Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases

Janine G. Borgaro and Zhenyu Zhu*

New England Biolabs Inc., 240 County Road, Ipswich, MA 01938, USA

Received December 10, 2012; Revised January 29, 2013; Accepted January 30, 2013

ABSTRACT

In T4 bacteriophage, 5-hydroxymethylcytosine (5hmC) is incorporated into DNA during replication. In response, bacteria may have developed modification-dependent type IV restriction enzymes to defend the cell from T4-like infection. PvuRts1I was the first identified restriction enzyme to exhibit specificity toward hmC over 5-methylcytosine (5mC) and cytosine. By using PvuRts1I as the original member, we identified and characterized a number of homologous proteins. Most enzymes exhibited similar cutting properties to PvuRts1I, creating a double-stranded cleavage on the 3' side of the modified cytosine. In addition, for efficient cutting, the enzymes require two cytosines 21–22 nt apart and on opposite strands where one cytosine must be modified. Interestingly, the specificity determination unveiled a new layer of complexity where the enzymes not only have specificity for 5β-glucosylated hmC (5βghmC) but also 5α-glucosylated hmC (5αghmC). In some cases, the enzymes are inhibited by 5βghmC, whereas in others they are inhibited by 5αghmC. These observations indicate that the position of the sugar ring relative to the base is a determining factor in the substrate specificity of the PvuRts1I homologues. Lastly, we envision that the unique properties of select PvuRts1I homologues will permit their use as an additive or alternative tool to map the hydroxymethylome.

INTRODUCTION

DNA modifications are present across many forms of life. One of the more commonly identified epigenetic modifications is cytosine methylation [5-methylcytosine (5mC)]. Depending on its location in the DNA, a 5mC modification performs a variety of biological roles, from protection against restriction enzymes to gene regulation. Prokaryotes contain restriction-modification systems where DNA methyltransferases modify the host DNA, and restriction enzymes serve as a protector against non-methylated foreign DNA. However, evolution has allowed several bacteriophages to survive in which modified bases that are resistant to many restriction enzymes are incorporated into their genome (1). One well-known example is bacteriophage T4. During replication, all cytosines are replaced with 5-hydroxymethylcytosine (5hmC), which is further modified by α and β glucosylation of the hydroxymethyl group (2). Even though 5hmC is resistant to most restriction enzymes, McrA (3), McrBC (4) and Type IV SauUSI (5) have been shown to specifically restrict its infection in vivo. Additionally, PvuRts1I (6) and GmrSD UT and CT (7) have shown to restrict DNA containing 5-glucoylhydroxymethylated DNA (5ghmC). T4 phage DNA consists of 30% beta glucosylated 5hmC and 70% alpha glucosylated 5hmC (8).

In eukaryotes, 5mC has been associated with the regulation of transcriptional activity and shown to affect fundamental processes such as development, imprinting and genome stability (9). Recently, 5hmC was discovered in human brain tissue and in mouse embryonic stem cells (10,11) and has subsequently generated much interest within the scientific community. 5hmC was identified as the oxidative product of 5mC, a reaction catalyzed by the ten eleven translocation (TET) family enzymes (12). Furthermore, mutations in human TET2 are associated with myeloid malignancies, further supporting the physiological relevance of 5hmC (13).

Even though the exact role of 5hmC in higher organisms is still unclear, current literature proposes two possible functions. It can serve as an intermediate for cytosine demethylation (14–16) or it may influence chromatin structure by altering the binding of methyl CpG binding proteins (17,18). To fully elucidate the biological function of this new modification, methods to map the hydroxymethylome need to be developed.

There are currently three reported methods for single base-resolution hydroxymethylome mapping. Two of these methods, o xoBS-seq and TAB-seq, use bisulfite sequencing coupled with either chemical or enzymatic oxidation, respectively (19,20). In these methods, 5mC and 5hmC are read as cytosine after bisulfite sequencing, while...
further oxidized products of 5hmC (5-formylcytosine or 5-carboxylcytosine) are deaminated and subsequently read as thymine (21). The third method, Aba-seq, uses the enzymatic properties of AbaSI (formally designated AbaSDFI), a member of the PvuRts1I restriction enzyme family shown to exhibit high specificity for 5hmC over 5mC and C, cleaving at a fixed distance away from the modification (22). Even though all three methods can map the 5hmC genome to base resolution, Aba-seq has certain advantages: using a restriction enzyme preserves the quality of the DNA, is semi-quantitative and allows less abundant 5hmC sites to be accurately identified (23).

Owing to the increasing evidence for the importance of 5hmC in mammalian epigenetics and the success of Aba-seq in mapping the 5hmC genome to base resolution, we have sought to determine the in vitro biochemical properties of PvuRts1I homologues identified in REBASE. We thus characterized >25 family members focusing on comparing their substrate selectivity for different forms of cytosine modifications, in addition to their cut sites and recognition site requirements. Interestingly, in addition to observing differential cutting on beta-glucosylated T4 DNA (T4β), we also observed differential specificities for alpha-glucosylated T4 DNA (T4α) among the homologues. For example, AbaSI cutting is greatly inhibited by 5α-glucosylated hmC (5αghmC) when compared with 5β-glucosylated hmC (5βghmC), while PvuRts1I cutting is enhanced on 5αghmC when compared with 5βghmC.

MATERIALS AND METHODS

Cloning, expression and purification of PvuRts1I homologues

C-terminally intein-tagged PvuRts1I homologous proteins were purified to high homogeneity from Escherichia coli strain T7 Express [New England Biolabs (NEB) #C2566] essentially as described (22). The sequences encoding the gene for a majority of the PvuRts1I enzyme family (Table 1) were optimized using Optimizer (24) and synthesized by Integrated DNA Technologies Inc (San Jose, CA, USA).

A large range of concentrations (0.016–4.5 mg/ml) were used for the characterization experiments due to variation in the expression levels of the proteins. The units of each enzyme used in the experiments could be calculated from the specific activities, resulting in a range of 1–400 units (Table 1). One unit of enzyme is defined as the amount to reach 1.2, after which protein expression was induced with 0.2 mM isopropylthio-β-galactoside. After incubating at 16°C overnight, cells were harvested by centrifugation, and the cell-free extract was loaded onto a 5 ml DEAE column, followed by a 5 ml HiTrap Heparin HP column, where the DEAE and HiTrap Heparin HP columns were pre-equilibrated in eluent A at pH 7.5 and the HiTrap Q HP column was pre-equilibrated in eluent A at pH 8.0. AGT was eluted from each column with a linear gradient of NaCl, and protein purity was analyzed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (PAGE).

Glucosylation assay for T4gt DNA

A standard glucosylation assay consisted of a fixed concentration of uridine diphosphate (UDP)-glucose [1-3H] (American Radiolabeled Chemical, Inc; ART 0525), 100 ng of T4gt DNA and varying concentrations from a 2-fold dilution series of AGT in NEB buffer 4 for 2 h at 37°C. The reactions were stopped by flash freezing in an ethanol/dry ice bath. The samples were processed by applying the thawed reaction mixture to a 2.5 cm DE81 membrane (GE Healthcare# 3658-325) under air pressure using a vacuum manifold (Millipore). The reaction was washed three times with 0.2 M ammonium bicarbonate, followed by three times with deionized water and finally, three times with ethanol. The membranes were dried, and the amount of tritium incorporation was determined by standard scintillation counting for 1 min (Perkin Elmer TriCarb 2900TR).

DNA substrates for specificity determination

Specificity of enzymes were determined on non-methylated lambda DNA (C), XP12 DNA [5mC (25)], phage T4gt DNA (containing non-glucosylated 5hmC), T4β (containing 5βghmC) and T4α (containing

Thyroglobulin (669 kDa), apoferritin (443 kDa), β-amylase (200 kDa), bovine serum albumin (BSA; 66 kDa) and carbonic anhydrase (29 kDa) in the equilibration buffer. A standard curve was generated by plotting the molecular masses on a logarithmic scale versus Ve (elution volume)/Vo. After calibration, the column was re-equilibrated in the same buffer, and the homologues, varying in concentration from 200 μg to 3 mg (depending on the stock concentrations) were applied to the column. Ve/Vo for each protein was calculated and the molecular weights were determined from the standard curve.

T4 α-glucosyltransferase

The pAl117-α-glucosyltransferase (AGT) plasmid containing the coding region for AGT was transformed into a dem- E. coli strain T7 Express (NEB # C2566). After selection on solid LB media containing ampicillin (100 μg/ml), individual colonies were used to inoculate 1 Luria broth (LB) media containing ampicillin (100 μg/ml). The culture was incubated at 37°C until the OD600 reached 1.2, after which protein expression was induced with 0.2 mM isopropylthio-β-galactoside. After incubating at 16°C overnight, cells were harvested by centrifugation, suspended in 25 ml of 20 mM Tris–HCl, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 100 mM NaCl buffer, pH 7.5 (eluent A) and sonicated at 4°C. Cell debris was removed by centrifugation, and the cell-free extract was loaded onto a 5 ml DEAE column, followed by a 5 ml HiTrap Heparin HP column and then a 5 ml HiTrap Q HP column, where the DEAE and HiTrap Heparin HP columns were pre-equilibrated in eluent A at pH 7.5 and the HiTrap Q HP column was pre-equilibrated in eluent A at pH 8.0. AGT was eluted from each column with a linear gradient of NaCl, and protein purity was analyzed by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (PAGE).
5zghmC). Non-methylated lambda DNA was purchased from Sigma (# D3654). XP12, T4wt and T4gt genomic DNA were purified from phage cultures. DNA containing either 5zghmC or 5zghmC was obtained by further modification of T4gt DNA by the T4 β-glucosyltransferase [BGT], NEB #M0357] and AGT, respectively. The relative specificities of the PvuRtsI1 homologues were determined by incubating 100 ng of each DNA substrate with a 2-fold serial dilution of each enzyme in Diluent E (250 mM potassium acetate, 10 mM Tris–acetate buffer, pH 8.0, 0.2 mg/ml BSA) in NEB buffer 4 for 20 min at room temperature. The reaction products were then resolved on a 0.8% agarose gel.

### Substrates for cleavage site determination and recognition site requirements

The DNA oligonucleotides containing a top-strand 5hmC modification and 3' fluorescein amidite (FAM) labels were synthesized by Integrated DNA Technologies. The sequences are as follows: 5'-CCA TAC ATA TCC TT ACT TCT CCT AA(5hmC) GTG GAT GAT AAA GGT AGT TTA TGT GGA-3' as a template in the presence of dhmCTP/dATP/dGTP/dTTP and Taq DNA Polymerase (NEB #M0273). Purification yielded a final concentration of 25–30 ng/μl of double-stranded oligonucleotide. Subsequently, both the 5' and 3'FAM-labeled double-stranded oligonucleotides were glucosylated by an overnight incubation with BGT. The cleavage sites were then determined by incubating 100–150 ng of double-stranded oligonucleotide with each enzyme for 30 min at room temperature. The reaction products were resolved using a 20% polyacrylamide 7 M urea denaturing gel.

The oligonucleotides containing 5hmC used in Figure 4 were synthesized by the NEB Organic Synthesis Division. Similar to the preparation of the 3'FAM oligonucleotides used for the cut site determination, equal concentrations of top and bottom strands were mixed to yield a 10 μM final concentration of double-stranded substrate. The oligonucleotides were annealed by heating to 95°C followed by gradually cooling the solution to room temperature. To determine the recognition-site requirements for the enzymes, five different synthetic oligonucleotides (A, C/C, C, 5mC and 5hmC) were synthesized. Synthetic oligonucleotide A contains an 5hmC modification and a non-cytosine residue 22 nt away and on the opposite strand, oligonucleotide C/C contains two cytosines 22 nt apart and on opposite strands, oligonucleotide C consists of an 5hmC modification and a cytosine 22 nt away and on the opposite strand, oligonucleotide 5mC contains an 5hmC modification and an 5mC modification 22 nt away and on the opposite strand and lastly, oligonucleotide 5hmC contains two 5hmC modifications 22 nt apart and on opposite strands (Figure 4). The sequences of the substrates are listed in Table 2. Each substrate (77 ng) was

![Table 1. PvuRtsI1 homologue information](image)

| Enzyme   | Specific activity units/mg | Length, similarity (%) and identity (%) | Source                                                                 | Genbank accession |
|----------|---------------------------|----------------------------------------|-----------------------------------------------------------------------|-------------------|
| PvuRtsI1 | 65 000                    | 293, 100, 100                          | Proteus vulgaris (Rts1)                                               | AP004237.1        |
| AbaSI    | 16 000                    | 317, 47.8, 35.4                        | Acinetobacter baumannii SDF                                          | CAP0265.1         |
| AbaHI    | 17 800                    | 317, 48.9, 36.5                        | Acinetobacter baumannii AB210                                        | EGK46224.1        |
| AbaAU    | 20 000                    | 317, 48.2, 35.7                        | Acinetobacter baumannii AB900                                        | ZP_04663830.1     |
| AbaCI    | 80 000                    | 317, 48.1, 35.7                        | Acinetobacter baumannii ATCC 19606                                   | ACQ010000053.1    |
| AbaDI    | 8400                      | 317, 47.8, 35.4                        | Acinetobacter baumannii DI27979                                     | AERZ01000045.1    |
| AbaBGI   | 11 000                    | 317, 48.2, 35.8                        | Acinetobacter baumannii ATCC 17978                                   | CP000521.1        |
| AbaTI    | 8500                      | 317, 47.8, 35.4                        | Acinetobacter baumannii MDR-TJ                                       | ABTJ_00252.1      |
| AbaUI    | 7600                      | 317, 48.1, 35.4                        | Acinetobacter baumannii UMB003                                       | AEMP01000006.1     |
| AcaAPI   | 4400                      | 317, 47.8, 35.0                        | Acinetobacter calcoaceticus PHEA-2                                   | ADY83307.1        |
| BhiDI    | 104 000                   | 311, 43.3, 32.0                        | Bifidobacterium bifidum NCIMB 41171                                  | ABQP01000008      |
| BmeDI    | 23 000                    | 300, 40.0, 30.5                        | Bacillus megaterium strain DSM319                                     | ADJ40396.1        |
| EsaMMI   | 2000                      | 305, 55.1, 43.1                        | Marine metagenome                                                     | EDB78928.1        |
| EsaNI    | 2200                      | 299, 47.8, 39.6                        | Marine metagenome                                                     | EBQ05151.1        |
| Mtc37I   | 5000                      | 265, 46.1, 36.1                        | Microbacterium testaceum SilB037                                     | BA74504.1         |
| PatT1    | 1400                      | 306, 51.4, 36.5                        | Pseudomonas atlantica T9c ATCC BAA-1087                               | ABG40021.1        |
| PfrCI    | 364                      | 292, 41.4, 29.3                        | Propionibacterium freudenreichii subsp. Shermanii CIRM-BIA1           | ADY83307.1        |
| PpeHI    | 13 000                    | 294, 72.5, 64.6                        | Proteus pennieri ATCC 3598                                            | EEG85915.1        |
| PxyI     | 93                       | 290, 46.6, 36.0                        | Paraprevotella xylaniphila YIT 11841                                  | EGG52627.1        |
| YkrI     | 9000                      | 299, 49.6, 36.1                        | Yersinia kristensenii ATCC33638                                       | EEP89557.1        |

*a*aminoo acid.

*b*GCG best fit result (Version 11.1, Accelrys Inc., San Diego, CA, USA).
incubated with enzyme for 30 min at room temperature. The reaction products were then resolved using a 5% agarose gel.

RESULTS

A number of PvuRts1I homologues were identified by blasting the PvuRts1I protein sequence against the NR and ENV_NR databases [(26), Table 1]. Seven homologues have identical sequences and are therefore omitted from the comparison.

Of the 28 hits from NCBI/BLAST, five were inactive when tested for activity against T4wt and T4gt in crude lysates and three were not purified using the methods described here. The remaining recombinant proteins were fused with a cleavable intein- and chitin-binding domain and were purified to near homogeneity.

Figure 1A shows an example of the purity of the homologues and indicates that they are relatively pure. The remaining homologues also show a similar level of purity to those pictured. Furthermore, size exclusion chromatography (Figure 1B) indicates that all of the homologues are likely dimers (Table 3).

In vitro conversion rate of T4 DNA by the α-glucosyltransferase

BGT can fully glucosylate T4 DNA in vitro (27). To determine the percent of glucosyl incorporation by AGT, the extent of glucosylation by AGT and BGT on a common substrate was compared. Incorporation was comparable, indicating an in vitro conversion rate of 5hmC to 5αghmC by AGT of 100% in T4 DNA.

Relative selectivity of the PvuRts1I homologues

AbaSI, a PvuRts1I homologue, has previously been characterized as a modification-dependent restriction endonuclease that recognizes 5hmC as well as 5ghmC with little to no activity toward 5mC and C (22). With the hopes of discovering an enzyme with even higher selectivity than AbaSI, we sought to characterize a set of 20 PvuRts1I homologues. A similar method to that used in the initial substrate selectivity determination of PvuRts1I (22) was used to determine the substrate selectivity of the PvuRts1I homologues. Each homologue was assayed on non-methylated lambda DNA (containing unmodified cytosines), phage XP12 DNA (containing 5mC), phage T4gt DNA (containing non-glucosylated 5hmC DNA), and...
phage T4β DNA (containing 5βghmC) and phage T4α DNA (containing 5αghmC). The relative selectivity for each enzyme is defined as the ratio of activity on the different modified cytosine substrates. The homologues with the highest selectivity include AbaUI, which exhibits a relative selectivity of 5hmC:5αghmC:5βghmC:5mC:C = 512:16:8192:8:ND (ND: non-detectable, meaning no apparent difference between cut and uncut substrate, Figure 2); AbaA1, which exhibits a relative selectivity of 5hmC:5αghmC:5βghmC:5mC:C = 1024:128:16384:16:ND and BbiDI, which exhibits a relative selectivity of 5hmC:5αghmC:5βghmC:5mC:C = 4096:4096:256:2:1. Figure 2 illustrates the comparison of the specific activities for all the homologues. In contrast to the β-glucosyl modification, the α-glucosyl modification resulted in varying effects on the relative selectivity of the homologues. For AbaSI, the 5αghmC modification had inhibitory effects of 1/500 when compared with 5βghmC. In contrast, for PvuRts1I, the 5αghmC modification enhanced selectivity by 32-fold when compared with 5βghmC. These modifications are important because even though they are not known to exist in the human genome, in vitro 5hmC can be converted to 5αghmC and 5βghmC by BGT and AGT, respectively.

### Cleavage properties

To determine the cleavage properties for the PvuRts1I homologues, two different substrates were designed, a hemi-glucosylhydroxymethylated oligonucleotide with 3′FAM labels (Figure 3A) and fully glucosylhydroxymethylated oligonucleotide with 5′FAM labels (Figure 3B), and subjected to enzyme digestion, which would allow the detections of each of the cleavage products. A 3′FAM-labeled substrate allows us to determine the cleavage pattern of a single 5ghmC site and the cleavage site on the same strand of the modification. The 5′FAM-labeled substrate will allow us to determine the cleavage site on the opposite strand of the modification and whether the cleavage properties are altered if there is a fully hydroxymethylated site. Denaturing PAGE allowed for the single-base resolution of the small digested fragments, which were subsequently compared with size markers and with the cleavage pattern produced by AbaSI.

It has previously been shown that AbaSI exhibits a double-stranded cleavage to the 3′ side of a cytosine modification at N11–13/N9–10 (22). If the enzymes exhibit a similar cutting pattern to AbaSI, cleavage of the hemi-glucosylhydroxymethylated 3′FAM-labeled substrate would result in two labeled fragments of 15(+/−) and 39(+/−) nt. Digestion of the 3′FAM-labeled substrate by the homologues showed the same cutting pattern as AbaSI, indicating a cleavage site of 12–13 nt away on the same strand of the modified cytosine (Figure 3A).

For the fully glucosylhydroxymethylated 5′FAM-labeled substrate, if AbaSI recognizes only one 5ghmC modification, cleavage would result in two labeled fragments of 17(+/−) and 39(+/−) nt. If AbaSI recognizes both 5ghmC modifications, cleavage would occur on both sides of the 5ghmC sites, resulting in two FAM-labeled fragments of 17(+/−) nt. Digestion of the

### Table 3. Oligomeric state of the homologues as determined by gel filtration

| Enzyme     | Determined molecular weight (kDa, DM) | Calculated monomeric molecular weight (kDa, CM) | Ratio (DM/CM) |
|------------|--------------------------------------|-----------------------------------------------|---------------|
| PvuRts1I   | 74 ± 8                               | 34.283                                        | 2.1 ± 0.2     |
| AbaSI      | 91 ± 8                               | 37.665                                        | 2.4 ± 0.2     |
| AbaHI      | 96 ± 8                               | 37.440                                        | 2.6 ± 0.2     |
| AbaAI      | 93 ± 8                               | 37.482                                        | 2.5 ± 0.2     |
| AbaCI      | 91 ± 8                               | 37.397                                        | 2.4 ± 0.2     |
| AbaDI      | 84 ± 8                               | 37.508                                        | 2.2 ± 0.2     |
| AbaBGI     | 77 ± 8                               | 37.412                                        | 2.1 ± 0.2     |
| AbaTI      | 85 ± 8                               | 37.509                                        | 2.3 ± 0.2     |
| AbaUI      | 93 ± 8                               | 37.582                                        | 2.5 ± 0.2     |
| AcaPI      | 96 ± 8                               | 37.312                                        | 2.6 ± 0.2     |
| BbiDI      | 84 ± 8                               | 36.337                                        | 2.3 ± 0.2     |
| BmeDI      | 84 ± 8                               | 39.821                                        | 2.1 ± 0.2     |
| EsaMMI     | 77 ± 8                               | 35.258                                        | 2.2 ± 0.2     |
| EsaNI      | 77 ± 8                               | 35.005                                        | 2.2 ± 0.2     |
| Mte371     | 96 ± 8                               | 36.464                                        | 2.6 ± 0.2     |
| PatTI      | 77 ± 8                               | 33.537                                        | 2.3 ± 0.2     |
| PfrCI      | 80 ± 8                               | 35.342                                        | 2.3 ± 0.2     |
| PpeHI      | 84 ± 8                               | 34.677                                        | 2.4 ± 0.2     |
| PxyI       | 80 ± 8                               | 35.314                                        | 2.2 ± 0.2     |
| YkrI       | 84 ± 8                               | 33.802                                        | 2.5 ± 0.2     |
5′FAM-labeled substrate by the homologues result in a mixture of products from cleavage on only one side [39(+/-), 17(+/-) nt, Figure 3B (1a, 1b)] and on both sides [17(+/-), Figure 3B (2)] of the modifications. From the accurate measurement of the 17-nt fragment, we can deduce that the cleavage on the opposite strand is 9–10 nt away from the modification. The 39(+/-) nt fragment in Figure 3B is labeled as an intermediate because if the reaction went to completion, only the 17(+/-) nt fragments will be observed. Due to the resolving power of the gel, only the size of the smaller fragments of DNA could be accurately determined. However, by simply subtracting the smaller fragment from the total length, the size of the larger fragment could also be derived. The cut site for all the enzymes is predominantly N11–13/N9–10 (Table 4). The digestion range for the cut site is owing to some enzymes exhibiting minimal base wobbling.
reaction went to completion, it would have been completely digested into smaller fragments. If only one modification is being recognized, we would expect to see equal intensities of the 39(+/-) and 17(+/-) nt bands. Instead, the 17(+/-) nt band is more intense, indicating that both modifications are being recognized. Incomplete enzyme digestion is often seen when using synthetic oligonucleotides compared with genomic DNA. In addition, as the cleavage site for the homologues on the fully hydroxymethylated substrate is the same as that reported for AbaSI on a hemi-hydroxymethylated substrate (22), we can conclude that the presence of two modifications does not alter the cleavage properties of the enzymes. Overall, our results suggest that most of the homologues do not affect the cleavage properties of the enzymes. Specifically, BmeDI has a long C-terminus of ~40 amino acids that extends beyond the end of the alignment of all the other homologues. This observation led us to create a phylogenetic tree based on the sequences of the homologues. The tree showed BmeDI on a branch of its own, indicating that it evolved differently from its other family members and supports our hypothesis that BmeDI is a unique enzyme. Further studies are required to determine the exact reason or amino acids that are responsible for the difference in cut site of BmeDI.

DISCUSSION

Enzymes have been used successfully for decades to answer important questions in biology. Here we present the characterization of a special class of enzymes specific for modified cytosines in DNA. These enzymes have been used as a defense mechanism in the struggle between unicellular organisms and their viruses. PvuRts1I was among the first of these enzymes identified and was shown to restrict T-even bacteriophages that contain 5hmC or 5ghmC (6). Characterization of the PvuRts1I family enzymes shows that, like AbaSI, all of the enzymes exhibit DNA-modification-dependent endonuclease activity with similar cleavage properties. Specifically, most of the homologues generate a double-stranded cut on the 3' side of the modified cytosine at a distance of CN$_{11-13}$ on the top strand and N$_{9-10}$G on the bottom strand (Figure 4). Additionally, for efficient cleavage, the enzymes require two cytosines separated by 21–22 nt where one cytosine must be modified (Figure 4). The observation that two cytosines are required for efficient cleavage agrees with our finding that the homologues form a dimer in solution.

There is one outlier, BmeDI, which generates a double-stranded cut on the 3' side of the modified cytosine at a distance of CN$_{2-3}$ on the top strand and N$_{0-1}$G on the bottom strand. After performing a sequence alignment with the PvuRts1I homologues, it is clear that the sequence for BmeDI is unique. Specifically, BmeDI has a long C-terminus of ~40 amino acids that extends beyond the end of the alignment of all the other homologues. This observation led us to create a phylogenetic tree based on the sequences of the homologues. The tree showed BmeDI on a branch of its own, indicating that it evolved differently from its other family members and supports our hypothesis that BmeDI is a unique enzyme. Further studies are required to determine the exact reason or amino acids that are responsible for the difference in cut site of BmeDI.

To compare and contrast the specificity of the PvuRts1I family members, we generated DNA substrates that contain different cytosine modifications. The specificity of this class of enzymes is especially important because the amount of 5hmC in the genome is extremely low in comparison with 5mC or C (28). All of the enzymes assayed exhibit different relative selectivities toward the various DNA substrates with minimal to non-existent cutting on C. Notably, AbaAI, AbaCI, AbaUI and AbaTI exhibit the highest selectivity between 5βghmC

Table 4. Cut sites of the PvuRts1I homologues

| Enzyme     | Top-strand cut | Bottom-strand cut |
|------------|----------------|-------------------|
| PvuRts1I   | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AbaSI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AbaHI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AbaAI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AbaCI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AbaDI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AbaBG1     | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AbaTI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AbaUI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| AcaPI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| BbiDI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| BmeDI      | C$_{2-3}$N$_{0-1}$G   | C$_{2-3}$N$_{0-1}$G   |
| EsaMMI     | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| EsaNI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| Mtc37I     | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| PfuTI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| PfrCI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| PpeHI      | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| PsyI       | C$_{11-13}$N$_{9-10}$G | C$_{11-13}$N$_{9-10}$G |
| YkrI       | C$_{10-13}$N$_{9-10}$G | C$_{10-13}$N$_{9-10}$G |

*a with respect to 5ghmC modification.*
and 5mC at 1000:1, while BbiDI exhibit the highest selectivity between 5αghmC and 5mC at 2000:1 (Figure 2). The observation of homologous enzymes exhibiting different relative selectivities toward their substrates is consistent with many examples in the literature where the difference of only a few amino acids can result in varied substrate specificity (29,30).

Interestingly, this comparison revealed an additional layer of complexity with the homologues exhibiting varied specificity toward 5αghmC and 5mC at 2000:1 (Figure 2). The observation of homologous enzymes exhibiting different relative selectivities toward their substrates is consistent with many examples in the literature where the difference of only a few amino acids can result in varied substrate specificity (29,30).

Figure 4. Activity of enzymes on different modified oligonucleotides. The sequence of the oligonucleotides can be found in Table 2. (A) Schematic of the five modified oligonucleotides used for the activity determination: A, C/C, C, 5mC and 5hmC. The two indicated residues on each substrate are 22 nt apart. (B) The extent of double strand cleavage on A, C/C, C, 5mC and 5hmC for each of the homologues is shown. All of the homologues have the highest activity on a substrate containing two 5hmC modifications, 22 nt apart (5hmC, blue). The activity decreases as the modification on the opposite strand changes from 5mC to C and there is no detectable cutting for most of the homologues in the absence of cytosine (A, yellow). The activities are normalized to cutting on 5hmC.

and 5mC at 1000:1, while BbiDI exhibit the highest selectivity between 5αghmC and 5mC at 2000:1 (Figure 2). The observation of homologous enzymes exhibiting different relative selectivities toward their substrates is consistent with many examples in the literature where the difference of only a few amino acids can result in varied substrate specificity (29,30).

Interestingly, this comparison revealed an additional layer of complexity with the homologues exhibiting varied specificity toward 5αghmC and 5mC DNA. For example, BbiDI, BmeDI and PvuRts1I show high selectivity for 5αghmC but are inhibited by 5βghmC, while AbaCI, AbaUI, AbaAI and AbaSI (to name a few) show high selectivity for 5βghmC but are inhibited by 5αghmC (Figure 2). T4wt DNA contains a mixture of α and β glucosylated 5hmC (8). To survive infection by T4-like phages, bacteria must contain enzymes with specificity toward either α or β 5ghmC. Enzyme active sites can be specific, and even a small change in the substrate or cofactor structure can have an immense effect on specificity. We believe the difference of the sugar ring conformation is likely attributable to differences in binding site specificity among the PvuRts1I homologues. This is supported by Gruber et al. (31) who determined that the UDP-galactopyranose mutase (UGM) has the ability to discriminate between two structurally similar substrates, UDP-galactopyranose (UDP-Galp) and UDP-glucose (UDP-Glc). Even though UDP-Galp and UDP-Glc differ only by the conformation of the sugar moiety, UGM discriminates against the latter during both binding and catalysis, which was attributed to the orientation of the sugar moieties in the active site. The crystal structure of AbaSI, once determined, will provide further insight into the mechanism of substrate specificity.

Lastly, the new observation of PvuRtsII homologues exhibiting either enhanced or inhibitory effects toward 5αghmC can present either an alternative or additive approach to map the hydroxymethylome. When designing such an experiment, it is imperative to have a high level of confidence that only 5hmC sites are being identified. This can be difficult because the amount of 5mC in the genome is much more abundant than that of 5hmC. Nevertheless, the unique characteristics of the PvuRtsII enzymes will provide this confidence. For example, AbaSI is strongly inhibited by α-glucosylation and enhanced by β-glucosylation. The DNA sample, which is β-glucosylated, will capture all 5hmC sites, in addition to sites that contain 5mC from low-level digestion by AbaSI. The DNA sample, which is α-glucosylated, will only capture sites that contain 5mC, and can serve as an experimental
control. These two sample preparations can be compared and the difference will identify only the 5hmC sites. Furthermore, as Aba-seq has already proved to be a successful method in mapping the hydroxymethylome (23), we envision a similar use for the homologues with high selectivity between 5ghmC and 5mC.

ACKNOWLEDGEMENTS

We would like to thank Drs Yu Zheng, Richard Morgan and Jurate Bitinaite for materials; Dr Zhiyi Sun for discussions and enzyme evolution analysis; Drs Geoffrey Wilson, William Jack and Richard Roberts for critical reviews and suggestions regarding the manuscript; and Jim Ellard and Donald Comb for support.

FUNDING

National Institutes of Health (NIH), SBIR [GM096723]. Funding for open access charge: NIH.

Conflict of interest statement. None declared.

REFERENCES

1. Warren,R.A. (1980) Modified bases in bacteriophage DNAs. Annu. Rev. Microbiol., 34, 137–158.
2. Kornberg,S.R., Zimmerman,S.B. and Kornberg,A. (1961) Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. J. Biol. Chem., 236, 1487–1493.
3. Mulligan,E.A. and Dunn,J.J. (2008) Cloning, purification and initial characterization of E. coli McrA, a putative 5-methylcytosine-specific nuclease. Protein Expr. Purif., 62, 98–103.
4. Sutherland,E., Coe,L. and Raleigh,E.A. (1992) McrBC: a novel structure for a dedicated target.
5. Lehman,J.R. and Pratt,E.A. (1960) On the structure of the DNA sequences.
6. Janosi,L., Yonemitsu,H., Hong,H. and Kaji,A. (1994) DNA methylation-mediated epigenetic control.
7. Rifat,D., Wright,N.T., Varney,K.M., Weber,D.J. and Black,L.W. (2008) Restriction endonuclease inhibitor IP1* of bacteriophage T4: a novel structure for a dedicated target. J. Mol. Biol., 375, 720–734.
8. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science, 324, 929–930.
9. Ito,S., D’Alessio,A.C., Taranova,O.V., Hong,K., Sowers,L.C. and Zhang,Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature, 466, 1129–1133.