Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 334 (2005) 1233–1240
BBRC
www.elsevier.com/locate/ybbrc

Protein fragment complementation in M.HhaI DNA methyltransferase

Wonchae Choe [a,b,1], Srinivasan Chandrasegaran [b,*], Marc Ostermeier [a,*]

[a] Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 N. Charles St. Baltimore, MD 21218, USA
[b] Department of Environmental Health Sciences, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205, USA

Received 24 June 2005
Available online 18 July 2005

Abstract

The 5mC DNA methyltransferase M.HhaI can be split into two individually inactive N- and C-terminal fragments that together can form an active enzyme in vivo capable of efficiently methylating DNA. This active fragment pair was identified by creating libraries of M.HhaI gene fragment pairs and then selecting for the pairs that code for an active 5mC methyltransferase. The site of bisection for successful protein fragment complementation in M.HhaI was in the variable region near the target recognition domain between motif VIII and TRD. This same region is the location of bifurcation in the naturally split 5mC methyltransferase M.AquI, the location for circular permutation in M.BssHII, and the location for previously engineered split versions of M.BspRI.

Keywords: M.HhaI; DNA methyltransferase; Protein fragment complementation; Leucine zipper; Incremental truncation; Protein engineering; M.AquI; M.BssHII; M.BspRI; DNA methylation

DNA methyltransferases (MTases) are DNA modifying enzymes that catalyze the transfer of a methyl group from *S*-adenosylmethionine to an adenine or cytosine base within a specific DNA sequence. DNA MTases are divided into three main groups based on the site of methylation. MTases that methylate the 5 position of cytosine are referred to as 5mC MTases. The modified base 5-methylcytosine is the only methylated base found in vertebrate and plant DNA. DNA methylation, especially in eukaryotes, has a profound effect on gene regulation. Epigenetic gene silencing by methylation is achieved by a non-mutational gene inactivation that occurs by the transfer of methyl groups to the cytosine residues of promoter sequences. This modification is faithfully propagated from mother to daughter cells. Because aberrations in cytosine-5-methylation are involved in many human genetic diseases, 5mC MTases are important targets for drug development to treat these diseases.

The crystal structure of *HhaI* cytosine-5-methyltransferase (M.HhaI: 5′GCGC3′) from *Haemophilus haemolyticus* was the first to be solved [1]. The structure of M.HhaI with its cognate sequence DNA showed that the enzyme flips the target cytosine residue that is to be methylated out of the DNA helix without any major distortion to the DNA [2]. M.HhaI has a large N-terminal domain and a smaller C-terminal domain connected by a 'long linker' region (approximately residues 175–191). The large domain is discontinuous, as the C-terminal end of the protein is a α-helix that is inserted into the large domain. Sequence comparison of 5mC MTases revealed 10 highly conserved protein motifs [3]. Motif I, also known as the FXGXG motif, is found in all SAM-dependent MTases and it appears to be involved in SAM binding; motif IV also known as the PC motif, which has a conserved cysteine residue, has been shown to be involved in enzyme catalysis; motifs II–V are

* Corresponding authors. Fax: +1 410 516 5510.
E-mail addresses: schandra@jhsph.edu (S. Chandrasegaran), oster@jhu.edu (M. Ostermeier).
[1] Present address: Department of Molecular Biology and Biochemistry, School of Medicine, Kyung Hee University, 1 Hoeki-Dong, Dongdaemoon-Ku, Seoul, 130-701, Korea.

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.07.017

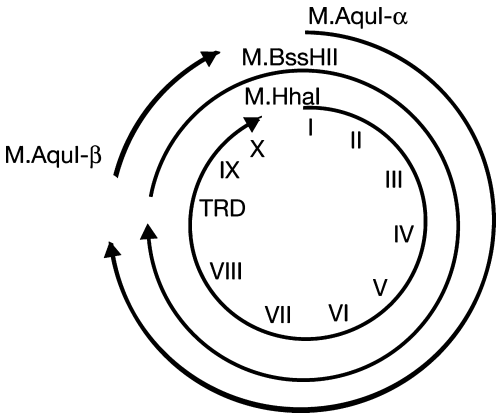

Fig. 1. Sequence comparison of M.HhaI with M.BssHII and M.AquI. The order of conserved motifs is circularly permuted in M.BssHII [6] and is split between two peptides in M.AquI [7].

likely responsible for the key interactions between the co-factor and the protein that are mediated through a set of hydrophobic amino acids [4,5]. The variable region between motifs VIII and IX (amino acids 171–271 in M.HhaI), which include a small sub region called the target recognition domain (TRD), shows the greatest heterogeneity in size and sequence among all 5mC MTases. Interestingly, 5mC MTases can have different assemblies of these motifs (Figs. 1 and 2).

Relative to most 5mC MTases, including M.HhaI, M.BssHII is circularly permuted within the variable region and the TRD, IX, and X motifs precedes motifs I–VIII [6]. The evolutionary origin of this circular permutation of motifs I–X within the 5mC methyltransferase family is not clear [10]. One possibility is that evolution could have proceeded by gene duplication, modifications of this gene pair (e.g., introduction of new stop and start codons) to create interacting fragment pairs, and finally fusion of these gene fragments in a new order. Such a mechanism is supported by the existence of a fragmented 5mC methyltransferase, M.AquI. M.AquI is comprised of two polypeptide chains coded for by partially overlapping reading-frames but is structurally organized quite similarly to other 5mC methyltransferases [7,11]. The 10 highly conserved motifs are present in the correct order on two distinct polypeptides: the α-polypeptide contains motifs I–VIII, and the β-subunit contains the TRD together with motifs IX and X. It has been shown that M.BspRI can be functionally bifurcated at several locations including within the variable region [9] (Fig. 2). The plasticity of the variable region to modifications is also evident in multi-specific mC5 MTases (those which methylate more than one specific DNA target) that have been shown to be tolerant to deletions, rearrangements and exchanges between different enzymes within this region [12]. In addition, some hybrids between M. HpaII and M.MspI in which the fusion was at the beginning or end of the variable region showed partial methylation activity [13].

Previous combinatorial searches for split fragments of *Escherichia coli* glycinamide ribonucleotide phosphotransferase [14] and aminoglycoside phosphotransferase (3′)-IIa [15] resulted in the identification of complementing fragments that were split both between and within sub-domains and at unexpected locations such as within secondary structural elements and within the active site. We were interested whether a combinatorial search for complementing protein fragments of M.HhaI would also result in fragment pairs split at similar locations to those naturally found in M.AquI and the engineered M.BspRI fragment pairs, or whether new bisection locations would be found. Furthermore, we also have a future interest in conferring additional properties to M.HhaI without affecting it primary function of site-specific methylation of DNA. Identification of bisection sites of complementing fragment partners is a first step towards this goal. Here, we report the identification of a bisection site within the M.HhaI molecule where new functionalities could potentially be introduced in the future without affecting its enzymatic activity.

Materials and methods

Bacterial strains and media. Escherichia coli K-12 strains ER1727 [*F′ proA + B + lacIq* Δ(*lacZ*)*M15*/*t-31 his-1 rpsL104* (*StrR*) Δ(*lacZ*)*r1 glnV44 xyl-7 mtl-2* Δ(*mcrBC-hsdRMS-mrr*)*2::Tn10mcrA1272::Tn10*(*-TetR*) *metB1 fhuA2*] and ER2267 [F′ *proA + B + lacIq* Δ(*lacZ*)*M15 zzf::mini-Tn*10 (KanR)/Δ(*argF-lacZ*)*U169 glnV44 e14-*(*McrA-*) *rfbD1? recA1 relA1? endA1 spoT1? thi-1* Δ(*mcrC-mrr*)*114::IS10*] were used as hosts in DNA subcloning experiments and library construction and were obtained from New England Biolabs (Beverly, MA). Library selection was carried using the *E. coli* strain GeneHogs [*F⁻ mcrA* Δ*(mrr-hsdRMS-mrcBC) f80lacZ*Δ*M15* Δ*lacX74 deoR recA1 endA1 araD139* Δ*(ara-leu)7697 galU galK rpsL(Str*^{R}*) nupG*] which was obtained from Invitrogen (Carlsbad, CA). Cells were grown in LB medium (1% w/v tryptone, 1% w/v NaCl, and 0.5% w/v yeast extract) at 37 °C. Chloramphenicol (75 μg/ml) and/or ampicillin (100 μg/ml) were added to media to maintain the plasmids, as needed.

Plasmids. Restriction enzyme digestions and ligations were performed as recommended by the manufacturer (New England Biolabs, Beverly, MA). DNA manipulations including subcloning and agarose gel electrophoresis were performed essentially as described elsewhere [16]. Plasmid DNA was purified by Qiagen plasmid midi prep kit. The following double-stranded oligonucleotide was digested with *Kpn*I and *Dra*III (sites underlined) and inserted into *Kpn*I/*Dra*III digested pDIM-N2 [17]: 5′-GCCAGC<u>GGTACC</u>ACCGAT***CCCGGG*** AAGCTTGGGGCGGAACAACTGAGGCGCGCCGCTTGGGGA AGAAAC**GAATTC**ATGATAA**ACTAGT**<u>CACTACGTG</u>GGAC-3′. This DNA sequence contains *Eco*RI and *Spe*I sites (bold) to allow for cloning of fragments from pDIM-C8. It also contains a *Sma*I site (bold and italics). The TetR operon was amplified from pACYC184 and inserted between the *Kpn*I and *Sma*I sites to create plasmid pDIM-N7. The gene encoding *M.HhaI* (Genbank Accession No. J02677, bases 437–1420) was amplified by PCR using chromosomal DNA purified from *Haemophilus heamolyticus* bacteria as template and primers complementary to the 5′ end (5′-GCGGGCCCATATGATTGAAAT AAAAGATAAACAG-3′) and 3′end (5′-GCGGGCCTCTAGATTA ATATGGTTTGAAATTTAATGA-3′) of the M.HhaI gene. The M.HhaI PCR product was digested with *Nde*I and *Xba*I and ligated into these same sites in pDIM-N7 under the control of lac promoter to

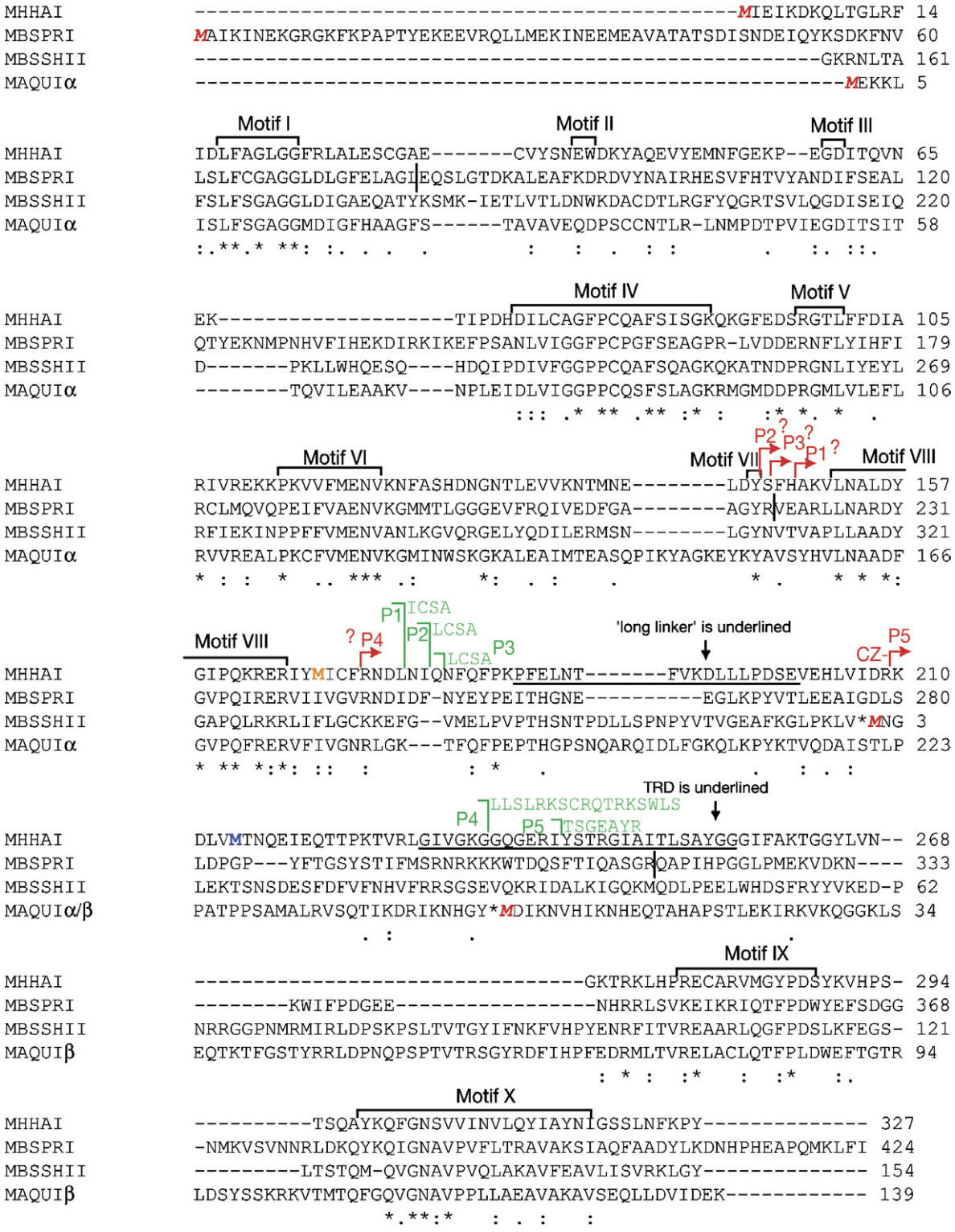

Fig. 2. Sequence alignment of M.HhaI, M.BspRI, M.BssHII, and M.AquI. Alignment was performed using ClustalW [8]. Start methionines are shown in red and italics. The sequences of complementing fragments of M.HhaI found in this study are indicated. The end of the N-terminal fragment is shown in green along with the sequence of the appended peptide. The beginning of the C-terminal fragment is shown in red along with whether the CZ zipper is attached. Variants P1–P4 have a question mark by the start of the C-terminal fragment to indicate that the exact start of translation is not known. What is indicated is the start of the amino acid sequence corresponding to the start of the gene fragment. Potential translation start sites for the C-terminal fragments of P1–P3 is shown in orange and for P4 is shown in blue. The approximate locations of the bifurcation points of M.BspRI resulting in some MTase activity [9] are indicated by the vertical black lines within the M.BspRI sequence. "*", identical amino acid residues in all sequences; ":", conserved substitutions are observed; "." semi-conserved substitutions are observed. (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this paper.)

create pMHhaI. The partial gene fragment coding for M.Hha[1–302] (i.e., fragment encoding amino acids 1–302 of M.HhaI) was amplified by PCR using primers (5′-G CGGGCCCATATGATTGAAATAAA AGATAAACAG-3′) and (5′-GCGGGCCTCTAGAGGATCCTTAA AATTGTTTATATGCTTGGCTGGT-3′). The PCR fragment containing M.Hha[1–302] was inserted in pDIM-N7 using *Nde*I and *Xba*I to create pDIM-N7-MHhaI[1–302] (Fig. 3A). The partial gene fragment coding for M.HhaI[29–327] was amplified by PCR using primers (5′-GGCCCCAGATCTGAATCTTGCGGTGCTGAGTGC-3′) and (5′-GCGGGCCACTAGTTAATATGGTTTGAAATTTAATGA-3′). The PCR fragment containing M.Hha[29–327] was inserted in pDIM-C8 using *Bgl*II and *Spe*I to create pDIM-C8-MHhaI[29–327] (Fig. 3A). Plasmid pAR-MHhaI[29–327] (Fig. 3A) was created by digesting pDIM-C8-MHhaI[29–327] with *Nco*I and *Spe*I and cloning the small fragment into a similarly digested, modified pAR4 [18] in which an *Spe*I site was inserted 5′ to the *Hin*dIII site.

Library construction. The creation of N-terminal and C-terminal incremental truncation libraries was carried out using time-dependent digestion with ExoIII as described elsewhere [17]. After cutting out the C-terminal truncation libraries by *Eco*RI and *Spe*I, they were cloned into *Eco*RI/*Spe*I cut N-terminal libraries to make the N-terminal/C-terminal crossed library (Fig. 3B).

HhaI and AscI protection assay. One microgram of purified plasmid DNA was digested with 20 U of either *Hha*I or *AscI* for 2 h at 37 °C in a total volume of 30 μl in the buffer recommended by the manufacturer. DNA digests were analyzed by using 1% agarose gel electrophoresis.

Library selection procedure. The purified plasmid DNA of the library (1 μg) was digested with 20 U of *HhaI* for 2 h at 37 °C in a total volume of 30 μl in the buffer recommended by the manufacturer. The DNA was treated with 100 U of ExoIII for 10 min at 22 °C to further degrade the *Hha*I digested DNA. The digested DNA was ethanol-precipitated and resuspended in water. GeneHogs cells were electroporated with 10–100 ng of the digested DNA. After incubation at 37 °C for 1 h, the cells were plated on 100 μg/ml ampicillin and 30 μg/ml tetracycline (1 μl on a small plate and the rest on a large 248 times 248 mm plate). Inoculums were prepared from individual colonies from the small plate and the plasmid DNA purified and test for methylation using the *Hha*I protection assay. The colonies from the large plate were recovered en mass. A portion of these cells was saved

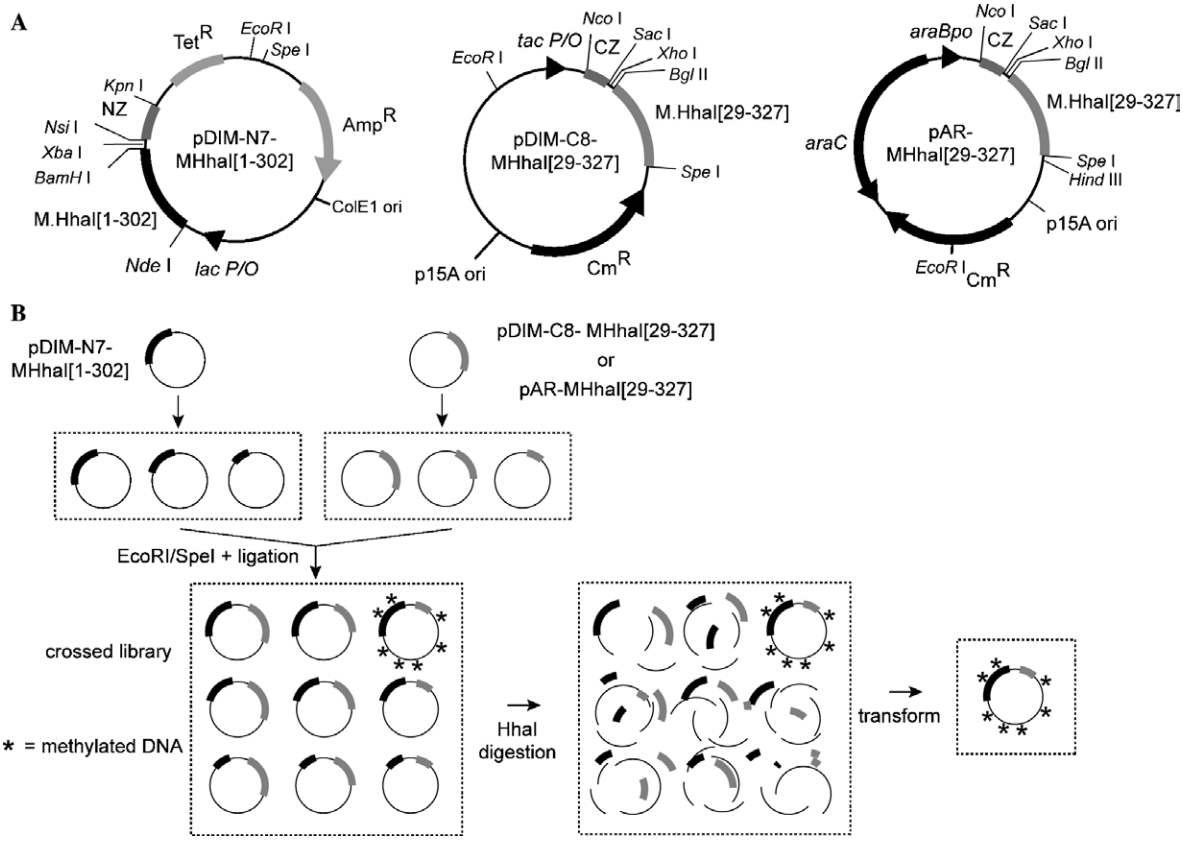

Fig. 3. Incremental truncation scheme and plasmid constructs. (A) Plasmids used for constructing incremental truncation libraries. Plasmids have compatible origins of replication (*ColE1 ori* and *p15A ori*). In pDIM-N7-MHhaI[1–302], the 5′ gene fragment to be truncated is cloned downstream from an IPTG inducible lac promoter (*lac* P/O). The vector also has an ampicillin resistance gene (Amp^R) and a tetracycline resistance gene (Tet^R) and DNA coding for the NZ leucine zipper. In the other two plasmids, 3′ gene fragment to be truncated is cloned downstream either from an IPTG inducible trp/lac hybrid promoter (*tac* P/O) or the arabinose promoter (*araBpo*). Both vectors also have an antibiotic resistance gene (chloramphenicol, Cm^R) and the DNA coding for the CZ leucine zippers. (B) Schematic depiction of library construction and selection. Individual truncation libraries are constructed using time-dependent ExoIII digestion [17] and then crossed by cloning the truncated gene library from pDIM-C8-MHhaI[29–327] or pAR-MHhaI[29–327] into the plasmid library created from pDIM-N7-MHhaI[1–302] using *Eco*RI/*Spe*I digestion, isolation of the desired fragments and ligation. If the gene fragment pairs code for an active MTase, the plasmid DNA will be methylated. The library plasmid DNA is then isolated and digested with *HhaI*, which will not digest methylated DNA. Transformation of this digested DNA into cells enriches for library clones coding for an active MTase.

as a frozen stock and a portion was used to purify the plasmid DNA purified as previously described elsewhere [17]. This plasmid DNA was used as the starting point for subsequent rounds of selection for resistance to *Hha*I digestion.

Results and discussion

Cloning of M.HhaI

The gene coding for M.HhaI was PCR-amplified from purified *H. haemolyticus* chromosomal DNA and cloned downstream from a leaky *lac* promoter on pDIM-N7 to create plasmid pMHhaI. Plasmid pMHhaI was transformed into the *mcrBC*$^-$ *E. coli* strain ER1727, which tolerates cytosine methylation. After confirming the gene had the expected DNA sequence, 5′- and 3′-fragments of the gene coding for amino acids 1–302 and 29–327 were cloned into pDIM-N7 and pDIM-C8 to make pDIM-N7-MHhaI[1–302] and pDIM-C8-MHhaI[29–327], respectively (Fig. 3A). These protein fragments lack motifs X and I, respectively and thus were expected to be inactive.

M.HhaI activity in cells bearing pMHhaI was confirmed in two-ways. First it was determined that the plasmid pMHhaI, when isolated from ER1727, could not be digested by the restriction enzyme *Hha*I (there are 24 *Hha*I sites in pMHhaI), whereas both plasmid pDIM-N2 and pDIM-N7-MHhaI[1–302] could be digested. Second, it was determined that whereas pDIM-N7-MHhaI[1–302] could efficiently transform both ER2267 (*mcrBC*$^-$) and DH5α-E (*mcrBC*$^+$), pMHhaI only could transform ER2267 efficiently. This is consistent with the known intolerance of *mcrBC*$^+$ strains to be transformed with methylated DNA. These experiments also confirm that the fragment M.HhaI[1–302] is incapable of methylating DNA.

Construction of library 1 incremental truncation libraries

A schematic of the library construction is shown in Fig. 3B. We sought protein fragment pairs of M.HhaI that could reassemble into an active 5mC MTase in vivo. We constructed incremental truncation libraries starting from the 3′ end of the gene fragment in pDIM-N7-MHhaI[1–302] and starting from the 5′ end of the gene fragment in pDIM-C8-MHhaI[29–327]. Truncations were performed such that fragments as small as M.HhaI[1–29] and M.HhaI [302–327] would be members of the N-terminal and C-terminal fragment libraries, respectively. In analogous work to identify fragments of the enzyme aminoglycoside phosphotransferase (3′)-IIa, we found that fragments could assemble only if they were fused to antiparallel zippers (no combination of fragments lacking leucine zippers was functional) [15]. Therefore, we designed the library construction such that the truncated genes would be fused to the DNA coding for these same antiparallel leucine zippers NZ and CZ [19]. These zippers could then assist in the assembly of protein fragment pairs that otherwise would not assemble. For the N-terminal fragments, 2/3 of the fusions of the truncated gene with the DNA coding for the leucine zippers would be out of frame and result in a short peptide extension to the N-terminal fragment. Thus, if a leucine zipper interfered in the assembly of two fragments, a corresponding construct lacking the leucine zipper but having a short peptide extension would be present in the library that might allow assembly.

After construction of the individual truncation libraries on pDIM-N7-MHhaI[1–302] and pDIM-C8-MHhaI[29–327], the C-terminal library was subcloned into the plasmids coding for the N-terminal truncation library (Fig 3B). This was done to ensure a physical linkage between the genes coding for potentially complementing protein fragments, which would facilitate the in vitro selection procedure. In order to reduce the potential for recombination within this "crossed-library" plasmid (i.e., recombination between two large overlapping fragments of the *M.HhaI* gene to create a full-length *M.HhaI* gene that would be a false positive), a gene coding for tetracycline resistance was included between the two truncated genes and tetracycline was added in addition to ampicillin to maintain this plasmid library.

Selection for active M.HhaI MTases from Library 1

Our in vitro selection method consisted of transforming the crossed-library plasmid into GeneHogs, purifying the plasmid DNA, digesting the DNA with *Hha*I to destroy plasmids that are not methylated and then retransforming them back into GeneHogs (Fig. 3B). At this point individual colonies were picked and the methylation state of the plasmid was checked by the *Hha*I protection assay. Additional rounds of digestion with *Hha*I and retransformation were performed if none of the plasmid DNA from randomly selected colonies was methylated. As an initial test of our selection strategy, we constructed a mock library consisting of purified plasmids pDIM-N7-MHhaI(1–302) and pMHhaI mixed in a molar ratio of 10,000:1 and performed a single round of selection. This DNA mixture was digested with an amount of *Hha*I that was more than sufficient to fully digest pDIM-N7-MHhaI(1–302). The digested DNA was transformed into ER2267 and 10 of the colonies that grew were examined as to the methylation state of its plasmid DNA (by the *Hha*I protection assay) and by sequencing and restriction digestion analysis. All 10 colonies contained pMHhaI DNA that was methylated.

The truncation library from the N-terminal fragment was constructed and verified to contain a wide variety of

truncation lengths using colony PCR and restriction digests. However, preparation of the C-terminal library was fraught with problems: relatively few transformants were obtained and there was a strong bias towards very short truncations or very long truncations. Despite this bias in the C-terminal library (which was determined to occur after introduction of the plasmid DNA into cells and not during the in vitro enzymatic manipulation of the DNA to make the libraries) these libraries were crossed into the same plasmid. This crossed library (Library 1) was subjected to three rounds of selection. Analysis of the plasmid DNA of more than 400 individual colonies indicated that almost all of the selected library members contained a recombined plasmid with a full-length *M.HhaI* gene. Four libraries members (P1–P4) were found in which there was no recombination (Fig. 2). However, in all four of these members the *tac* P/O (intended to drive the expression of the C-terminal fragment) and the DNA coding for the CZ leucine zipper had been deleted. There was no clear start codon for the C-terminal fragment. None of N-terminal fragments were fused to NZ, instead they had C-terminal peptide appendages. Several experiments were carried out that verified that methylation required both gene fragments on the plasmid, which included (1) retransformation of the plasmids into fresh ER2267 cells and testing methylation in the transformants via the *Hha*I protection assay (2) sequencing and restriction mapping to make sure that the plasmid did not contain a reassembled full *M.HhaI* gene and (3) constructing plasmid variants in which individual gene fragments were deleted and showing that the methylation did not occur with these constructs via the *Hha*I protection assay. We speculate that low level transcription from a spurious promoter on the plasmid is occurring followed by internal translation initiation within or slightly before the C-terminal gene fragment. Extended transcription from the Tet promoter is unlikely because it has a strong transcription terminator. Although translation initiation within the C-terminal gene fragment need not be occurring at an ATG codon, we note that for all four positives (P1–P4) there is an ATG codon within the region of overlap between the two gene fragments that might serve as the translation initiation site for the C-terminal fragment (Fig. 2).

Construction of Library 2 and selection for active M.HhaI MTases

Our inability to make a diverse library of C-terminal fragments as well as the deletion of the *tac* P/O in the selected clones suggests that many of the C-terminal fragments of M.HhaI are toxic to *E. coli*, perhaps owing to their DNA binding activity. In order to test this hypothesis and to create a less biased library, we cloned the incremental truncation-ready MHhaI[29–327] DNA into a new plasmid downstream from an arabinose promoter (in order to tightly control the C-terminal library expression) to create pAR-MHhaI[29–327]. When an incremental truncation library from this plasmid was created, the number of transformants was much larger than in the library created from the corresponding plasmid with the *tac* promoter and the distribution of fragment lengths was not biased. This library was subcloned into the previous one constructed for the N-terminal fragment to make a second crossed library (Library 2). Library 2 underwent three rounds of selection as before, the only difference being that 0.1% arabinose was added to induce expression of the C-terminal fragment. Analysis of the plasmid DNA of 30 individual colonies revealed that 29 of the selected library members contained a recombined plasmid with a full-length *M.HhaI* gene. However, one library member (P5) was found in which there was no recombination (Fig. 2). Unlike the previous positives, the promoter and leucine zipper for the C-terminal fragment of P5 was not deleted. P5 did not contain a leucine zipper on the N-terminal fragment but had a peptide extension instead. Experiments identical to those performed with P1–P4 were performed to confirm the heterodimeric nature of the DNA methylation activity of P5. We also constructed a plasmid pair in which the N-terminal fragment was coded for on pDIM-N7 and the C-terminal fragment was coded for on pAR4. Only cotransformants of these plasmids had methylated DNA.

Characterization of P5 heterodimeric M.HhaI methyltransferase

In order to test whether or not the small polypeptide appended to the N-terminal fragment and the C-terminal leucine zipper were essential for protein fragment complementation in P5, we constructed variants lacking these appendages and tested their DNA methylation ability via the *Hha*I protection assay (Fig. 4A). These experiments where carried out with the N-terminal fragment gene fragment on pDIM-N7 and the C-terminal gene fragment on pAR4 in order to facilitate the mixing and matching of different constructs. The CZ zipper was found to be essential for complementation with N-terminal constructs containing the peptide extension and could not be replaced by an unrelated protein (zinc finger QQR [20]). Conversely, deletion of the N-terminal fragment's peptide extension did not have a negative effect on complementation with C-terminal fragments containing the CZ zipper. Deletion of both the peptide extension and CZ resulted in partial protection (>75%) of the plasmid DNA from the *Hha*I protection assay. The results with *Hha*I digestion were confirmed by similar results obtained with the *AscI* protection assay. *AscI* digestion is also blocked by M.HhaI activity and only a single *AscI* enzyme site exists in each plasmid.

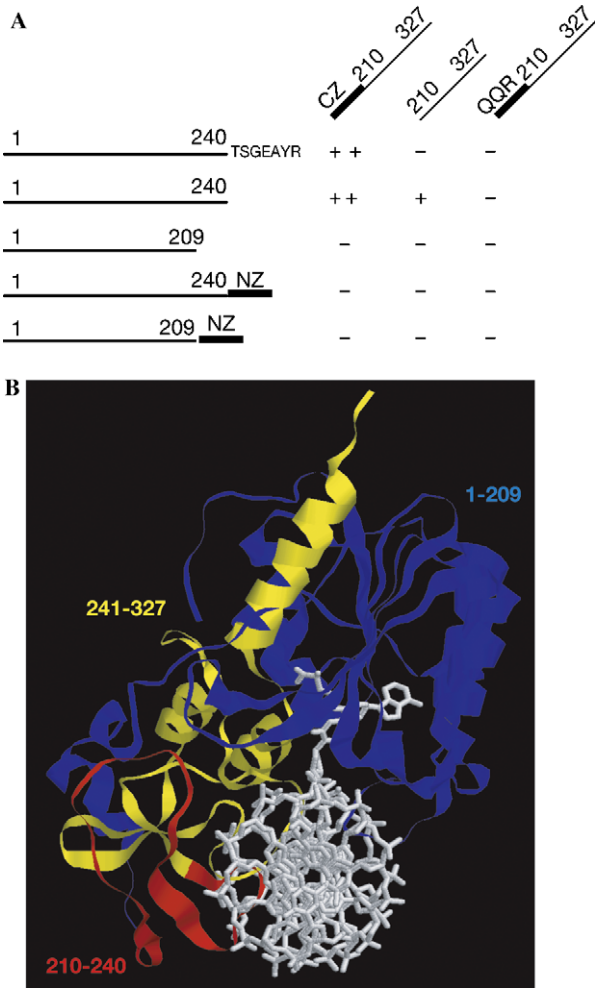

Fig. 4. Characterization of P5. (A) In vivo methylation activity of different variants of P5. N-terminal fragments are indicated in the rows and C-terminal fragments are indicated in the columns. The M.HhaI amino acids numbers at the start and end of fragments are shown. Amino acids sequences appended to the beginning or end of the fragments are shown. Amino acid sequences of the fragments are based on DNA sequencing. ++, 100% methylation; +, 75–100% methylation; –, no methylation. (B) Ribbon diagram of the crystal structure [2] of M.HhaI bound to DNA (white) color-coded to indicate the fragments of P5. Amino acids solely on the N-terminal fragment are indicated in blue. Amino acids solely on the C-terminal fragment are indicated in yellow. Amino acids on both the C-terminal and N-terminal fragment are shown in red. (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this paper.)

There are 30 amino acids of overlap between the two complementing protein fragments of P5. We constructed variants in which the fragment pairs lacked any overlap. We also constructed fragment pairs in which the NZ leucine zipper was appended to the N-terminal fragment, either with or without the 30 amino sequence overlap. None of these constructed pairs conferred any DNA methylation activity as judged by the *Hha*I or *AscI* protection assays (Fig. 4A). We postulate that dimerization of NZ and CZ, if it occurs, might prevent proper association of N-terminal and C-terminal halves of M.HhaI or distort its final conformation of the associated protein fragments. Alternatively the N-terminal antiparallel leucine zipper could interfere with the DNA binding or the catalytic activity of the reconstituted enzyme.

Comparison of M.HhaI fragment pairs with M.AquI, M.BssHII, and M.BspRI

Interestingly, the split point in P5 resides in between motif VIII and TRD domain, which is the same region where M.AquI is split and the site where M.BssHII is circularly permuted. This indicates an inherent tolerance to disruption of the peptide backbone in this region that has implications for the folding, stability and protein engineering of 5mC MTases. In M.HhaI, this region includes two of the five β-sheets within the small domain (Fig. 4B) and is after the 'long linker,' which at the outset might have been considered as a likely candidate site for successful bisection for complementing fragments. Although the unknown start sites for the C-terminal fragments of P1–P4 prevents any definitive statements about their nature, it appears that P1–P3 are split between motif VII and the 'long linker' and P4 is split somewhere in a region that includes the long linker.

Previous studies have shown that several different fragment pairs of M.BspRI are active together but not alone. One of these pairs is in the same region as P5 and another is near the start of the C-terminal DNA fragment in P1–P3 (Fig. 2). Partially overlapping fragments of M.BspRI showed complete methylation of substrates while non-overlapping fragments appeared to show only partial methylation of the substrate. We also observed a requirement for overlapping fragments for full methylation. When we constructed variants of P5 lacking the overlap, no methylation was observed. Our selection procedure preferentially selects for complete methylation since the library undergoes several rounds of HhaI digestion to recover complementing fragment partners. Overlapping fragment pairs have been seen in many other proteins as well [15]. We speculate that the overlap may offer benefit in the folding and assembly of the enzyme by protecting the exposed hydrophobic surfaces in the fragments before they find their complementing partner.

In summary, this study shows that the incremental truncation methodology could be used to identify active complementing protein fragment pairs of a monomeric mC5 MTase that could methylate DNA in vivo. Fragment pair P5 was split in the variable region between motifs VIII and the TRD, consistent with polypeptide breaks previously reported in natural MTases (M.AquI and BssHII) and engineered split MTases (M.BspRI). In the future, new functional domains may be able to be inserted in this region of M.HhaI without affecting its primary function of site-specific methylation, thereby conferring additional properties to the enzyme.

Acknowledgments

We thank Charles P. Scott for providing plasmid pAR4 and Andrew Hamilton for the DNA encoding leucine zippers NZ and CZ We also thank Chi V. Dang for his support and encouragement of this work. This study was made possible in part through the support from the National Institutes of Health (GM 53923 and GM 72794), The W.W. Smith Charitable Trust, and the Maryland Cigarette Restitution Fund Research Grant to the Johns Hopkins Medical Institutions (FY 03). Wonchae Choe was supported in part by an NIH/NHLBI postdoctoral training grant (T32 HL07525).

References

[1] X. Cheng, S. Kumar, S. Klimasauskas, R.J. Roberts, Crystal structure of the *Hha*I DNA methyltransferase, Cold Spring Harb. Symp. Quant. Biol. 58 (1993) 331.

[2] S. Klimasauskas, S. Kumar, R.J. Roberts, X. Cheng, *Hha*I methyltransferase flips its target base out of the DNA helix, Cell 76 (1994) 357–369.

[3] J. Posfai, A.S. Bhagwat, G. Posfai, R.J. Roberts, Predictive motifs derived from cytosine MTases, Nucleic Acids Res. 17 (1989) 2421–2435.

[4] X.D. Cheng, Structure and function of DNA MTases, Annu. Rev. Biophys. Biomol. Struct. 24 (1995) 293.

[5] U.T. Sankpal, D.N. Rao, Mutational analysis of conserved residues in *Hha*I DNA methyltransferase, Nucleic Acids Res. 30 (2002) 2628–2638.

[6] S.Y. Xu, J.P. Xiao, J. Posfai, R. Maunus, J. Benner, Cloning of the BssHII restriction-modification system in *Escherichia coli*: BssHII methyltransferase contains circularly permuted cytosine-5 methyltransferase motifs, Nucleic Acids Res. 25 (1997) 3991.

[7] C. Karreman, A. deWaard, Agmenellum quadruplicatum M.AquI, a novel modification methylase, J. Bacteriol. 172 (1990) 266–272.

[8] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson, Multiple sequence alignment with the Clustal series of programs, Nucleic Acids Res. 31 (2003) 3497–3500.

[9] G. Posfai, S.C. Kim, L. Szilak, A. Kovacs, P. Venetianer, Complementation by detached parts of GGCC-specific DNA MTases, Nucleic Acids Res. 19 (1991) 4843–4847.

[10] J.M. Bujnicki, Sequence permutations in the molecular evolution of DNA MTases, BMC Evol. Biol. 2 (2002) 3.

[11] H. Pinarbasi, E. Pinarbasi, D. Hornby, Recombinant alpha and beta subunits of M.AquI constitute an active DNA methyltransferase, J. Biochem. Mol. Biol. 35 (2002) 348–351.

[12] J. Walter, T.A. Trautner, M. Noyer-Weidner, High plasticity of multispecific DNA MTases in the region carrying DNA target recognizing enzyme modules, EMBO J. 11 (1992) 4445–4450.

[13] S. Mi, R.J. Roberts, How M.MspI and M.HpaII decide which base to methylate, Nucleic Acids Res. 20 (1992) 4811–4816.

[14] M. Ostermeier, A.E. Nixon, J.H. Shim, S.J. Benkovic, Combinatorial protein engineering by incremental truncation, Proc. Natl. Acad. Sci. USA 96 (1999) 3562–3567.

[15] D.E. Paschon, Z.S. Patel, M. Ostermeier, Enhanced catalytic efficiency of aminoglycoside phosphotransferase (3′)-IIa achieved through protein fragmentation and reassembly, J. Mol. Biol. (submitted)..

[16] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

[17] D. Paschon, M. Ostermeier, Construction of protein fragment complementation libraries using incremental truncation, Methods Enzymol. 388 (2004) 103–116.

[18] C.P. Scott, E. Abel-Santos, M. Wall, D.C. Wahnon, S.J. Benkovic, Production of cyclic peptides and proteins in vivo, Proc. Natl. Acad. Sci USA 96 (1999) 13638–13643.

[19] I. Ghosh, A.D. Hamilton, L. Regan, Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein, J. Am. Chem. Soc. 122 (2000) 5658.

[20] Y. Shi, J.M. Berg, Specific DNA–RNA hybrid binding by zinc finger proteins, Science 268 (1995) 282–284.