.*Cvi*PI expression, another compatible plasmid, pLysS, was transformed into the same host. pLysS contains a gene encoding lysozyme, which is an inhibitor of T7 RNA polymerase (18). The low level of lysozyme produced from the plasmid inhibits any T7 RNA polymerase activity resulting from leaky repression of the *LacUV5* promoter and thereby allows more stringent control of MTase production.

Induction of cells at an OD_{600} of ~0.4–0.6 for 2.5–4 h with 0.4 mM IPTG led to an increase in a protein of 41 kDa, the predicted size of M.*Cvi*PI protein (Fig. 3a). The majority of the induced protein was insoluble. Production of large amounts of active protein will require additional investigations into solubilization and renaturation of the enzyme that are beyond the scope of this report. Following purification of the soluble enzyme by Ni^{2+}–agarose and phosphocellulose column chromatography, a dominant single band was observed on SDS gel electrophoresis after Coomassie blue staining (Fig. 3a). A semi-quantitative but highly specific MTase activity assay was performed after each step of purification. The assay measures selectively only those 5-^{me}C MTases which modify the internal cytosine in a GpGpCpC context and thereby make DNA resistant to digestion by the *Hae*III restriction endonuclease (22,23). Throughout the purification, enzyme activity paralleled the presence of the 41 kDa protein band, consistent with its identity as the 5-^{me}C MTase M.*Cvi*PI. The purified enzyme after the final phosphocellulose column chromatography step exhibited an Ado-Met-dependent MTase activity (Fig. 3b). While we cannot compare a specific activity of the purified M.*Cvi*PI with other methyltransferases, the isolated M.*Cvi*PI enzyme shows specificity that is unique. It also lacks contaminating nucleic acid degrading or modifying activities that would impair its use in control of restriction endonuclease activity in cloning or in chromatin structure mapping.

Expression of M.*Cvi*PI in yeast shows its potential for chromatin mapping

The M.*Cvi*PI gene was cloned and expressed in *S.cerevisae*, a eukaryotic organism lacking endogenous methylation of DNA, enabling unambiguous detection of *de novo* modification. Expression of the protein in yeast was under control of the *GAL1* promoter, a strong yeast promoter tightly regulated by carbon source, that is repressed in glucose and induced in galactose (27). Transfection of the cloned gene together with its controllable promoter into the genomic *LYS2* locus created a single copy, stable integrant similar to that used in our previous studies of DNA methylation and chromatin structure in *S.cerevisiae* (2). A region located near the 3′-end of the coding sequence of *STE6* was chosen for bisulfite genomic sequencing, as it is known to be devoid of positioned nucleosomes which may obscure determination of the methyltransferase specificity. After induction by growing yeast cells in medium containing galactose, genomic DNA was purified and deaminated to determine the methylation pattern, as previously described (2). As shown in Figure 4, M.*Cvi*PI recognizes and methylates GpC in yeast, just as it does in *E.coli*. Within the sequence resolvable on the gel, M.*Cvi*PI methylated every GpC site. Like other MTases (2), the extent of modification of individual, specific sites is context dependent. In spite of variability in the modification level, studies of several regions of yeast genomic DNA demonstrated that the M.*Cvi*PI enzyme can methylate cytosine in any GpC context, independent of the flanking nucleotide sequences. Successful expression of M.*Cvi*PI in yeast confirms its potential for *in vivo* chromatin mapping studies. In addition, the fact that the GpC dinucleotide is slightly over-represented in the yeast genome increases the resolution of chromatin mapping using M.*Cvi*PI. Within chromosomes I, III, IV and XI, comprising 17% of the total yeast genome, there is one GpC site every 27.9 bp, as compared with one CpG site every 35.7 bp. Thus, in combination, M.*Cvi*PI and M.*Sss*I lead to a resolution of one naturally occuring site every 15.6 bp in chromatin mapping in *S.cerevisiae* (6,28–30).

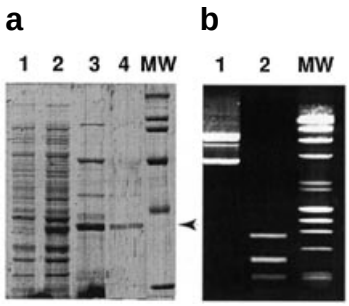

Figure 3. Purification of M.*Cvi*PI and assay for activity *in vitro*. (**a**) SDS–polyacrylamide gel showing the proteins after each step of purification. From 1 l of non-induced (lane 1) or IPTG-induced (lane 2) cells, total protein from 40 and 20 µl cells, respectively, was extracted and analyzed on the gel. Lane 3 contains 10 µl of the 200 µl total eluate from the Ni^{2+}–agarose column. The final, purified enzyme eluted with 0.3–0.4 M NaCl from the phosphocellulose column was analyzed in lane 4 (10 µl of 200 µl total eluate). The arrow indicates the position of the enzyme band. The molecular weight marker used was the broad range standard from Bio-Rad; the 200, 116, 97.4, 66, 45 and 31 kDa species are visible. (**b**) *In vitro* MTase assay of purified M.*Cvi*PI after phosphocellulose column chromatography. Purified enzyme [1 µl of the 200 µl phosphocellulose eluate analyzed in lane 4 of (a)] was assayed for MTase activity as described in Materials and Methods. pTZ18U DNA (1 µg) was treated with the enzyme in the presence of Ado-Met and subsequently digested with *Hae*III (lane 1). As a control, the same reaction was performed in the absence of Ado-Met (lane 2). The molecular marker is a λ/*Hin*dIII plus ϕX174RF/*Hae*III digestion mixture. Note the resistance to digestion by the methylation-sensitive restriction endonuclease in the sample treated with M.*Cvi*PI and Ado-Met.

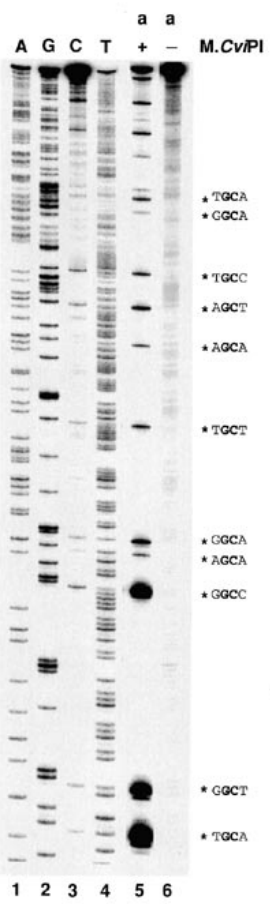

Figure 4. Methylation activity of M.*Cvi*PI expressed in yeast. Genomic DNA was purified from yeast cells grown in medium containing galactose and subsequently deaminated to determine the methylation pattern. Lanes 1–4 contain the sequencing lanes for 45871–45331 m.u. of chromosome XI, a region near the 3′-end of the coding sequence of *STE6*. The methylation pattern of DNAs purified from MXY108 yeast cells is shown in lane 5. Every GpC site present in the region (indicated by the asterisks followed by the GpC site and its flanking bases) is modified by the methyltransferase. As a control, DNA isolated from the parental **a** cell line, lacking the methyltransferase gene, is devoid of methylation (lane 6), demonstrating that the MTase activity is encoded by the M.*Cvi*PI gene.

Summary

We document the occurrence of a novel cytosine-5-DNA methyltransferase, named M.*Cvi*PI, from *Chlorella* virus NYs-1 that recognizes the sequence GpC. The GpC specificity of M.*Cvi*PI is only the second dinucleotide recognition sequence reported for a DNA MTase. Although the genes for ~15 6-^{me}A MTases have been cloned from *Chlorella* viruses, the M.*Cvi*PI gene is only the second 5-^{me}C DNA MTase gene isolated from these viruses. As such, the M.*Cvi*PI gene and its product are of interest in terms of enzymatic mechanisms of small site DNA recognition, cytosine methyltransferases and the evolution of modification systems in algal viruses. From a practical point of view, the identification of methyltransferases such as M*Cvi*PI, having small recognition sites that occur with a high frequency in eukaryotic genomes, and expressing these enzymes in eukaryotic cells is highly important in on-going efforts to develop high resolution methods for analysis of chromatin structure and protein–DNA interactions in living cells.

ACKNOWLEDGEMENTS

We thank members of the Simpson laboratory for helpful discussions, Dwight Burbank for supplying *Chlorella* viral genomic DNA and Mike Nelson for critical reading of the manuscript. This work was supported by NIH grants GM52908 (R.T.S.) and GM32441 (J.V.E.).

REFERENCES

1 Kladde,M.P. and Simpson,R.T. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 1360–1465.
2 Kladde,M.P., Xu,M. and Simpson,R.T. (1996) *EMBO J.*, **15**, 6290–6300.
3 Xu,M., Simpson,R.T. and Kladde,M.P. (1998) *Mol. Cell. Biol.*, **18**, 1201–1212.
4 Kladde,M.K., Xu,M. and Simpson,R.T. (1998) *Methods Enzymol.*, in press.
5 Renbaum,P., Abrahamove,D., Fainsod,A., Wilson,G., Rottem,S. and Razin,A. (1990) *Nucleic Acids Res.*, **18**, 1145–1152.
6 Dujon,B., Alexandrakl,D., André,B., Ansorge,W., Baladron,V., Ballesta,J.P.G., Banrevl,A., Bolle,P.A., Bolotin-Fukuhara,M., Bossler,P. *et al.* (1994) *Nature*, **369**, 371–378.
7 Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) *Nature*, **365**, 512–520.
8 Giniger,E., Varnum,S.M. and Ptashne,M. (1985) *Cell*, **40**, 767–774.
9 Tazi,J. and Bird,A. (1990) *Cell*, **60**, 909–920.
10 Nelson,M., Zhang,Y. and Van Etten,J.L. (1993) *DNA Methylation: Molecular Biology and Biological Significance*. Birkhauser-Verlag Press, Basel, Switzerland, pp. 186–211.
11 Nelson,M., Burbank,D.E. and Van Etten,J.L. (1998) *J. Biol. Chem.*, in press.
12 Shields,S.L., Burbank,D.E., Grabherr,R. and Van Etten,J.L. (1990) *Virology*, **176**, 16–24.
13 Posfai,J., Bhagwat,A.S., Posfai,G. and Roberts,R.J. (1989) *Nucleic Acids Res.*, **17**, 2421–2435.
14 Cheng,X., Kumar,S., Posfai,J., Pflugrath,J.W. and Roberts,R.J. (1993) *Cell*, **74**, 299–307.
15 Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) *Cell*, **82**, 143–153.
16 Finnegan,E.J. and Dennis,E.S. (1993) *Nucleic Acids Res.*, **21**, 2383–2388.
17 Dy,L., Chalasani,S. and Essani,K. (1993) *Gene*, **131**, 87–91.
18 Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
19 Maas,R. (1983) *Plasmid*, **10**, 296–301.
20 Rose,M.D., Winston,F. and Hieter,P. (1990) *Methods in Yeast Genetics: A Laboratory Course Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
21 Frommer,M., Macdonald,L.E., illar,D.S., Collis,C.M., Watt,F., Grigg,G.W., Molloy,P.L. and Paul,C.L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1827–1831
22 Mann,M.B. and Smith,H.O. (1977) *Nucleic Acids Res.*, **4**, 4211–4221.
23 Backman,K. (1980) *Gene*, **11**, 169–171.
24 Schuster,A.M., Burbank,D.E., Meister,B., Skrdla,M.P., Meints,R.H., Hattman,S., Swinton,D. and Van Etten,J.L. (1986) *Virology*, **150**, 170–177.
25 Zhang,Y., Nelson,M. and Van Etten,J.L. (1992) *Nucleic Acids Res*, **20**, 1637–1642.
26 Raleigh,E.A. (1987) *Methods Enzymol.*, **152**, 130–141.
27 Johnston,M., Flick,J.S. and Pexton,T. (1994) *Mol. Cell. Biol.*, **14**, 3834–3841.
28 Bussey,H., Kaback,D.B., Zhong,W., Vo,D.T., Clark,M.W., Fortin,N., Hall,J., Ouellette,B.F., Keng,T., Barton,A.B. *et al.* (1995) *Proc. Natl Acad. Sci. USA*, **92**, 3809–3813.
29 Oliver,S.G., van der Aart,Q.J., Agostoni-Carbone,M.L., Aigle,M., Alberghina,L., Alexandraki,D., Antoine,G., Anwar,R., Ballesta,J.P., Benit,P. *et al.* (1992) *Nature*, **357**, 38–46.
30 Jacq,C., Alt-Morbe,J., Andre,B., Arnold,W., Bahr,A., Ballesta,J.P., Bargues,M., Baron,L., Becker,A., Biteau,N. *et al.* (1997) *Nature*, **387** (suppl. 6632), 75–78.
31 Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.