

# A directed evolution design of a GCG-specific DNA hemimethylase

Rūta Gerasimaitė, Giedrius Vilkaitis and Saulius Klimašauskas\*

Laboratory of Biological DNA Modification, Institute of Biotechnology, Graičiūno 8, LT-02241 Vilnius, Lithuania

Received July 30, 2009; Revised and Accepted September 2, 2009

## ABSTRACT

**DNA cytosine-5 methyltransferases (C5-MTases) are valuable models to study sequence-specific modification of DNA and are becoming increasingly important tools for biotechnology. Here we describe a structure-guided rational protein design combined with random mutagenesis and selection to change the specificity of the HhaI C5-MTase from GCGC to GCG. The specificity change was brought about by a five-residue deletion and introduction of two arginine residues within and nearby one of the target recognizing loops. DNA protection assays, bisulfite sequencing and enzyme kinetics showed that the best selected variant is comparable to wild-type M.HhaI in terms of sequence fidelity and methylation efficiency, and supersedes the parent enzyme in transalkylation of DNA using synthetic cofactor analogs. The designed C5-MTase can be used to produce hemimethylated CpG sites in DNA, which are valuable substrates for studies of mammalian maintenance MTases.**

## INTRODUCTION

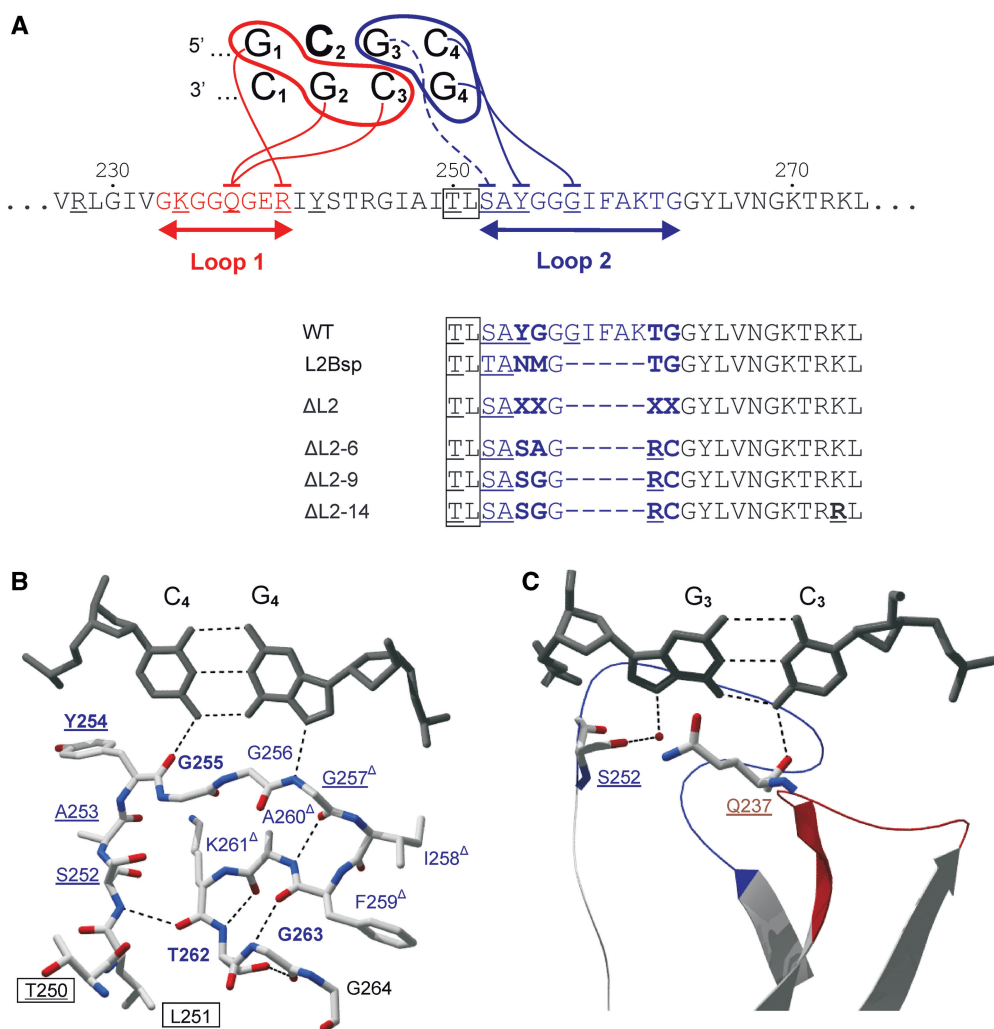
Recognition of specific target sequences in the genome by dedicated proteins plays key roles in controlling the flow of genetic information in the cell. On the other hand, the ability of such proteins to target specific genomic loci makes them valuable tools for modern molecular biology and nanotechnology. In this regard, enzymes capable of both recognizing and modifying a specific DNA sequence are especially interesting. One important class of such enzymes are DNA methyltransferases (MTases), which catalyze the transfer of the methyl group from the cofactor *S*-adenosyl-L-methionine (AdoMet) to their target nucleotides within 2–8 bp sequences in DNA (1). DNA MTases recognizing short sequences have been used as non-destructive chromatin foot-printing agents *in vivo* or *in vitro* (2). Targeted heritable gene silencing can be achieved by MTases fused with sequence-specific Zn-finger domains (3,4). Formation of

irreversible inhibitory complexes between DNA cytosine-5 methyltransferases (C5-MTases) and nucleotide analogs such as 5-fluorocytosine in DNA can be used for the design of covalently functionalized DNA-based nanostructures (5). Recently, novel approaches for sequence-specific functionalization and labeling of DNA using synthetic analogs of the cofactor AdoMet have been proposed (6). More than 900 different MTases are known recognizing over 200 different targets (7). Since the repertoire of naturally occurring enzymes still lacks many useful sequences and sequence types, engineering of enzymes with novel predetermined specificities is increasingly desirable.

C5-MTases, which modify the fifth position of the cytosine ring in the target sequence, proved valuable models to study sequence-specific recognition of DNA modifying enzymes and epigenetic phenomena in higher eukaryotes. C5-MTases share a common mechanism of catalysis which involves flipping of the target cytosine from the DNA helix to the active site followed by its covalent activation. This is manifested by their high degree of sequence and structural homology in the larger catalytic domain. In contrast, the recognition of the specific DNA target is carried out by a highly variable region that folds to form a smaller domain. Thus the catalytic and DNA recognition functions are largely segregated in two distinct domains (8,9). This notion is partially supported by the construction of catalytically active hybrid enzymes by swapping target recognition domains (TRDs) between MTases with different specificities (10,11).

In spite of low sequence conservation in TRDs, certain structural similarities indicate that C5-MTases use similar strategies for their target recognition. Crystal structures of M.HhaI (12) and M.HaeIII (13) show that these enzymes recognize their targets via two distinct recognition loops (Loops 1 and 2) which form multiple base-specific contacts with two distinct segments of their target sequences (Figure 1). Modular organization of TRDs themselves suggested that novel specificities could potentially be created by swapping loop regions between different MTases, however, such experiments typically yielded enzymes with diminished catalytic activity and relaxed target specificity (14,15). During earlier work on swapping segments of TRD between monospecific DNA

\*To whom correspondence should be addressed. Tel: 370 5 2602114; Fax: 370 5 2602116; Email: klimasau@ibt.lt



**Figure 1.** Recognition of the GCGC target sequence by the HhaI methyltransferase. (A) Top—schematic representation of contacts between the target recognition loops (Loops 1 and 2) of M.HhaI and the DNA bases in the GCGC target site; DNA contacting residues are underlined; lines represent direct H-bonds to the DNA bases, dotted lines indicate nucleobase contacts through a water molecule; residues of the conserved TL dipeptide are boxed; target cytosine C<sub>2</sub> is shown in bold; residues in Loops 1 and 2 are colored red and blue, respectively; bottom—aligned Loop 2 sequences of from WT M.HhaI and its engineered variants with altered target specificity; ΔL2 represents the randomized library; randomized positions are bold; an additional mutation outside Loop 2 is bold underlined. (B) and (C)—stick models depicting interactions of M.HhaI with the fourth and third target G:C base pair, respectively, based on a crystal structure of the M.HhaI-DNA-AdoHcy complex (PDB code 3mht). Deleted residues are marked with Δ, other coding as in A.

C5-MTases we found that hybrid MTases, in which the C-terminal recognition loop (Loop 2) of M.HhaI was exchanged, often retained the ability to methylate DNA although with lower efficiency (15). One such hybrid, M.HhaI-L2Bsp (Figure 1), obtained by replacing recognition Loop 2 of M.HhaI with a short fragment from M.Bsp6I (target site GCNGC) showed a marked preference to methylate GCG targets. The asymmetric nature of the GCG sequence allows the generation of hemimethylated CG sites in DNA, which can be uniquely used to study the action of eukaryotic maintenance DNA methyltransferases (16).

Here we report detailed analysis of the sequence specificity of the hybrid M.HhaI-L2Bsp MTase, as well as application of directed evolution to dramatically improve its sequence fidelity and catalytic efficiency. Moreover we show that the newly constructed GCG-specific MTase

is able to transfer extended groups from synthetic AdoMet analogs (17,18) paving the way for its potential use for studies of eukaryotic CpG methylation and targeted DNA labeling.

## MATERIALS AND METHODS

### Materials

All restriction enzymes, DNA polymerases, Exonuclease I, Proteinase K, Shrimp alkaline phosphatase, λ DNA (*dam*<sup>-</sup>, *dem*<sup>-</sup>), pBR322 DNA and kits for molecular biology were obtained from Fermentas and were used according to manufacturer's instructions. Sodium bisulfite, AdoMet and poly[dG-dC]-poly[dG-dC] were purchased from Sigma. Commercial AdoMet was further purified to remove contaminating AdoHcy by passing through a short

column of C<sub>18</sub>-Reversed phase Silica gel 100 (Fluka). [*methyl*-<sup>3</sup>H]AdoMet was obtained from Amersham Biosciences. Oligonucleotides were from MWG Biotech or Fermentas (HPSF grade). The AdoPentyn cofactor was synthesized starting from pentyn-2-ol-1 and purified to an over 85% chiral purity of the *S,S*-isomer (Lukinavicius et al., manuscript in preparation) following previously described procedures (17,19).

### Plasmids, strains and protein expression

Expression of mutant MTases was carried in *Escherichia coli* strains ER1727 (F' lac proA<sup>+</sup>B<sup>+</sup> lacI<sup>q</sup> Δ(lacZ)M15/fhuA2 Δ(lacZ)r1 glnV44 trp-31 mcrA1272::Tn10 (Tc<sup>r</sup>) his-1 rpsL104 (Str<sup>r</sup>) xyl-7 mtl-2 metB1 Δ(mcrC-mrr)102::Tn10 (Tc<sup>r</sup>), ER2267 (F' proA<sup>+</sup>B<sup>+</sup> lacI<sup>q</sup> Δ(lacZ)M15 zff::mini-Tn10 (Kan<sup>R</sup>)/Δ(argF-lacZ)U169 glnV44 e14<sup>-</sup>(McrA<sup>-</sup>) rfbD1? recA1 endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10) and ER2566 (F<sup>-</sup> λ<sup>-</sup> fhuA2 [lon] ompT lacZ::T7 gene 1 gal sulA11 Δ(mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10) (TetS) endA1 [dcm]) (20) (New England Biolabs). ER2566 was modified by transferring an episome from ER2267 to endow cells with a *lacI<sup>q</sup>* gene (G. Mitkaite, unpublished observations). WT M.HhaI, HhaI-L2Bsp and HhaI-ΔL2 mutants were all expressed as an enhanced solubility (21) variant Δ324GH<sub>6</sub>, in which the C-terminal FKPY tetrapeptide is substituted by a C-terminal Gly-His<sub>6</sub> tag by polymerase chain reaction (PCR) mutagenesis using pΔ324G as template and the following primers: 5'-TTTT CGCAATGATCTCAATATTC-3' (direct) and 5'-TTAG TGGTGGTGGTGGTGGCCATTTAATGATGA AC-3' (reverse; 21 nts coding for a His<sub>6</sub> tag and a stop codon are underlined) to give plasmid pΔ324GH<sub>6</sub>. pHHΔBE was constructed by digesting pΔ324GH<sub>6</sub> with R.BspTI and R.Eco91I, blunt-ending with Klenow fragment and recirculation (Figure S1) to give a 36 nt in-frame deletion in the hhaIM gene, involving recognition Loop 2 of M.HhaI.

*Escherichia coli* cells were grown in minimal M9 medium supplemented with histidine, methionine, tryptophan (10 μg/ml each), thiamine (6 μg/ml), carbenicillin (100 μg/ml) and kanamycin (30 μg/ml) at 37°C overnight. For protein overexpression and purification, the ER2267 strain bearing an appropriate plasmid was grown at 37°C until OD<sub>600 nm</sub> ~0.6–0.8, then the culture was cooled down to 16°C and IPTG added to a 0.4 mM concentration. Before harvesting, the culture was further cultivated at 16°C overnight. For the determination of L2Bsp specificity *in vivo*, the mutant MTase was overexpressed in ER1727: the cells were grown in LB medium supplemented with ampicillin (100 μg/ml) and tetracycline (10 μg/ml) at 37°C until OD<sub>600 nm</sub> ~0.6–0.8, then IPTG was added to a final concentration 0.4 mM, the cells were cultivated at 37°C for additional 2 h and harvested by centrifugation.

### Construction of the randomized MTase library

For the HhaI-ΔL2 library construction pHHΔBE linearized with R.BspTI was used as template in the PCR reaction with the following primers: 5'-ATTACCT

TAAGTGCTNNSNNSGGANNNSNNSGGTTACCTAG TAAACGGG-3' (direct; randomized codons are shown in bold; the BspTI site is italicized) and 5'-ATCAACAGGA GTCCAAGCTCAGC-3' (reverse). The resulting 267 bp PCR product was cloned into pHHΔBE using the R.BspTI and R.HindIII sites (Figure S1). The ligation mixture was deproteinized with chloroform and precipitated with ethanol. One hundred and fifty microliters of ER2566 *lacI<sup>q</sup>* cells were electroporated with 1 μg of the ligated DNA, inoculated into 200 ml of supplemented M9 minimal medium and grown at 37°C overnight. The resulting total plasmid DNA was isolated, sequenced and used in biochemical selection.

### Selection of the HhaI-ΔL2 library

Total plasmid DNA was digested with R.Hin6I and R.Bsh1236I, deproteinized, ethanol precipitated and ~0.5 μg was electroporated into 150 μl of ER2566 *lacI<sup>q</sup>* cells. The cells were plated onto five M9 agar plates and incubated at 37°C overnight. The resulting colonies were washed off from the plate, total plasmid DNA was isolated and the biochemical selection repeated.

### Protein expression and purification

Cells were disrupted by sonication in buffer A (20 mM Na-PO<sub>4</sub> pH 7.4, 1 M NaCl and 1 mM PMSF). The supernatant was loaded onto 5 ml HiTrap Chelating<sup>TM</sup> column (Amersham Pharmacia Biotech AB) and eluted with a 3–300 mM linear gradient of imidazole in buffer A. After extensive dialysis to remove AdoMet, the protein was concentrated and stored at –20°C in a buffer containing 20 mM Na-PO<sub>4</sub> pH 7.4, 0.5 mM EDTA, 100 mM NaCl, 2 mM 2-mercaptoethanol and 50% glycerol. Concentration of proteins was determined using a Coomassie G-250 assay with BSA as standard. The molecular mass of each mutant was verified by electrospray mass spectrometry.

### Analysis of methylation specificity by restriction endonuclease cleavage

Methylation reactions were typically performed in a 20 μl of reaction buffer (50 mM MOPS pH 7.4, 0.5 mM EDTA, 15 mM NaCl, 0.2 mg/ml BSA, 2 mM 2-mercaptoethanol) containing 1.2 μg λ DNA and 300 μM AdoMet (or synthetic analog AdoPentyn). The amount of MTase was varied from 3.1 μM to 12 nM in 2-fold dilutions which corresponded to MTase: GCG sites ratio from 1:1 to 1:512 (equivalent to MTase:GCGC ratios from 8:1 to 1:64). Methylation reaction was allowed to proceed for 1 h then the MTase was inactivated by heating at 80°C for 15 min. DNA protection assays were performed as previously described (17).

### Analysis of methylation specificity by bisulfite sequencing

For the determination of the methylation specificity *in vivo*, plasmid DNA isolated from overexpressing cells was analyzed. For the analysis of methylation specificity *in vitro*, 5 μg of pBR322 was methylated with 775 nM and 48 nM L2Bsp or ΔL2 mutant (MTase:target sites ratio of



1:4 and 1:64, respectively) in 120  $\mu$ l of reaction mixture for 1 h at 37°C. Approximately 0.5–1  $\mu$ g DNA fragmented with R.Alw44I or R.PagI in 20  $\mu$ l was denatured by adding 3  $\mu$ l of 2N NaOH and incubating for 30 min at 37°C, and 500  $\mu$ l of freshly prepared sodium bisulfite solution was added. Samples were incubated for 5 h at 55°C, and the temperature was raised to 95°C for 3 min every hour to maintain denaturation of DNA and then processed as described previously (22).

PCR was performed using modified-DNA specific primers (Table S2). In the case of specificity *in vivo* analysis of L2Bsp, all six fragments indicated in Table S2 were analyzed. *In vitro* methylation specificity was investigated after PCR amplification of an upper strand fragment corresponding to the 3369–3967 positions of pBR322 using NG-VP1 and NG-VA3 primers. PCR product was subcloned into R.SmaI-digested pUC19. DNA from individual clones was sequenced using M13/pUC dir (–46), 22-mer and/or M13/pUC rev (–46), 24-mer primers (Fermentas) and analyzed with BiQ Analyzer software v0.91 beta (23).

### Steady-state kinetic analysis

Methylation reactions were carried out in the methylation buffer (50 mM Tris–HCl pH 7.4, 0.5 mM EDTA, 10 mM NaCl, 2 mM 2-mercaptoethanol and 0.2 mg/ml of BSA) at 37°C.  $K_M^{\text{AdoMet}}$  measurements of HhaI- $\Delta$ L2 MTases were performed with constant MTase (4 nM) and poly[dG-dC]-poly[dG-dC] DNA (1.5  $\mu$ M double-stranded GCGC sites) and varying [methyl-<sup>3</sup>H]AdoMet (0.5 Ci/mmol) concentration.  $K_M^{\text{DNA}}$  measurements were performed with constant MTase (25 or 50 pM) and [methyl-<sup>3</sup>H]AdoMet (500 nM or 750 nM; 16.1 Ci/mmol) and varying DNA concentration. Reactions were incubated for 10–60 min at 37°C and processed as previously described (21,24,25). Data were analyzed by non-linear regression fitting to a Michaelis–Menten equation using Grafit 5.0.6 (Erithacus Software) or Dynafit software (26).

### Fluorescence spectroscopy

Affinity of mutant MTases towards cofactor AdoMet was determined by measuring tryptophan fluorescence quenching upon cofactor binding as described previously (24). Titration data were fitted to the equation for single site binding using Grafit 5.0.6.

## RESULTS

### Sequence specificity of M.HhaI-L2Bsp *in vivo*

In preliminary studies, we found that the previously constructed HhaI-L2Bsp hybrid MTase (further referred to as L2Bsp) preferentially methylates GCG targets (15). To determine the *in vivo* specificity of the L2Bsp variant at single nucleotide resolution, we performed bisulfite sequencing of plasmid DNA isolated from *E. coli* ER1727 cells overexpressing L2Bsp. Plasmid DNA was treated with bisulfite and PCR-amplified fragments were cloned into R.SmaI-digested pUC19. Both strands of three fragments from the pMB1 replicon and the

$\beta$ -lactamase gene were analyzed in individual clones. The total length of analyzed regions spanned 1577 nucleotides which included 32 different GCG sites. The methylation status of each GCG site was assayed by sequencing 16–32 independent clones. Control sequencing of a plasmid encoding no active M.HhaI variant revealed no non-converted cytosines besides those in three CCWGG sites methylated by the endogenous EcoDcm MTase. Consistent with earlier results (15), the GCG sites proved the major targets of L2Bsp (Figure 2), although the extent of methylation at individual GCG sites varied from 40 to 100% (Figure S2); no clear correlation between the methylation efficiency and the nature of flanking sequences could be established. However, significant methylation was observed at non-GCG sites: certain GCA sites were methylated as efficiently as 40%, and GCC and GCT were occasionally found to be methylated up to 25%. We concluded that the specificity of L2Bsp is more degenerate than was thought previously and should be defined as GC[G/a]. Therefore the utility of L2Bsp as a molecular tool was limited due to its promiscuous specificity, which manifested as substantial methylation of sequences outside the desired CG consensus.

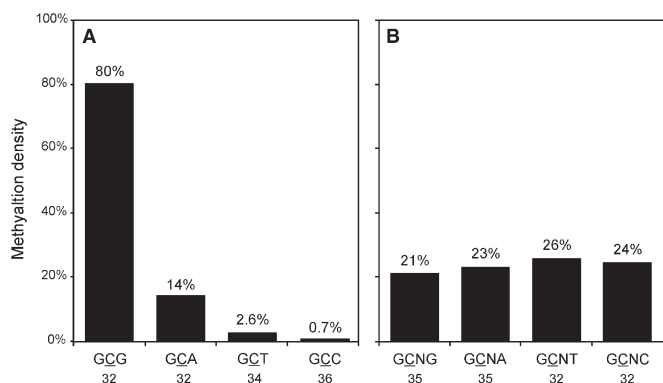
### Strategy for designing a proper GCG-specific methyltransferase

In the L2Bsp variant, the recognition Loop 2 of M.HhaI has become shorter by five amino acids. Based on the crystal structure of the M.HhaI-DNA complex (Figure 1) we concluded that the truncation of the Loop 2 and the associated loss of recognition contacts to the fourth G:C pair is most likely responsible for the change in sequence specificity of the L2Bsp hybrid towards the GCG sites. We also assumed that the new sequence elements that come from M.Bsp6I are too short and out of context to form any structural elements for DNA target recognition. Thus, in order to obtain a more efficient GCG-specific MTase, we decided to retain a five-residue deletion and to optimize the sequence of the truncated Loop 2 by directed molecular evolution. We went on to create a protein library in which four residues in the truncated Loop 2 were randomized; a Gly residue was retained in the center of the random region (XXGXX) (Figure 1) to ensure a certain degree of folding flexibility in the redesigned loop.

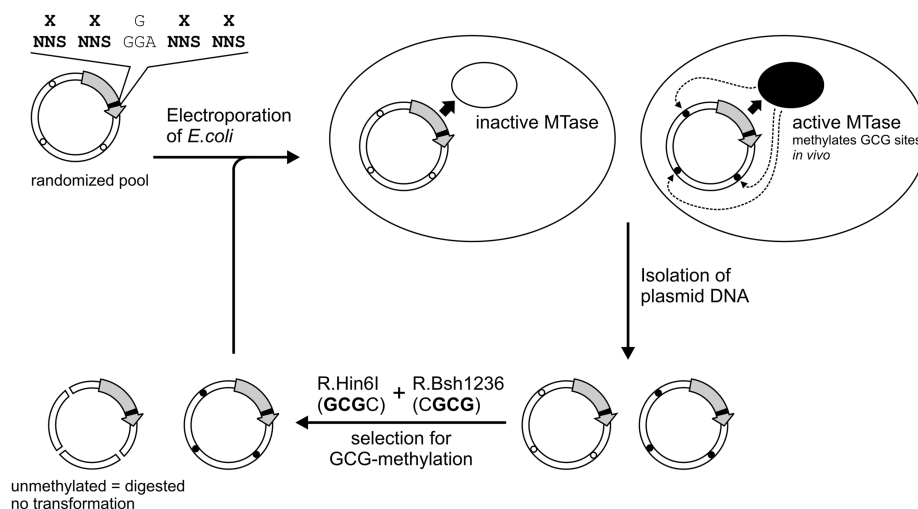
Mutagenesis was carried out by PCR with a degenerate mutagenic primer carrying four NNS codons in positions to be randomized (Figure 3). Linearized pHH $\Delta$ BE, which contained a 36 nt deletion in the Loop 2 and encoded a catalytically inactive protein, was used as the PCR template (Figure S1). The lack of homology to the degenerate part of the primer precluded potential sequence bias from the template. The randomized PCR fragment was inserted in-frame into pHH $\Delta$ BE, and transformation of ER2566 *lacI<sup>r</sup>* *E. coli* cells with the resulting plasmid pool yielded a total of  $\sim 6 \times 10^6$  clones in liquid culture (determined by aliquot plating). The cells were grown at 37°C in M9 minimal medium with background level of transcription (no IPTG added to growth medium). These growth conditions permitted complete methylation of

plasmid DNA (protection against R.Hin6I cleavage *in vitro*) in cells expressing WT M.HhaI, but little methylation was rendered by L2Bsp (not shown). The total plasmid DNA was isolated, and sequencing of the randomized region showed no significant nucleotide bias in the four codons. Restriction endonuclease mapping confirmed the correct overall structures of the plasmid DNA (data not shown).

Selection was carried out by digesting the total plasmid DNA with two 5-methylcytosine sensitive restriction endonucleases, R.Hin6I [GCGC; cytosines whose methylation blocks DNA cleavage are underlined, methylation sensitivity profiles provided in ref. (7)] and



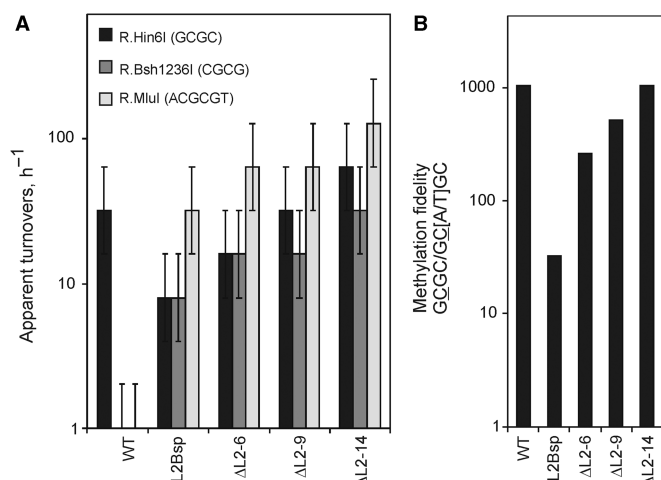
**Figure 2.** Bisulfite sequencing analysis of the *in vivo* sequence specificity of M.HhaI-L2Bsp. Plasmid DNA was isolated from *E. coli* cells overexpressing M.HhaI-L2Bsp and analyzed by bisulfite sequencing. The analyzed region spanned 1577 bases and contained 134 GCN sites. Methylation of each site was assayed by sequencing 16–32 independent clones and the methylation density at each position was determined as a ratio of methylated cytosines over the total number of sequence reads. An average methylation density at GCX (A) and GCNX (B) sequences is shown (the target cytosine residues are underlined) and the number of individual target sites is indicated underneath each sequence.



**Figure 3.** Directed evolution of GCG-specific C5-MTases. A schematic representation of random library selection for active GCG-specific MTases. The MTase encoding gene is shown as a grey arrow, the recognition Loop 2 is shown in black; open and filled circles indicate nonmethylated and methylated GCG sites, respectively.

R.Bsh1236I (CGCG) (14 and 8 target sites on the pHhaI- $\Delta$ L2 plasmid, respectively). The digested plasmid DNA was again transformed into *E. coli* cells and  $\sim 2 \times 10^4$  clones were obtained. Plasmid DNA from nine individual colonies was analyzed with R.Bsh1236I, which showed only four transformants expressing active methyltransferases. The remaining clones were combined, a total plasmid DNA was isolated and the selection procedure was repeated. After the second round of selection,  $\sim 1500$  transformants were obtained. Thirty-five individual clones were tested with R.Bsh1236I and most showed a higher degree of protection than the original L2Bsp expressing plasmid. Sequencing of 28 clones revealed only two sequence variants (SGGRC—22 clones and SAGRC—1 clone) in the randomized region (Figure 1A). Notably, nearly all possible codons for Ser, Gly and Arg were found in the mutant genes indicating that the selected variants did not arise from a single precursor or due to nucleotide bias in the library. A third protein variant (five clones) contained an additional inadvertent mutation, K273R, 10 codons outside of the randomized region.

Representatives of all three variants were selected (referred to as  $\Delta$ L2-6,  $\Delta$ L2-9 and  $\Delta$ L2-14) and corresponding plasmids were analyzed with R.Bsh1236I to compare their methylation efficiency *in vivo* under background level of transcription with that of the original L2Bsp (Figure S3). All three selected variants showed a much stronger protection against the cleavage as compared to L2Bsp; notably, the highest degree of protection came from the  $\Delta$ L2-14 variant, which contained the K273R mutation. All three  $\Delta$ L2-mutant MTases (clones  $\Delta$ L2-6,  $\Delta$ L2-9 and  $\Delta$ L2-14), WT M.HhaI and the L2Bsp control were expressed in *E. coli* and purified to near-homogeneity by Ni<sup>2+</sup> chelating affinity column chromatography for further characterization *in vitro*.



**Figure 4.** DNA protection analysis of the *in vitro* specificity of the HhaI- $\Delta$ L2 MTases. (A) Apparent number of enzymatic turnovers executed at different target sites. Serial 2-fold dilutions starting with equimolar amounts of MTase and target sites were used to methylate bacteriophage  $\lambda$  DNA, which was then challenged with a set of 5-methylcytosine-sensitive restriction endonucleases (Figure S4). Enzymatic turnover rates (turnovers per hour) were estimated based on a minimal molar ratio of MTase to its target sites that is required for a complete protection of DNA in 1 h. (B) The sequence fidelity of the  $\Delta$ L2-methyltransferases, expressed as the ratio of methylation turnover rates at GCGC to GC[A/T]GC sites.

#### Catalytic efficiency and sequence-specificity of selected MTases

For initial characterization of the catalytic activity and sequence-specificity of the M.HhaI mutants, a DNA protection assay was used (17). Serial 2-fold dilutions starting with equimolar amounts of MTase and target sites were used to methylate bacteriophage  $\lambda$  DNA, which was then challenged with a set of individual 5-methylcytosine-sensitive restriction endonucleases (listed in Table S1) and analyzed by agarose gel electrophoresis. Enzymatic turnover rates were estimated based on the minimal molar ratio of MTase to its target sites that is required for complete protection of DNA in 1 h (end-point assay).

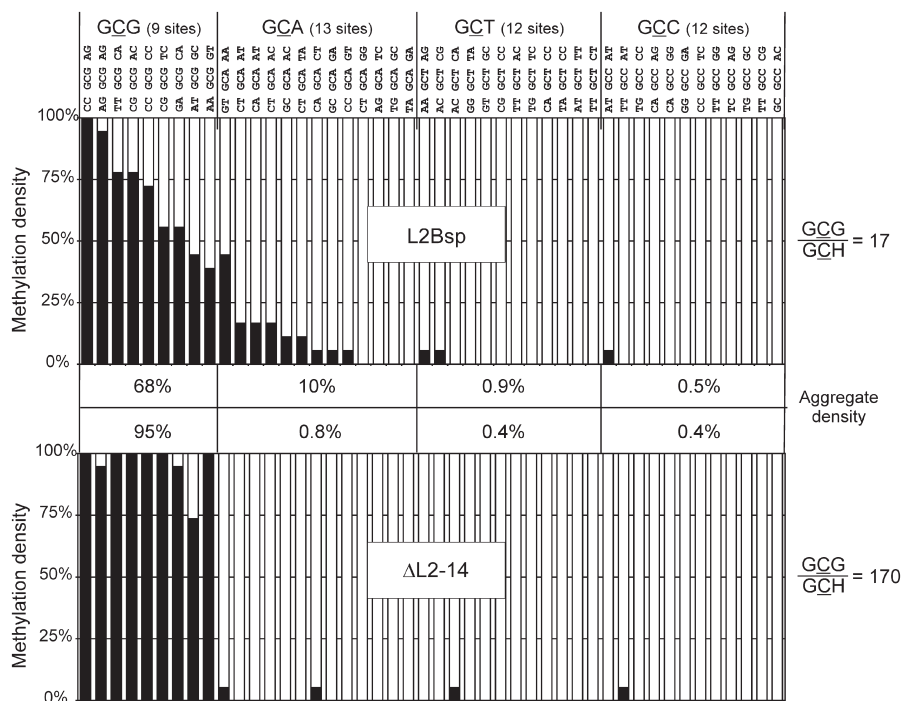
All mutant MTases methylated DNA *in vitro* and complete protection from R.Hin6I could be readily achieved although at different enzyme dilutions (Figure S4A). In support of our *in vivo* results, L2Bsp proved the least active mutant, whereas the  $\Delta$ L2-14 was the most efficient enzyme. The mutant MTases methylated a broader range of targets than the WT HhaI (Figure 4A). A complete protection from GCGC-specific REases that do not have GCGC in their sites (R.Bsh1236I (CGCG) and R.MluI (ACGCGT)) was achieved, consistent with the specificity switch from GCGC (WT) to GCGN. Again,  $\Delta$ L2-14 was found to be the most efficient MTase, whereas  $\Delta$ L2-6 and  $\Delta$ L2-9 showed very similar methylation efficiency. Similarly, off-target methylation at GC[A/T] sites was assessed by digestion with R.BseXI (GC[A/T]GC). We found that methylation of these sites was evident only at high MTase concentrations with the selected mutants (Figure S4). Since no complete protection could be achieved at the non-cognate sites, direct comparison of the methylation efficiency at specific

and non-specific sites was only possible by theoretical calculation of end-point dilutions assuming that, on average, five 2-fold dilutions separate a starting and a full protection point (data not shown). The derived ratios of the target/off-target methylation, which can only provide rough estimates are shown in Figure 4B, indicating that the  $\Delta$ L2-14 mutant is similarly faithful as the WT M.HhaI, whereas the L2Bsp hybrid shows a 10-fold lower sequence fidelity.

Since the use of restriction endonucleases for specificity assessment is undermined by the limited repertoire of available sequence specificities and by significant errors inherent in the determination of end-point fragmentation patterns, bisulfite sequencing of the modified DNA was employed to define more precisely the recognition targets of the  $\Delta$ L2-MTases. pBR322 DNA was methylated *in vitro*, treated with bisulfite and the region of interest was amplified with converted-DNA specific primers. We analyzed a 538 bp fragment in the  $\beta$ -lactamase gene (upper strand positions 3398–3935 on pBR322) which contains 9 GCG sites and 37 GC[A/T/C] sites. For initial analysis, the PCR-amplified fragment was directly sequenced which allowed us qualitatively describe sites as methylated, unmethylated or partially methylated but gave no information about the methylation density of a particular site. When small amount of MTases were used for reaction (MTase:GCG molar ratio 1:64) eight, seven and five out of nine GCG sites were completely methylated by  $\Delta$ L2-14,  $\Delta$ L2-9,  $\Delta$ L2-6, respectively (other GCG sites were nearly fully or partially methylated). No methylation in other positions was observed, and no non-converted cytosines were present in a control sample derived from pBR322 incubated without MTase. In contrast, L2Bsp methylated only three out of nine GCG motifs completely, while weak partial methylation at several GCA positions was clearly evident. Such a wide spread of modification densities observed with the different enzymes appeared suitable for quantitative analysis of their methylation specificity.

The specificity of L2Bsp and  $\Delta$ L2-14 was investigated more thoroughly by cloning the PCR amplified fragments and sequencing individual clones. A total of 18 and 19 clones were sequenced for the two MTases, respectively. The methylation density at a particular site was determined as the ratio of number of methylated cytosines found to a total number of reads. The bisulfite sequencing data are summarized in Figure 5. In addition to GCG, L2Bsp methylated GCA sites quite efficiently and displayed several cytosines in GCC and GCT sites. Consistently with the direct PCR product sequencing data, only two out of nine GCG sites were methylated completely under these conditions. As the amount of MTases in the methylation reaction was increased (MTase:GCG ratio 1:4) some methylated cytosines were found outside the GCG sequences in DNA methylated by all variants. 5-methylcytosines were detected almost exclusively in GCN sites indicating that the third position of the target was the most degenerate. Analysis of DNA methylation by  $\Delta$ L2-14 revealed only a tiny fraction of methylated cytosines outside the GCG targets (4 occurrences in 703 reads through GC[A/T/C] sites) (Figure 5). Comparison of





**Figure 5.** Bisulfite sequencing analysis of *in vitro* specificity of the L2Bsp and  $\Delta$ L2-14 MTases. pBR322 DNA was methylated with L2Bsp (top) or  $\Delta$ L2-14 (bottom) at a MTase to GCG target sites ratio of 1:64 and subjected to bisulfite modification. Methylation densities at individual 46 G<sub>C</sub>N sites were determined by sequencing of a 538 nt pBR322 fragment in individual clones obtained after cloning the bisulfite-converted DNA. The methylation density is expressed as a ratio of methylated cytosines observed to a total number of sequence reads.

relative methylation densities of GCG versus non-GCG sites,  $R_{\text{spec}}$ , again shows that the engineered MTase is  $\sim$ 10-fold more specific than the original L2Bsp hybrid ( $R_{\text{spec}} = 170$  and 17, respectively).

### Kinetic characterization of selected MTases

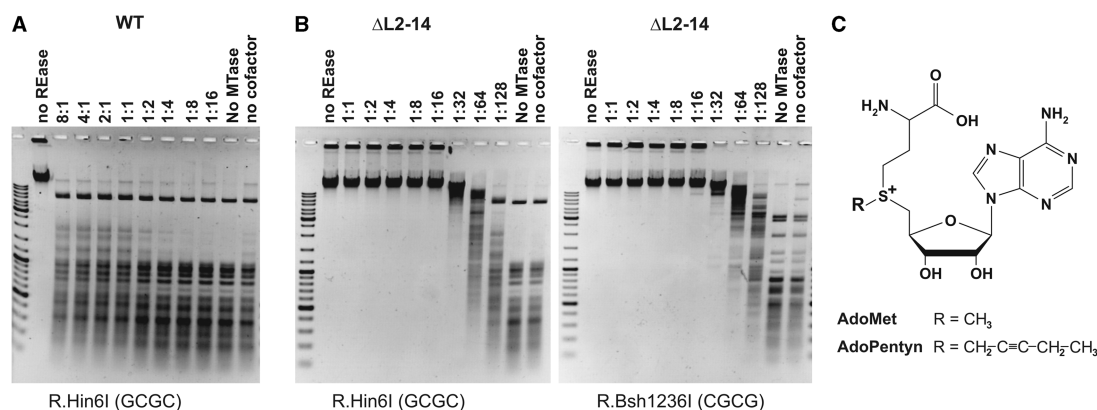
In order to better understand the observed changes in catalytic efficiency and specificity, kinetic parameters were determined for L2Bsp, the three  $\Delta$ L2-mutants and WT M.HhaI. Steady-state kinetic analyses were carried out with poly[dG-dC]-poly[dG-dC] DNA as previously described (27,28). Our data presented in Table 1 show that all variants, including the WT HhaI, have similar  $k_{\text{cat}}$  values (smaller than 4-fold difference), whereas the most prominent differences between the mutants are manifested in  $K_{\text{M}}^{\text{DNA}}$ . With its  $K_{\text{M}}^{\text{DNA}}$  of nearly four orders of magnitude higher than that of WT M.HhaI, L2Bsp clearly indicated its DNA binding capacity to be severely impaired. This notion is in good agreement with our DNA binding studies, which showed no detectable MTase-DNA complexes in gel shift assays even in the presence of AdoHcy (data not shown). Upon optimization of the Loop 2, the interaction between MTases and DNA has improved significantly, as  $K_{\text{M}}^{\text{DNA}}$  decreases in the order L2Bsp  $\gg$   $\Delta$ L2-6  $\geq$   $\Delta$ L2-9  $>$   $\Delta$ L2-14 (Table 1). Altogether, our experiments show that the  $\Delta$ L2-14 variant is the most catalytically efficient GCG-specific MTase.

In addition to increased  $K_{\text{M}}^{\text{DNA}}$ , all mutants showed substantially higher  $K_{\text{M}}^{\text{AdoMet}}$  values as compared to WT M.HhaI (Table 1). This may appear somewhat surprising as changes in the TRD are not expected to disrupt any

**Table 1.** Kinetic and thermodynamic parameters of WT and truncated variants of M.HhaI

M.HhaI variant	$K_{\text{D}}^{\text{AdoMet}}$ ( $\mu\text{M}$ )	$K_{\text{M}}^{\text{AdoMet}}$ ( $\mu\text{M}$ )	$K_{\text{M}}^{\text{DNA}}$ (nM)	$k_{\text{cat}}$ ( $\text{min}^{-1}$ )	$k_{\text{cat}}/K_{\text{M}}^{\text{DNA}}$ ( $\text{M}^{-1} \text{s}^{-1}$ )
WT	$4.2 \pm 0.2$	$0.03 \pm 0.01$	$0.17 \pm 0.02$	$0.89 \pm 0.05$	$9 \times 10^7$
L2Bsp	$6.1 \pm 0.3$	$4.0 \pm 0.5$	$1300 \pm 200$	$0.27 \pm 0.01$	$3 \times 10^3$
$\Delta$ L2-6	$8.4 \pm 0.9$	$1.2 \pm 0.1$	$13.5 \pm 4.2$	$0.45 \pm 0.01$	$6 \times 10^5$
$\Delta$ L2-9	$7.5 \pm 0.8$	$1.3 \pm 0.2$	$4.8 \pm 1.4$	$0.23 \pm 0.01$	$8 \times 10^5$
$\Delta$ L2-14	$6.9 \pm 0.5$	$0.9 \pm 0.1$	$1.7 \pm 0.3$	$0.98 \pm 0.03$	$1 \times 10^7$

protein contacts with the cofactor. To rule out this possibility, we determined  $K_{\text{D}}^{\text{AdoMet}(\text{binary})}$  in the binary complex by monitoring Trp41 fluorescence changes upon cofactor binding (28). Indeed we found that  $K_{\text{D}}^{\text{AdoMet}(\text{binary})}$  is virtually unaffected by the deletion of the Loop 2 (Table 1) and is similar for all mutants and the WT M.HhaI. This suggested that the observed increases in  $K_{\text{M}}^{\text{AdoMet}}$  may be related to a faster release of cofactor from the closed ternary complex. It was previously shown that the rate limiting step in the catalytic cycle of M.HhaI is the dissociation of the ternary product complex (MTase-methylated DNA-AdoHcy), which leads to faster product formation in the first turnover (burst) as compared to the steady state rate (24). We therefore, performed a similar pre-steady-state kinetic experiment (Figure S5), but observed no pre-steady-state burst with the  $\Delta$ L2-14 mutant. Altogether, our findings indicate that the rate limiting step in this mutant is different from WT M.HhaI, consistent with an enhanced cofactor exchange in the ternary complex (25).



**Figure 6.** Enzymatic transalkylation of DNA using synthetic cofactor analogs.  $\lambda$  DNA was incubated with decreasing amounts (two-fold serial dilutions) of WT (A) or  $\Delta$ L2-14 (B) M.HhaI in the presence of cofactor analog AdoPentyn for 1 h at 37°C, then digested with R.Hin6I or R.Bsh1236I and analyzed by agarose gel electrophoresis. Numbers above lanes indicate molar ratios of MTases to their target sites (GCGC or GCG, respectively). (C) Chemical structure of the AdoMet cofactor and its synthetic analog AdoPentyn.

### DNA transalkylation using synthetic AdoMet analogs

Synthetic AdoMet analogs with extended sulfonium-bound propargylic side chains have been used for methyltransferase-directed sequence-specific derivatization and labeling of DNA (17,18). The novel approach, named mTAG, envisions many useful applications. However, as the side chain is extended in length, wild-type DNA MTases, such as M.HhaI, become increasingly inefficient. Simple steric engineering of the cofactor binding pocket (replacing bulky amino acids with Ala or Ser residues) enabled M.HhaI to use such cofactor analogs as alkyl donors (17). Tyr254, which is part of the Loop 2, was one of such replacements around the cofactor pocket (Lukinavičius et al., in preparation). Meanwhile, the  $\Delta$ L2-MTases inherently contain a Ser at this position, suggesting that they may be more active than the WT M.HhaI in the transalkylation reactions. Therefore, the  $\Delta$ L2-14 mutant was tested for its ability to transfer a pentynyl chain from a synthetic cofactor analog, AdoPentyn (Figure 6C). DNA protection assays showed that the apparent alkylation rate of the engineered MTase is  $\sim$ 8 turnovers per hour, which is an at least 100-fold improvement as compared to WT M.HhaI (compare Figure 6A and B), and is only  $\sim$ 8-fold lower than the rate observed with AdoMet (see Figure S4). Digestion of the modified DNA with R.Bsh1236I indicated that the sequence specificity of  $\Delta$ L2-14 remained unaltered with both cofactors (Figure 6B). These experiments demonstrate that the designed MTase can catalyze an efficient transfer of methyl groups as well as extended linear chains to the GCG sites in DNA.

### DISCUSSION

The sequence specificity of C5-MTases is largely defined by their TRD. X-ray structures of reaction complexes for two C5-MTases (M.HhaI GCGC and M.HaeIII GGCC) are available to date (12,13), which revealed that a target DNA sequence is recognized via two recognition loops located in the TRD; the 5' part of the target site (on the target strand) is contacted by the

N-terminal recognition loop (Loop 1), whereas the 3' part of target sequence is recognized by Loop 2 (Figure 1). Attempts to create enzymes with novel specificities by recombining loop regions among multi-specific (14) or mono-specific (15) C5-MTases showed that changes in the recognition Loop1 always lead to inactive enzymes, whereas changes in Loop 2 were often tolerated, but yielded enzymes with diminished catalytic activity and degenerate target specificity. The specificity of the chimeric MTases typically resembled the sequence defined by Loop 1. For example, our previously constructed hybrid M.HhaI-L2Bsp, in which Loop 2 of M.HhaI was replaced by a putatively equivalent region from M.Bsp6I (GCNGC), turned out to be a weak GCG-methylating enzyme (15). Altogether, one can conclude that an exchanged foreign element becomes functionally inactive in the context of another enzyme and that dysfunction of Loop 1, which carries structural elements associated with target base flipping, would be more critical for the catalytic activity than inactivity of Loop 2, which interacts with distal bases in the target sequence. Therefore, for the loop exchange approach to succeed, a proper 'accommodation' of the transferred element is required. This could in principle be achieved by using rational structure-based design and/or directed evolution. Given that scarce structural information is currently available, precise identification of recognition loop boundaries and DNA contacting residues using computational methods may be challenging (28) and thus *in vitro* selection of random libraries presents a powerful alternative.

In this work, our ultimate target sequence specificity was GCG. Starting from M.HhaI, our task appeared as a 'functional deletion' of Loop 2, which is responsible for recognition of the fourth base pair (Figure 1). This recognition is mediated by two hydrogen bond contacts from the main chain atoms of the Loop 2. Directed evolution of DNA-contacting residues in Loop 2 did not yield any active mutants with altered target specificities (29) suggesting that the most promising way to disable Loop 2 is its truncation. As this was in part



achieved in the L2Bsp hybrid, our task was thus reduced to structural ‘accommodation’ of the largely dysfunctional Loop 2. However, at this starting point, the presence of the truncated Loop 2 in the L2Bsp hybrid somehow precluded optimal functioning of Loop 1, which manifested in a substantial promiscuity at the third nucleotide (methylation of GC[A/C/T] sequences along with the GCG target sites). While the exact reason for the enhanced off-target specificity is not clear, two contributing factors can be envisioned. Recognition of the GCG trinucleotide by M.HhaI relies on Loop 1, except for one water mediated contact between the side chain hydroxyl of Ser252 and the N7 atom of the G3 (Figure 1C). It is thus possible that replacement of Ser252 for threonine in the hybrid perturbs the recognition of the C3:G3 pair. Although the water mediated contact to N7 of G3 is not discriminatory *per se* (any purine base could make such a contact), it may aid the recognition of the C3 by Gln237 via proper positioning of the base pair within the protein scaffold. The second possibility is that Loop 2 induces structural perturbations of adjacent regions in the protein and the target DNA leading to altered interactions between Loop 1 and the DNA. Both possibilities were taken into consideration in the selection design.

It came as no surprise that elimination of two hydrogen bonds from the protein–DNA interface upon truncation of the Loop 2 resulted in a decreased affinity of the L2Bsp hybrid towards DNA. This is manifested by its poor catalytic efficiency, a lack of a detectable MTase–DNA complex band in gel-shift assays (not shown), and by increased  $K_M^{DNA}$ . Although such changes seem inevitable when an enzyme with a lower specificity is designed, a possible solution to this problem is to compensate for the lost specific contacts by introducing new nonspecific contacts (or enhancing existing ones) between the DNA and the protein. This is well illustrated by the occurrence of an arginine residue in all selected  $\Delta$ L2-MTases (see Figure 1A). Although we cannot predict the exact conformation of the truncated Loop 2, the proximity of the added Arg to the phosphodiester backbone is likely to account for  $\sim$ 100-fold lower  $K_M^{DNA}$  values in  $\Delta$ L2–6 or  $\Delta$ L2–6 versus L2Bsp. Moreover, the most active variant,  $\Delta$ L2–14, contained another substitution outside the randomized region, which resulted in a several-fold improvement in both  $k_{cat}$  and  $K_M^{DNA}$ . The latter mutation maps to the IX conserved motif, where either Arg or Lys is typically present in other DNA MTases. M.HhaI–DNA crystal structures show that Lys273 points towards the DNA although its positively charged nitrogen atoms is  $\sim$ 5 Å apart from the phosphodiester backbone (not shown). The longer side chain of arginine may bring the positively charged moiety closer to the phosphate leading to an enhanced interaction with the DNA.

In addition to increased  $K_M^{DNA}$ , all the mutants showed 100-fold higher  $K_M^{AdoMet}$  values as compared to WT M.HhaI (Table 1). Higher  $K_M^{AdoMet}$  may result from destabilizing the closed conformation of the catalytic loop (residues 81–99 in M.HhaI) or otherwise enhancing accessibility of the cofactor pocket both leading to a faster cofactor exchange in the ternary reaction/product complex (25,28). Inspection of the crystal structures (12)

indicates that such Loop 2 truncations, and replacement of Tyr254 with a smaller residue (Thr or Ser) in particular, would surely create a wider solvent channel in the cofactor pocket. Moreover, these mutations remove a stabilizing contact between Gln82 in the catalytic loop and Tyr254 in TRD, which is likely to shift the equilibrium of the catalytic loop towards an open conformer. Since essentially no improvement of  $K_M^{AdoMet}$  occurred upon evolution of the truncated Loop 2 (only 4-fold lower values in  $\Delta$ L2-MTases as compared to L2Bsp), one can conclude that either the lost structural feature cannot be recovered in the current structural framework, or further rounds of directed evolution under conditions of low AdoMet concentrations are required. From the point of practical utility, this parameter is quite satisfactory for both *in vivo* and *in vitro* applications. In addition, this endows the  $\Delta$ L2-MTases with a valuable feature to accommodate AdoMet analogs in the active site for targeted transfer of extended groups to DNA (6,18) (Figure 6).

The GCG specificity is unique as no natural C5-MTase is known to recognize this target. Methylation of this asymmetric target leads to the formation of hemimethylated CG sites, which are preferred substrates for eukaryotic maintenance DNA methyltransferases (1). Previously, *in vivo* methylated plasmid DNA obtained from cells overexpressing M.HhaI–L2Bsp was used as a substrate to study the processivity of the mouse Dnmt1 MTase (16). Thus in the context of recently reported specificity changes of M.SinI (30) and M.HaeIII (31), the obtained GCG-MTase has a high potential to become a valuable molecular tool for studies of various aspects of eukaryotic DNA methylation. Although the catalytic efficiency ( $k_{cat}/K_M^{DNA}$ , see Table 1) of the designed MTase is 10-fold lower than that of the WT M.HhaI (28,32), it is comparably efficient or even supersedes certain C5-MTases such as M.HaeIII ( $3 \times 10^4$ ) (31), M.SinI ( $3 \times 10^5$ ) (30) or M.SssI ( $10^4$ – $10^5$ ) (33,34), some of which are widely used to produce methylated DNA molecules for epigenomic and biochemical studies. Its sequence fidelity *in vitro* is also comparable with that of currently characterized WT C5-MTases (30–32). The methylation fidelity defined as the ratio of methylation of target/off-target sites is typically around two orders of magnitude when nano- to micromolar concentrations of enzyme is used [Figures 4 and 5; (32)]. At these practically useful concentrations and typical  $K_M^{DNA}$  values in the nM range ( $[MTase] > K_M^{DNA}$ ), the definition of fidelity as a ratio of  $k_{cat}/K_M^{DNA}$  values for specific over non-specific sites is not particularly informative, since the specificity is largely controlled by  $k_{cat}$  with little contribution from the DNA binding affinity.

In summary, this work represents the first example of enzyme engineering effort leading to a dual specificity change in a DNA methyltransferase. The newly designed MTase is an efficient sequence-specific enzyme that is able to use synthetic AdoMet analogs for the transfer of extended groups onto DNA. The endowed unique features envision useful practical applications of the designer MTase, such as (i) *in vitro* or *in vivo* generation of hemimethylated CpG-sites for studies of maintenance methylation in mammals and (ii) attaching larger

chemical entities to DNA for synthesis of DNA-based nanoparticles.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The authors are grateful to Gražvydas Lukinavičius for his gift of the AdoPentyn cofactor, Goda Mitkaitė for constructing the *E. coli* ER2566 *lacI<sup>r</sup>* (Km<sup>r</sup>) strain, Miglė Gudeliauskaitė for her help with construction of the pΔ324GH<sub>6</sub> plasmid, and Fermentas Life Sciences for valuable support. They also thank Janusz Bujnicki and Jan Kosinski for structural analyses and stimulating discussions during early stages of the project.

## FUNDING

Howard Hughes Medical Institute grants [5500317, 55004123 to S.K.]; and the Lithuanian State Science and Studies Foundation [P-03/2007 to S.K.]. Funding for open access charge: Lithuanian State Science and Studies Foundation [P-03/2009 to S.K.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Grosjean, H. (2009) *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*. Landes Bioscience, Austin.
- Kilgore, J.A., Hoose, S.A., Gustafson, T.L., Porter, W. and Kladdé, M.P. (2007) Single-molecule and population probing of chromatin structure using DNA methyltransferases. *Methods*, **41**, 320–332.
- Carvin, C.D., Parr, R.D. and Kladdé, M.P. (2003) Site-selective in vivo targeting of cytosine-5 DNA methylation by zinc-finger proteins. *Nucleic Acids Res.*, **31**, 6493–6501.
- Smith, A.E. and Ford, K.G. (2007) Specific targeting of cytosine methylation to DNA sequences in vivo. *Nucleic Acids Res.*, **35**, 740–754.
- Singer, E.M. and Smith, S.S. (2006) Nucleoprotein assemblies for cellular biomarker detection. *Nano Lett.*, **6**, 1184–1189.
- Klimasauskas, S. and Weinhold, E. (2007) A new tool for biotechnology: AdoMet-dependent methyltransferases. *Trends Biotechnol.*, **25**, 99–104.
- Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2007) REBASE—enzymes and genes for DNA restriction and modification. *Nucleic Acids Res.*, **35**, D269–D270.
- Lauster, R., Trautner, T.A. and Noyer-Weidner, M. (1989) Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.*, **206**, 305–312.
- Posfai, J., Bhagwat, A.S., Posfai, G. and Roberts, R.J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.*, **17**, 2421–2435.
- Trautner, T.A., Balganes, T.S. and Pawlek, B. (1988) Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. *Nucleic Acids Res.*, **16**, 6649–6658.
- Klimasauskas, S., Nelson, J.L. and Roberts, R.J. (1991) The sequence specificity domain of cytosine-C5 methylases. *Nucleic Acids Res.*, **19**, 6183–6190.
- Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
- Reinisch, K.M., Chen, L., Verdine, G.L. and Lipscomb, W.N. (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell*, **82**, 143–153.
- Lange, C., Wild, C. and Trautner, T.A. (1996) Identification of a subdomain within DNA-(cytosine-C5)-methyltransferases responsible for the recognition of the 5' part of their DNA target. *EMBO J.*, **15**, 1443–1450.
- Vilkaitis, G. and Klimasauskas, S. (1998) Construction and analysis of monospecific DNA cytosine-C5 methyltransferases with chimeric target recognition domains. *Biologija*, 51–54.
- Vilkaitis, G., Suetake, I., Klimasauskas, S. and Tajima, S. (2005) Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J. Biol. Chem.*, **280**, 64–72.
- Dalhoff, C., Lukinavičius, G., Klimasauskas, S. and Weinhold, E. (2006) Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat. Chem. Biol.*, **2**, 31–32.
- Lukinavičius, G., Lapiene, V., Stasevskij, Z., Dalhoff, C., Weinhold, E. and Klimasauskas, S. (2007) Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG). *J. Am. Chem. Soc.*, **129**, 2758–2759.
- Dalhoff, C., Lukinavičius, G., Klimasauskas, S. and Weinhold, E. (2006) Synthesis of S-adenosyl-L-methionine analogs and their use for sequence-specific transalkylation of DNA by methyltransferases. *Nat. Protocols*, **1**, 1879–1886.
- Dila, D., Sutherland, E., Moran, L., Slatko, B. and Raleigh, E.A. (1990) Genetic and sequence organization of the mcrBC locus of *Escherichia coli* K-12. *J. Bacteriol.*, **172**, 4888–4900.
- Daujotyte, D., Vilkaitis, G., Manelyte, L., Skalicky, J., Szyperski, T. and Klimasauskas, S. (2003) Solubility engineering of the HhaI methyltransferase. *Protein Eng.*, **16**, 295–301.
- Vilkaitis, G. and Klimasauskas, S. (1999) Bisulfite sequencing protocol displays both 5-methylcytosine and N4-methylcytosine. *Anal. Biochem.*, **271**, 116–119.
- Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J. and Lengauer, T. (2005) BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, **21**, 4067–4068.
- Vilkaitis, G., Merkiene, E., Serva, S., Weinhold, E. and Klimasauskas, S. (2001) The mechanism of DNA cytosine-5 methylation. Kinetic and mutational dissection of HhaI methyltransferase. *J. Biol. Chem.*, **276**, 20924–20934.
- Merkiene, E. and Klimasauskas, S. (2005) Probing a rate-limiting step by mutational perturbation of AdoMet binding in the HhaI methyltransferase. *Nucleic Acids Res.*, **33**, 307–315.
- Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.*, **237**, 260–273.
- Wu, J.C. and Santi, D.V. (1987) Kinetic and catalytic mechanism of HhaI methyltransferase. *J. Biol. Chem.*, **262**, 4778–4786.
- Vilkaitis, G., Dong, A., Weinhold, E., Cheng, X. and Klimasauskas, S. (2000) Functional roles of the conserved threonine 250 in the target recognition domain of HhaI DNA methyltransferase. *J. Biol. Chem.*, **275**, 38722–38730.
- Lee, Y.F., Tawfik, D.S. and Griffiths, A.D. (2002) Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation. *Nucleic Acids Res.*, **30**, 4937–4944.
- Timar, E., Groma, G., Kiss, A. and Venetianer, P. (2004) Changing the recognition specificity of a DNA-methyltransferase by in vitro evolution. *Nucleic Acids Res.*, **32**, 3898–3903.
- Cohen, H.M., Tawfik, D.S. and Griffiths, A.D. (2004) Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. *Protein Eng. Des. Sel.*, **17**, 3–11.
- Youngblood, B., Buller, F. and Reich, N.O. (2006) Determinants of sequence-specific DNA methylation: target recognition and catalysis are coupled in M.HhaI. *Biochemistry*, **45**, 15563–15572.
- Subach, O.M., Maltseva, D.V., Shastry, A., Kolbanovskiy, A., Klimasauskas, S., Geacintov, N.E. and Gromova, E.S. (2007) The stereochemistry of benzo[a]pyrene-2'-deoxyguanosine adducts affects DNA methylation by SssI and HhaI DNA methyltransferases. *FEBS J.*, **274**, 2121–2134.
- Rathert, P., Raskó, T., Roth, M., Slaska-Kiss, K., Pingoud, A., Kiss, A. and Jeltsch, A. (2007) Reversible inactivation of the CG specific SssI DNA (cytosine-C5)-methyltransferase with a photocleavable protecting group. *ChemBioChem*, **8**, 202–207.