9294–9305 *Nucleic Acids Research, 2011, Vol. 39, No. 21* *Published online 3 August 2011*
doi:10.1093/nar/gkr607

Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine

Hua Wang, Shengxi Guan, Aine Quimby, Devora Cohen-Karni, Sriharsa Pradhan, Geoffrey Wilson, Richard J. Roberts, Zhenyu Zhu* and Yu Zheng*

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

Received May 27, 2011; Revised July 6, 2011; Accepted July 7, 2011

ABSTRACT

PvuRts1I is a modification-dependent restriction endonuclease that recognizes 5-hydroxymethylcytosine (5hmC) as well as 5-glucosylhydroxymethylcytosine (5ghmC) in double-stranded DNA. Using PvuRts1I as the founding member, we define a family of homologous proteins with similar DNA modification-dependent recognition properties. At the sequence level, these proteins share a few uniquely conserved features. We show that these enzymes introduce a double-stranded cleavage at the 3′-side away from the recognized modified cytosine. The distances between the cleavage sites and the modified cytosine are fixed within a narrow range, with the majority being 11–13 nt away in the top strand and 9–10 nt away in the bottom strand. The recognition sites of these enzymes generally require two cytosines on opposite strand around the cleavage sites, i.e. 5′-CN$_{11–13}$↓N$_{9–10}$G-3′/3′-GN$_{9–10}$↓N$_{11–13}$C-5′, with at least one cytosine being modified for efficient cleavage. As one potential application for these enzymes is to provide useful tools for selectively mapping 5hmC sites, we have compared the relative selectivity of a few PvuRts1I family members towards different forms of modified cytosines. Our results show that the inherently different relative selectivity towards modified cytosines can have practical implications for their application. By using AbaSDFI, a PvuRts1I homolog with the highest relative selectivity towards 5ghmC, to analyze rat brain DNA, we show it is feasible to map genomic 5hmC sites close to base resolution. Our study offers unique tools for determining more accurate hydroxymethylomes in mammalian cells.

INTRODUCTION

Modification-dependent restriction endonucleases are widely present in bacterial genomes and are thought to protect hosts from invading bacteriophages containing modified DNA (1). Among many examples are the T-even phages, in which only 5-hydroxymethylcytosines (5hmC) are incorporated into the genome during replication and further modified to 5-glucosylhydroxymethylcytosine (5ghmC) by glucosyltransferases (1). Although T4 wild-type DNA is resistant to most regular restriction enzymes, there are types of modification-dependent restriction enzymes that are able to restrict their infection *in vivo*, including PvuRts1I (2,3) among a few others. For a long time, the detailed *in vitro* biochemical properties of PvuRts1I remained obscure (4).

In mammalian genomes, it is commonly believed that 5-methylcytosine (5mC) is the major form of epigenetic base modification. Recently, the observation of 5hmC as the enzymatic oxidative product of 5mC in mammalian genomes (5,6) has added an extra layer of complexity to the current understanding of epigenetic regulation and spurred rising interest in determining its genomic locations and metabolism. However, although the modified base 5hmC was discovered in bacteriophages >50 years ago (7), there are few useful methods, either enzymatic or chemical, to specifically recognize 5hmc residues and pinpoint their locations in DNA, largely due to their close structural similarity to 5mC. For example, 5mC-dependent endonucleases, such as the the MspJI family (8) or McrBC (9), do not distinguish 5mC and

*To whom correspondence should be addressed. Tel: +1 978 380 7441; Email: zhengy@neb.com
Correspondence may also be addressed to Zhenyu Zhu. Tel: +1 978 380 7237; Fax: +1 978 380 7518; Email: zhuz@neb.com

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

5hmC; 5mC-sensitive endonucleases, such as MspI or HpaII, etc., in most cases are equally affected by 5mC and 5hmC (4). The widely used bisulfite conversion method cannot differentiate between 5mC and 5hmC and reports both forms indistinguishably (10,11). Recently, the availability of 5hmC-specific antibodies has enabled a few enrichment-based methods [e.g. hMeDIP (12)]. However, the format of the experiment, based on affinity pull-down, may limit the range of its application, and the resolution of the data is still far from base resolution.

Given mounting evidence for the importance of 5hmC in mammalian epigenetics and the previous experimental observations that PvuRts1I is able to specifically recognize 5hmC both *in vivo* and *in vitro* (3), we have set out to investigate the *in vitro* biochemical properties of PvuRts1I and its homologs identified in REBASE (4). During the course of our study, Szwagierczak *et al.* (13) reported that recombinant PvuRts1I selectively cleaves 5hmC-containing DNA substrates and that the double-stranded cleavage sites are at $N_{11\text{–}12}/N_{9\text{–}10}$ on the 3′-side of the recognized 5hmC site. In addition, the authors notice that PvuRts1I prefers to cleave at symmetric sites $5'\text{-}^{hm}CN_{11\text{–}12}\downarrow N_{9\text{–}10}G\text{-}3'/3'\text{-}GN_{9\text{–}10}\downarrow N_{11\text{–}12}{}^{hm}C\text{-}5'$, suggesting a likely *in-cis* dimerization cleavage process (13). Still, there are a number of questions left unanswered. For example, it is not clear whether PvuRts1I is applicable for mapping genomic 5hmC sites along with needing details concerning its practical use. In this regard, a quantitative description of substrate selectivity on 5hmC versus 5mC or unmodified cytosine is crucial, because in most human tissue DNA, the level of 5hmC is usually on the order of 0.01% of the total nucleotide (14). During our investigation, we have observed that PvuRts1I is sensitive to different purification procedures, such that certain ions used in the buffer may quickly inactivate most of the enzyme in crude lysates (see ‘Results’ section). We have thus optimized purification conditions to obtain highly active enzymes. Furthermore, we have observed that in certain reaction conditions (e.g. reaction buffer or high enzyme concentration), PvuRts1I starts to digest 5mC and 5hmC indiscriminately (Figure 3 in Results). This raises the concern of a possibly elevated false discovery rate if it is used improperly, which must be carefully monitored during its application.

In this article, we systematically characterized the enzymatic properties of several members in the PvuRts1I family. In particular, we focus on comparing their substrate selectivity on different forms of cytosine modifications and evaluating their suitability in mapping genomic 5hmC sites. As one of the conclusions, we show that by using AbaSDFI, a homolog of PvuRts1I with much higher substrate selectivity, it is possible to map genomic 5hmC sites close to base resolution.

METHODS AND MATERIALS

Cloning, expression and purification

Genes in the PvuRts1I family, including PvuRts1I, PpeHI and AbaSDFI (Supplementary Table S1), were synthesized using the optimized *Escherichia coli* codon set from Integrated DNA Technologies Inc. They were then sub-cloned into pTXB1, and overexpressed in *E. coli* strain T7 Express (NEB #C2566). Cells were grown at 30°C in LB medium with ampicillin to late log phase and induced by IPTG at 16°C overnight. Cells were harvested by centrifugation and re-suspended in 0.5 M KOAc, 10 mM Tris–OAc (pH 8.0) (column buffer). After sonication and centrifugation, the clear supernatant was loaded onto a chitin column (NEB #S6651), which was equilibrated with the column buffer containing 0.1% Triton-X100. The column was washed with 50 column volumes of the column buffer. For intein cleavage, the column was flushed with the column buffer containing 30 mM DTT and incubated at 4°C overnight. Fractions containing the purified protein were eluted from the column using the column buffer.

Activities of enzymes were assayed on either T4 gt DNA or T4 wt DNA, depending on the preference of each enzyme. One unit of the enzyme is defined as the amount to digest 1 µg of substrate DNA (T4 gt or T4 wt) to completion in NEB buffer 4 at 23°C within 20 min.

Salt sensitivity of PvuRts1I enzymes in crude lysate

To test the enzyme sensitivity to different salts in crude lysates, 1.5 ml PvuRts1I-expressing *E. coli* cells from overnight culture were spun down and supernatant was removed, then 1.5-µl 1-M Tris–acetate (pH 8.0) and 150 µl of a 1-M solution of each different salt, all buffered to pH 8.0 by its own ion type, were added. Cells were then sonicated, spun again and left at 23°C for 6 h. The supernatant was then diluted in 10-, 100- or 1000-fold by diluent (250 mM KOAc, 10 mM Tris–acetate, pH 8.0 and 200 µg/ml BSA). Of diluted supernatant, 3 µl was tested for activity by incubating with 125 ng T4 gt DNA at 23°C for 20 min in NEB buffer 4. The reactions were stopped by adding 6× loading dye and visualized on a 1% agarose gel (Figure 3).

Preparation of DNA substrates

To prepare the DNA substrates used in Figures 2C and Figure 3, DNA fragments were PCR-amplified from the T4 gt genomic or pUC19 DNA by using dATP/dGTP/dTTP mixed with dhmCTP (Bioline #BIO-39046), dmCTP (NEB #N0356S) or dCTP, respectively. PCRs were carried out using Phusion polymerase (NEB #M0530). The DNA fragment containing 5ghmC was obtained by further modification of 5hmC DNA fragment by the T4 β-glucosyltransferase (NEB #M0357). All PCR primers are listed in the Supplementary Table S2.

The synthetic oligonucleotides used in Figure 4 were made from PCR by using primers hmCG_ACGT_F and hmCG_ACGT_R on the hmCG_ACGT_template in the presence of dhmCTP /dATP/dGTP/dTTP (Supplementary Table S3). The oligonucleotide sequence is designed so that there is only one CG site (underlined in Supplementary Table S3), which contains the 5hmC in the top strand. Before PCR, each primer is individually labeled by using γ-^{33}P-ATP (Perkin-Elmer) and T4 polynucleotide kinase (NEB #M0201), followed by purification using G-25 columns (GE Healthcare) to make the 5′-end labeled substrates. PCR products were purified from the QIAGEN Nucleotide Removal kit. To make the 3′-end labeled substrate, purified unlabeled PCR product was incubated with Taq polymerase (NEB #M0273) and α-^{33}P-dATP (Perkin-Elmer). This way both of its 3′-ends were labeled. As the final step, all labeled DNA fragments were further modified by T4 β-glucosyltransferase.

The synthetic oligonucleotides containing 5hmC used in Figure 5 were synthesized in-house (Supplementary Table S4). Each oligo was resuspended in H_2O to 20 µM. Equal volumes of top strand and bottom strand were then mixed. The final concentration of double-stranded substrate is at 10 µM. To be used as AbaSDFI substrate, each double-stranded oligo was glucosylated using T4 β-glucosyltransferase. In Table 4, 5hmC_21C_top pairs with 5hmC_215hmC_bottom as substrate used in Figure 5B. Similarly, 5hmC_21C_top pairs with 5hmC_21mC_bottom (Figure 5C); 5hmC_21C_top pairs with 5hmC_21C_bottom (Figure 5D); 5hmC_nonC_top pairs with 5hmC_nonC_bottom (Figure 5E); C_21C_top pairs with 5hmC_21C_bottom (Figure 5F).

Relative selectivity of the PvuRts1I enzymes

In each digestion series, 125 ng substrate DNA was digested by PvuRts1I, PpeHI or AbaSDFI in a 2-fold serial dilution in NEB buffer 4 with additional KOAc (final concentration 250 mM) for 20 min at 23°C. Addition of KOAc was found to significantly inhibit the enzyme activity on 5hmC, 5mC and C, with less effect on 5ghmC. The ratio of the relative selectivity is determined by the comparison of the extent of digestion on different substrates.

Sequence dependence on T4 genomic DNA

Of T4 gt DNA, 0.9 µg was digested by PvuRts1I and purified using spin columns. Digested DNA was then treated with T4 DNA polymerase (NEB #M0203) for end-polishing. The DNA fragments were ligated to dephosphorylated linear pUC19 (linearized by HincII). Colonies were picked after transformation to NEB Turbo competent cells (C2984) and the inserts were sequenced.

Genomic mapping of 5hmC sites

A total of 2 µg rat brain genomic DNA from mixed tissue was glucosylated using the T4 β-glucosyltransferase at 37°C overnight. After heat inactivation at 65°C for 20 min, the DNA was then precipitated using isopropanol and re-suspended in 20 µl water. The digestion was completed in a total volume of 30 µl, with 3 µl NEB buffer 4, 6 µl KOAc (2 M, pH = 8), 20 µl glucosylated gDNA and 100 U of AbaSDFI, at room temperature for 1 h. The DNA was then precipitated using isopropanol and re-suspended in 20 µl water. Ligation was performed in a total volume of 10.5 µl, with 1 µl ligation buffer, 8 µl DNA, 0.5 µl of double-stranded adaptor (P1b_top_2N+P1b_bottom for the 2N library or, P1b_top_3N+P1b_bottom for the 3N library, both at 10 µM, see Supplementary Table S4 for sequences), and 1 µl T4 ligase (NEB #M0202), at room temperature for overnight. The ligated DNA was then resolved on a 1% low-melting agarose gel (Lonza #50080) with a DNA size marker. The gel piece containing DNA fragments within the 1–3 kb size range was excised and digested using β-agarase (NEB #M0392). Adaptor-specific PCR was prepared by using primer P1XbaIcloningprimer (Supplementary Table S4) and Phusion DNA polymerase. After PCR, the DNA fragments were cloned into the PmeI site in pNEB193 and individually sequenced in 96-well format (2 plates for the 2N library and 3 plates for the 3N library). The cloned genomic fragments were identified by trimming the adaptor sequence (but leaving the randomized 2 N or 3N nt) and aligned to the rat reference genome (REFSEQ ID: NC_005109.2) using BLASTN (15). The ends of each cloned genomic fragment signify half of the enzymatic cleavage sites. The other half of each cleavage site was inferred by extracting the adjacent 30-nt sequences from the reference genome and joined to the cloned sequence for analysis (Figure 6).

RESULTS

Using PvuRts1I protein sequence as the query, we searched the NR and ENV_NR databases at NCBI using BLAST (15) and identified a number of homologs; collectively, we call them the PvuRts1I enzyme family. Using both *in vivo* phage restriction assays against T4 phages and *in vitro* digestion assays on modified DNA, we evaluated the activity of each homolog and summarized our results in Supplementary Table S1. In the following, we focus our discussion on three representative members in the family: PvuRts1I from *Proteus vulgaris* Rts1, PpeHI from *Proteus penneri* ATCC 35 198 and AbaSDFI from *Acinetobacter baumannii SDF*. All enzyme entries can be found in REBASE (4).

Conserved sequence features define the PvuRts1I family

For a long time, PvuRts1I was placed into the ‘weirdo’ class of restriction endonucleases in REBASE (4), mainly due to its unique biological properties and lack of detailed experimental characterization. With our recent screening efforts, we have identified a number of active PvuRts1I homologs from complete bacterial genome sequences as well as environmental sequences (Supplementary Table S1 and Figure 1). These genes are significantly similar to each other at the sequence level, yet no previously known conserved domains can be identified in the family. Examination of the multiple sequence alignment (Supplementary Figure S2) of the PvuRts1I family protein sequences does not reveal the hallmarks of the usual catalytic motifs that are often observable in the restriction endonucleases, such as PD…(D/E)XK or HNH motifs, etc. (16). Figure 1A shows a schematic sequence conservation profile at the amino acid level abstracted from the multiple sequence alignment (Supplementary

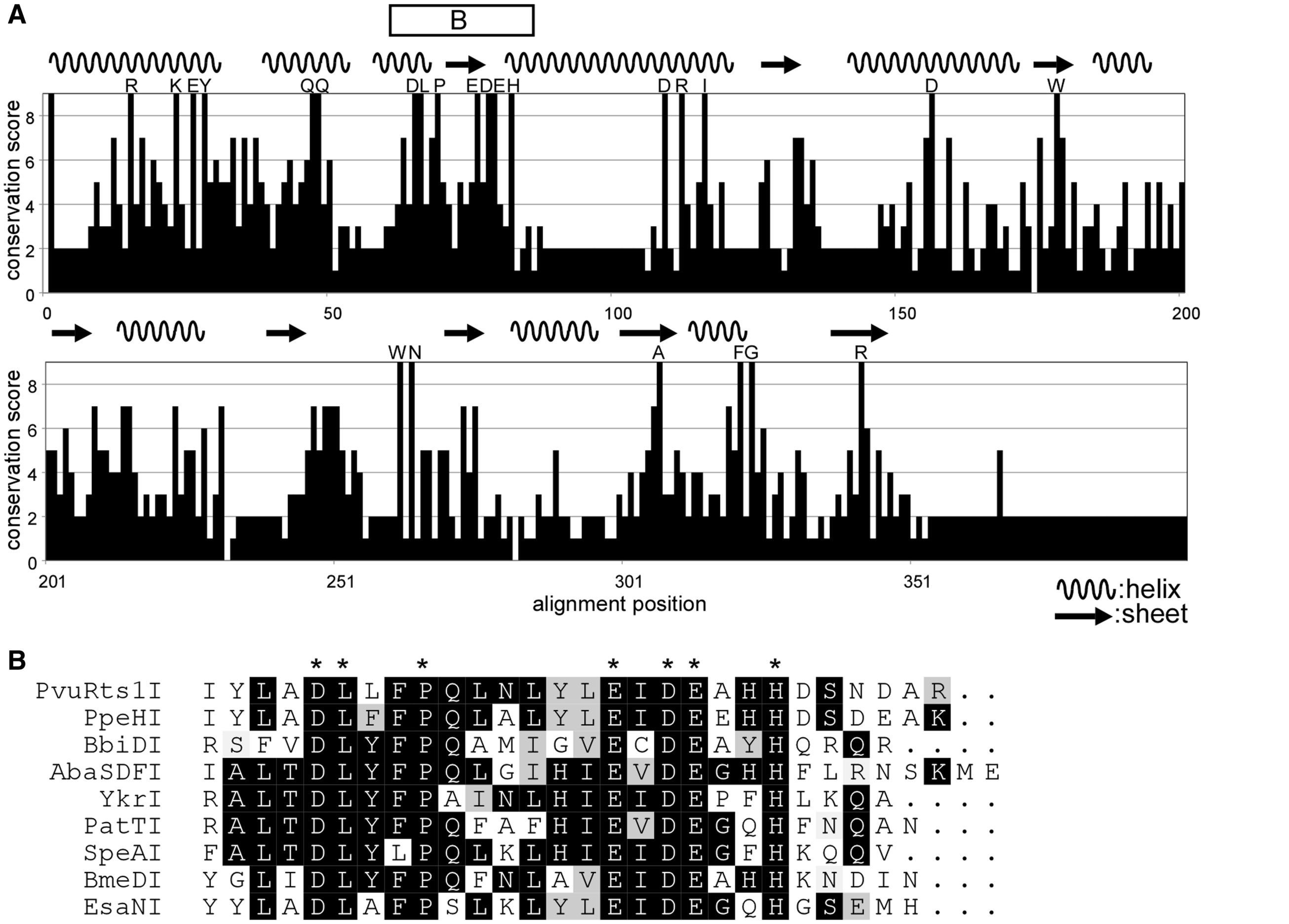

Figure 1. Sequence features of the PvuRts1I family. (**A**) Sequence conservation profiles for the PvuRts1I family. Multiple Sequence Alignment and secondary structure predictions were generated by PROMALS (18). Conservation scores (in 0–9 scale) were calculated at each aligned position by AL2CO (17) and plotted. Absolutely conserved amino acids are shown on top of the conservation profile. A detailed multiple sequence alignment in box B is shown in (B). (**B**) A putative motif with conserved amino acids for binding metal ion and catalysis. The absolutely conserved amino acids are indicated by stars.

Figure S2) (17). The scale of the conservation is from 0 to 9, with 9 being most conserved. The absolutely conserved residues and the predicted secondary structure elements are shown on the top of the profile in Figure 1A (18). It appears that the N-terminal region of the PvuRts1I family is more evolutionarily constrained, with more conserved residues and more well-defined structural elements than the C-terminal region. In the absence of previously known catalytic motifs, we attempted to identify potential catalytically important residues based on conservation and observed enzymatic properties. The activities of the PvuRts1I enzymes are dependent on Mg^{2+} in the reaction buffer, suggesting the possible involvement of metal-ion chelating residues. Figure 1B shows a multiple sequence alignment encompassing a conserved cluster of negatively charged residues in the N-terminal region (box B in Figure 1A). It is likely that this region may be responsible for metal ion binding and can act as the catalytic center. Systematic mutagenesis experiments and structure determination are needed in the future to test the above speculations.

Enzyme purification and the *in vitro* modification-dependent activity

All genes were synthesized using optimized *E. coli* codons. We first fused a few genes with a 6×His-tag, either at the N- or C-terminus, to facilitate quick purification. To our surprise, although a high level of cytosine modification-dependent activity was detected in the crude lysate of the expression clones, a large portion of the activity was quickly lost after purification, even though the target protein was successfully recovered and purified. It appeared to us that the loss of activity may be due to the specific chemicals used during purification. We then investigated the sensitivity of PvuRts1I enzymes in crude lysates to different salt concentrations, as shown in Figure 2A for PvuRts1I. Indeed, a high concentration of imidazole salts, as routinely used for eluting the His-tagged protein from chelating columns, leads to the loss of the majority of the PvuRts1I activity in crude lysates (Figure 2A, lanes d and e). High concentrations of another anion, Cl^{-}, which is commonly used to increase the ionic strength of the buffer, also seem to

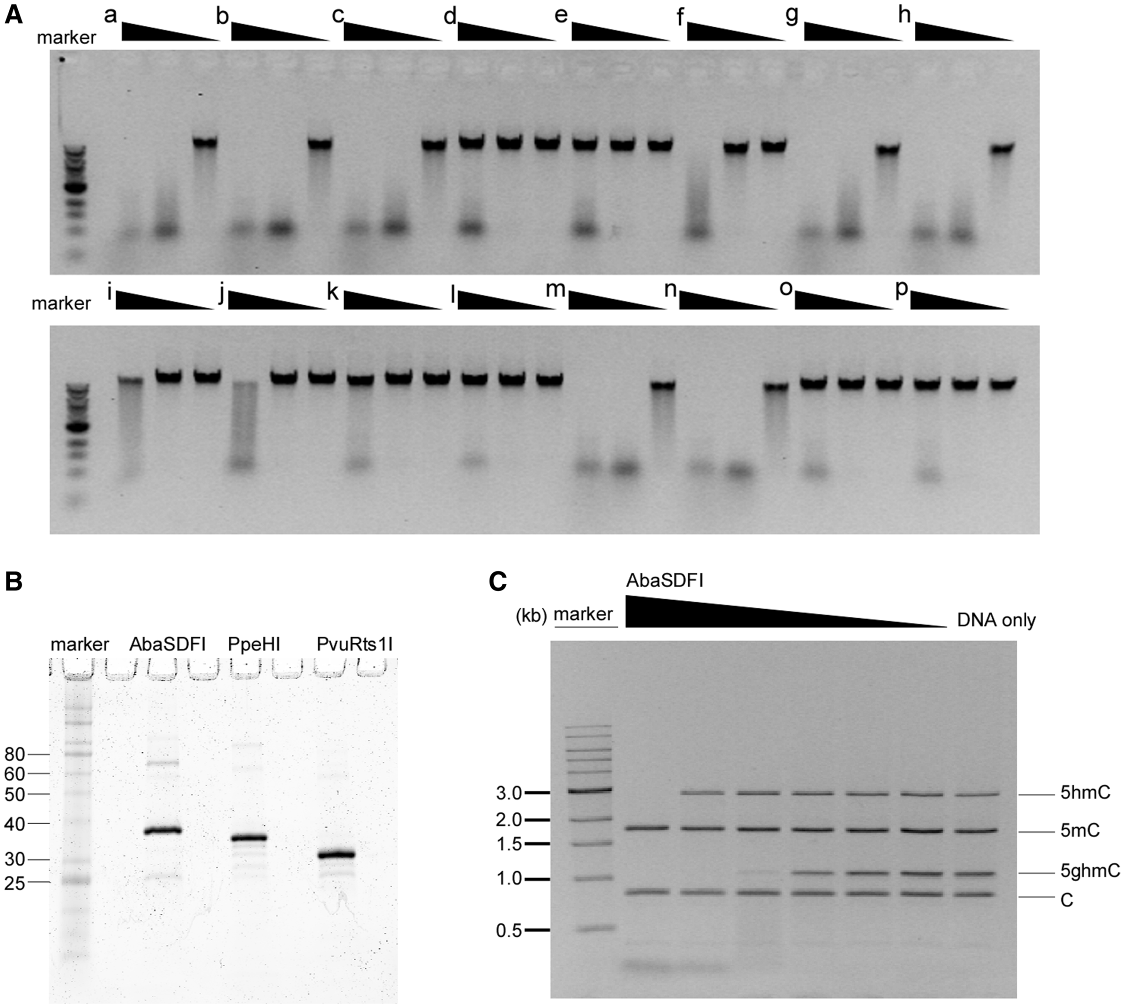

Figure 2. Purification of PvuRts1I family enzymes and *in vitro* modification-dependent activity of AbaSDFI. (**A**) Sensitivity of PvuRts1I to different salts in the crude lysate. a, potassium acetate; b, sodium phosphate; c, potassium phosphate; d, imidazole chloride; e, imidazole citrate; f, sodium citrate; g, ammonium citrate; h, ammonium sulfate; i, sodium chloride; j, potassium chloride; k, cesium chloride; l, calcium chloride; m, sodium sulfate; n, potassium sulfate; o, sodium carbonate; p, sodium nitrate. (**B**) SDS–PAGE of purified AbaSDFI, PpeHI and PvuRts1I. Of each protein, 1 µg was loaded onto the gel. (**C**) *In vitro* modification-dependent activity of AbaSDFI. 4 PCR products with different forms of cytosine were used as substrate for a 5-fold titration of AbaSDFI digestion.

inhibit activity. Most of these activity losses appear irreversible, since dilution or buffer change cannot restore the lost enzymatic activities. Since the presence of high concentrations of NaCl or KCl can adversely affect PvuRts1I enzyme in crude lysates, many common salt gradient elution purification schemes cannot be used as the first step. In Supplementary Figure S1, we show the salt sensitivity of the two other enzymes, PpeHI and AbaSDFI. It appears that the sensitivity profile of each enzyme to a specific salt also varies.

To find a mild and universal purification method, we expressed the recombinant protein fused with a cleavable intein and a chitin-binding domain (CBD) (19). First, the fusion protein was bound to the chitin column under mild conditions; then, the CBD tag of the fusion protein was cleaved off by the embedded intein in the presence of dithiothreitol (DTT) (19). Using this strategy, we obtained each wild-type enzyme in highly active form and close to homogeneity on an SDS-PAGE gel (Figure 2B). The activity of each enzyme was assayed on wild-type T4 (containing 5ghmC, referred to as T4 wt hereafter) or a mutant phage T4 gt (containing 5hmC, referred to as T4 gt hereafter) genomic DNA (see 'Materials and Methods' section). Table 1 lists the basic properties of each purified enzyme. These preparations were then used in the following characterization experiments.

Figure 2C demonstrates the modification-dependent activity of AbaSDFI on a set of DNA fragments designed to test modification selectivity when a choice is offered. Each differently sized fragment carries one form of cytosine at all C locations—5ghmC, 5hmC, 5mC or unmodified C. Under such competitive digestion conditions, it can be seen that AbaSDFI digests 5ghmC- and 5hmC-containing DNA, but prefers the former, and does not act on either 5mC- or C-containing DNA (Figure 2C).

Table 1. Purified members in the PvuRts1I family

Enzyme name	Final concentration (mg/ml)	Unit concentration (substrate) (U/ml)	Specific activity (U/mg)
PvuRts1I	11	5 300 000 (T4gt)	480 000
PpeHI	41	2 600 000 (T4gt)	65 000
AbaSDFI	17	1 300 000 (T4wt)	76 000

Relative substrate selectivity on different forms of cytosine modification

It is known that wild-type restriction endonucleases sometimes exhibit activity on non-canonical sites under certain *in vitro* conditions, e.g. high enzyme concentrations or extended incubation times, etc. (20). These so-called 'star' activities usually do not impair the fitness of the bacterial hosts from which these enzymes originate and are thus not selected against by nature, as the *in vivo* concentration of the enzymes is relatively low. Similarly, restriction endonucleases known to recognize one particular modification may exhibit activity on other modifications, as long as these modifications are not present in the hosts or are not deleterious. A good example is ScoMcrA, which recognizes both DNA phosphorothioation and methylation (21). To use PvuRts1I-like enzymes to map 5hmC sites in the mammalian genome, it is important to know their relative selectivity to different modified cytosines, as C, 5mC and 5hmC all exist in the genome and 5hmC constitutes only a tiny fraction of the cytosine pool (14). Enzymes with low relative substrate selectivity can result in a high false discovery rate.

To quantify the relative selectivity of the PvuRts1I enzymes on different cytosine modifications, we adopted an approach similar to that previously used for regular restriction endonucleases (20). For example, as shown in Figure 3A, with an increasing 2-fold titration of purified PvuRts1I, the enzyme shows a different activity profile on each substrate DNA. PvuRts1I acts on 5hmC and 5ghmC DNA almost equally. When the enzyme concentration is relatively high, PvuRts1I starts to digest DNA containing only 5mC and C as well. From a practical standpoint of mapping 5hmC sites, this is undesired. We define quantitatively the relative selectivity of each enzyme as the ratio of specific activity on different forms of modified cytosines. For example, the relative selectivity for PvuRts1I is 5hmC:5ghmC:5mC:C = 2000:2000:8:1 (Figure 3A). Similarly, the relative selectivity for PpeHI is 5hmC:5ghmC:5mC:C = 128:256:2:1 (Figure 3B); the relative selectivity for AbaSDFI is 5hmC:5ghmC:5mC:C = 500:8000:1:ND (ND: none detected) (Figure 3C). Figure 3D shows the comparison of the three enzymes' relative selectivity normalized based on the activity towards 5mC. From the comparison, we conclude that among the active PvuRts1I-like enzymes we characterized, AbaSDFI has the best discriminative power on 5ghmC over 5mC and C. In addition, only AbaSDFI does not have detectable activity towards unmodified cytosine (Figure 3D). These properties were used in our 5hmC site mapping experiment (see below). Here, we consider 5ghmC equally important as 5hmC because although 5ghmC is not known to be present in the mammalian genome, *in vitro* 5hmC can be converted to 5ghmC essentially completely using the T4 β-glucosyltransferase (22).

Cleavage properties of PvuRts1I enzymes

To investigate the cleavage positions of PvuRts1I enzymes near the modified sites, we individually labeled oligonucleotide substrates at either the 5′- or the 3′-ends (Figure 4A). In the example shown in Figure 4B, the enzyme used was AbaSDFI and the recognition site is a hemi-5ghmC site in the top strand. In Figure 4B, left panel, the top-strand-labeled substrate (lane 2) and the bottom-strand-labeled substrate (lane 1) were separately digested by AbaSDFI. The digested products were resolved on a denaturing polyacrylamide gel to single base resolution for small fragments and compared with synthetic size markers for the bottom strand cleavage site (Figure 4B, left panel). It can be seen that AbaSDFI cleaves both the top strand and the bottom strand on the 3′-side of the recognition site, producing a large-labeled fragment from the top-strand-labeled substrate (lane 2) and a short-labeled fragment from the bottom-strand-labeled substrate (lane 1). The bottom strand cleavage products are of a size that allows discrimination at single-base resolution. The distance from the bottom strand cleavage site to the modified cytosine is predominately 10 nt for this particular substrate, with minor cleavage plus or minus 1 nt.

To precisely map the top strand cleavage site, α-^{33}P-dATP was incorporated into the 3′-ends of both strands by the non-templated polymerization activity of Taq polymerase. The AbaSDFI-digested products were resolved by PAGE (Figure 4B, right panel) and compared with synthetic size markers (Figure 4A). From lane 3 in Figure 4B right panel, the top-strand cleavage site can be deduced to be 12 or 13 nt away from the 3′-side of the modified cytosine for this particular substrate.

Overall, our results suggest that AbaSDFI generates a double-stranded cleavage on the 3′-side and away from the modified cytosine. The substrate tested in Figure 4 is hemi-modified. We have additionally tested fully modified sites and observed activity. For a fully modified site, AbaSDFI cleaves on both sides of the site, essentially carving a small fragment from the DNA, like enzymes in the MspJI family (8). The difference is that the cleavage distance from the recognition site for MspJI (N_{12}/N_{16}) is longer than that of PvuRts1I. Thus, whereas MspJI can produce 32-mer fragments from the fully modified sites, the length of such small fragments for PvuRts1I is ~24 bp, which may provide only limited

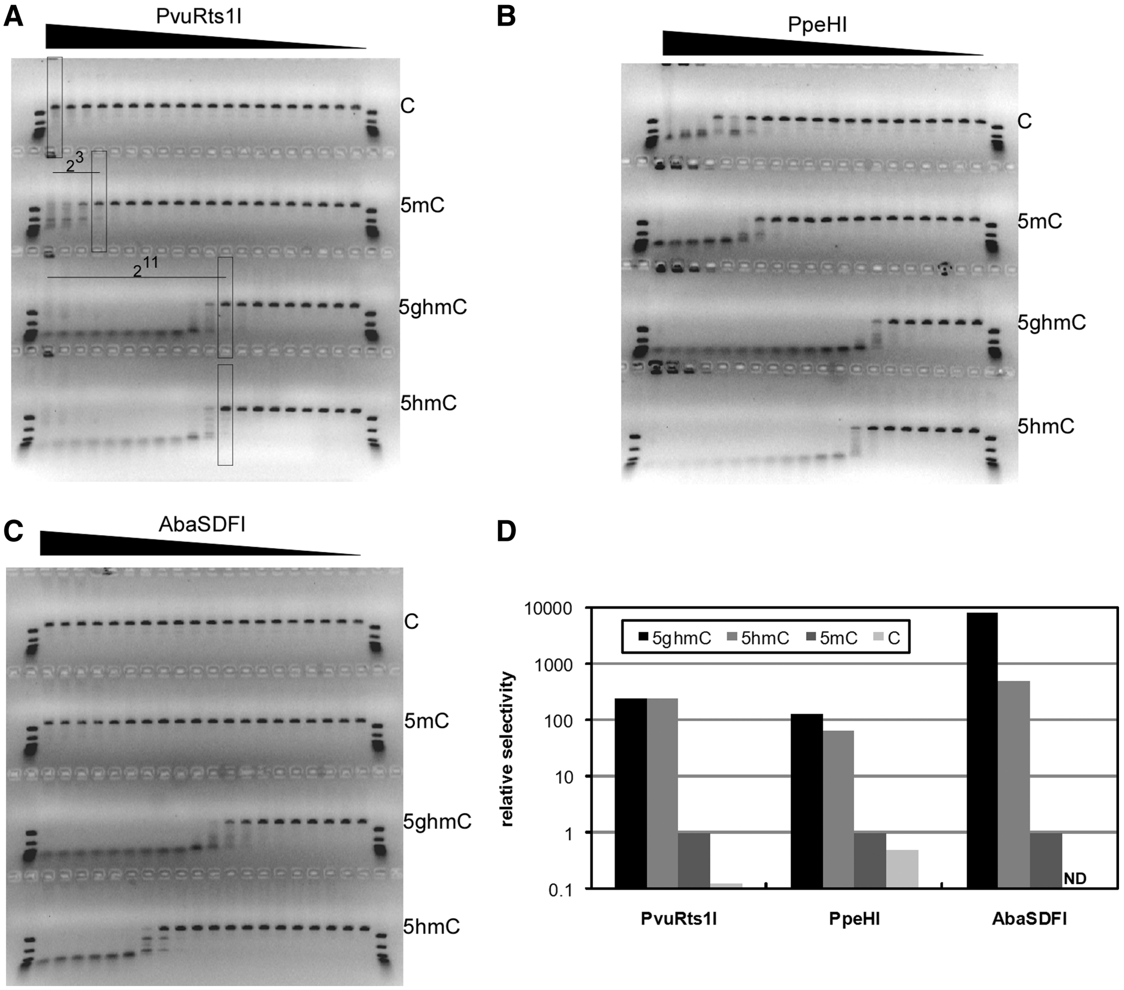

Figure 3. Relative selectivity of PvuRts1I, PpeHI and AbaSDFI on unmodified cytosine (**C**), 5mC, 5hmC and 5ghmC (see 'Materials and Methods' section for description of methods). In each gel, the amount of enzyme is titrated from left (high) to right (low). All DNA substrates were made by PCR. (**A**) PvuRts1I, the approximate relative selectivity is 5hmC:5ghmC:5mC:C = 2000:2000:8:1; (**B**) PpeHI, the approximate relative selectivity is 5hmC:5ghmC:5mC:C = 128:256:2:1; (**C**) AbaSDFI, the approximate relative selectivity is 5hmC:5ghmC:5mC:C = 500:8000:1:ND (none detected). (**D**) Comparison of the relative selectivity on 5ghmC, 5hmC, 5mC and C among PvuRts1I, PpeHI and AbaSDFI. The relative selectivity is plotted in log scale and normalized based on the 5mC activity.

resolution power in the human genome if sequenced. Since the amount of 5hmC is usually very low, we have not visually observed the appearance of the 24-mer band from the digestion of different genomic DNAs using PvuRts1I-family enzymes. Another intriguing observation according to our experiment is that AbaSDFI generates a mixture of fragments with either 2- or 3-base 3′-overhang, which provides the basis for our later genomic mapping experiment (see below).

Sequence dependency of the PvuRts1I enzymes

To determine whether PvuRts1I enzymes require other sequence elements in addition to modified cytosines, we initially cloned and sequenced the digested T4 wt or T4 gt genomic DNA fragments. The T4 wt or gt genomic DNA provides a complex, fully modified substrate, thus allowing identification of preference for sequence context. Briefly, T4 gt DNA was digested by PvuRts1I to completion. The digested DNA fragments were blunt-ended by DNA polymerases and cloned into pUC19 for sequencing (see 'Materials and Methods' section for details). After mapping the inserts to the T4 genome, the sequences encompassing the ends of the inserts, which signify the cleavage sites of PvuRts1I, are subject to further analysis for compositional bias. The identified consensus recognition sites are shown in Supplementary Figure S3. Two 5hmCs on opposite strands are significantly enriched on either side of the cleavage site, with distance to the cleavage site either 12 nt (top strand cut) or 9–10 nt (bottom strand cut) (Supplementary Figure S3). The distances between the two 5hmCs are either 21 nt (47% of all cases) or 22 nt (45% of all cases), which reflects the variable cleavage positions of the enzyme. Importantly, there is little compositional bias in the adjacent positions of the two 5hmC (Supplementary Figure S3), suggesting the possibility that

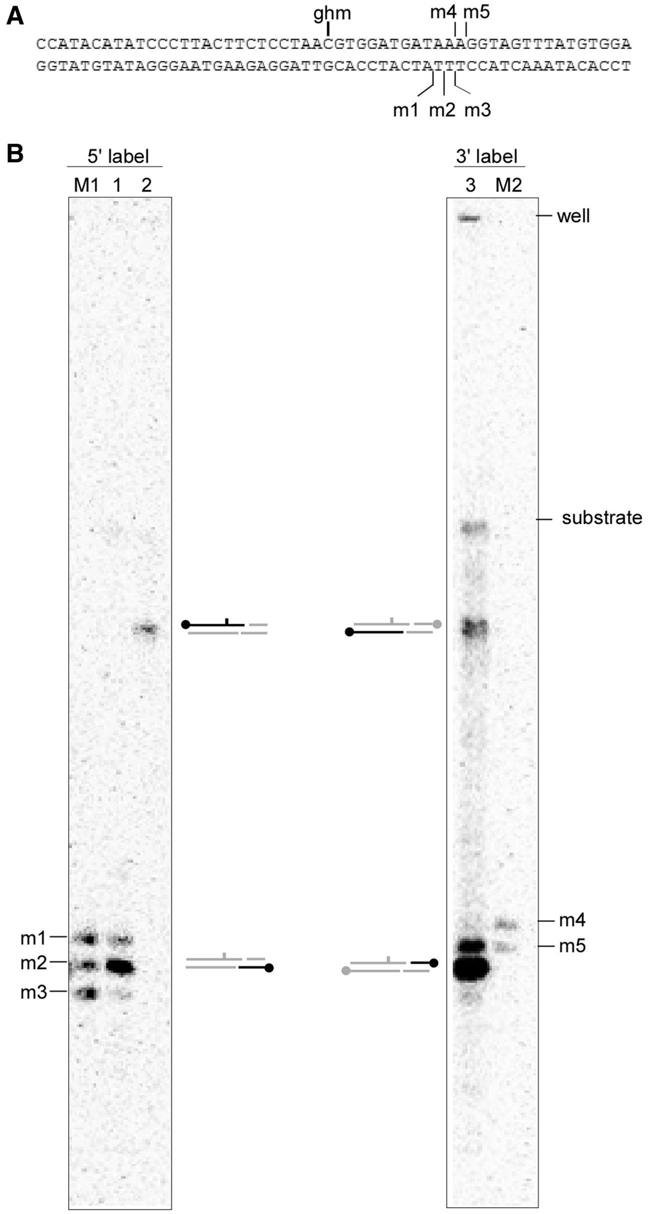
A
ghm
m4 m5
CCATACATATCCCTTACTTCTCCTAACGTGGATGATAAAGGTAGTTTATGTGGA
GGTATGTATAGGGAATGAAGAGGATTGCACCTACTATTTCCATCAAATACACCT
m1 m2 m3
B
5' label
M1 1 2
3' label
3 M2
well
substrate
m4
m5
m1
m2
m3

Figure 4. Cleavage position of AbaSDFI near a hemi-5ghmC site. (**A**) The structure of the oligonucleotide used. The modified cytosine and the positions of the synthetic markers are indicated. (**B**) Digested labeled oligonucleotides resolved in a 20% polyacrylamide 7 M urea denaturing gel. Left panel: lane 1, bottom strand 5′-labeled; lane 2, top strand 5′-labeled; M1, synthetic markers indicated on the bottom strand in (A). Right panel: lane 3, 3′-labeled on both strands; M2, synthetic markers indicated on the top strand in (A). Schematic drawings of the digested products are shown on the side of the gels corresponding to digested bands.

PvuRts1I primarily recognizes the two 5hmCs. Our findings in this experiment are consistent with the study published from Szwagierczak *et al.* (13).

The symmetrical configuration of the recognition sites suggests a possible cleavage process in which two individual monomers bind, one to each modified site, then interact through dimerization leading to double-stranded cleavage. On the other hand, it is important to realize that all cytosines in the T4 wt or T4 gt DNA are in the form of 5ghmC or 5hmC. To provide further experimental support for the dimerization hypothesis, we tested whether there is a dependence on the modification status of the suitably placed cytosines on opposite strands. Figure 5 compares the activity of AbaSDFI on synthetic oligonucleotides containing designed sites with one constant 5ghmC and another base, either as 5ghmC (Figure 5B), 5mC (Figure 5C), unmodified C (Figure 5D) or no cytosine properly placed in the opposite strand (Figure 5E). As a control, Figure 5F shows that AbaSDFI does not act on non-modified DNA substrate. By comparing the cleavage efficiency in Figure 5, it can be concluded that the cleavage efficiency decreases ~25-fold when one of the two 5ghmCs in the recognition site changes to 5mC or unmodified C. However, the cleavage efficiency drops dramatically when there is no cytosine within the suitable distance range in the opposite strand (Figure 5A and E). Supplementary Figure S4 shows the activity of PvuRts1I on the same set of oligonucleotide substrates (without glucosylation). Similar to AbaSDFI, PvuRts1I prefers sites with two properly placed 5hmC (Supplementary Figure S4B). Sites with one 5mC or one C are digested with a lower efficiency (Supplementary Figure S4CD). The efficiency further drops on sites with only one 5hmC (Supplementary Figure S4E). Consistent with the results in Figure 3, PvuRts1I even digests unmodified DNA substrate in high concentration (Supplementary Figure S4F). These results further support the high substrate selectivity of AbaSDFI.

Overall, it appears that the PvuRts1I-family of enzymes recognizes two cytosines on opposite strands, which are separated by 21 or 22 nt and at least one cytosine needs to be suitably modified as 5hmC or 5ghmC.

Mapping 5hmC sites in mammalian genomic DNA using AbaSDFI

The property of introducing a double-stranded cleavage at a narrowly specified distance from 5hmC sites by the PvuRts1I family enzymes suggests a potential application for mapping genomic 5hmC sites. As a proof-of-principle experiment, we chose the enzyme AbaSDFI, which has the highest relative selectivity on 5ghmC versus 5mC or C. Briefly, we first glucosylated rat brain genomic DNA, using recombinant T4 β-glucosyltransferase (see 'Materials and Methods' section for details). AbaSDFI was then used to digest the glucosylated genomic DNA. The digested DNA was then ligated with a double-stranded adaptor with either a 2- or 3-base randomized 3′-overhang, which are referred as the 2 N or the 3 N libraries hereafter. The ligated DNA was size-selected from 1 to 3 kb on an agarose gel and PCR-amplified with an adaptor-specific primer. The amplified DNA was cloned into pUC19 for sequencing the inserts. The sequenced inserts were then aligned to the rat reference genome to identify the cleavage sites at both ends.

Figure 6 summarizes the analysis of the sequence fragments around the 122 identified cleavage sites in the 2 N library. One of the advantages of using the 2N adaptor is

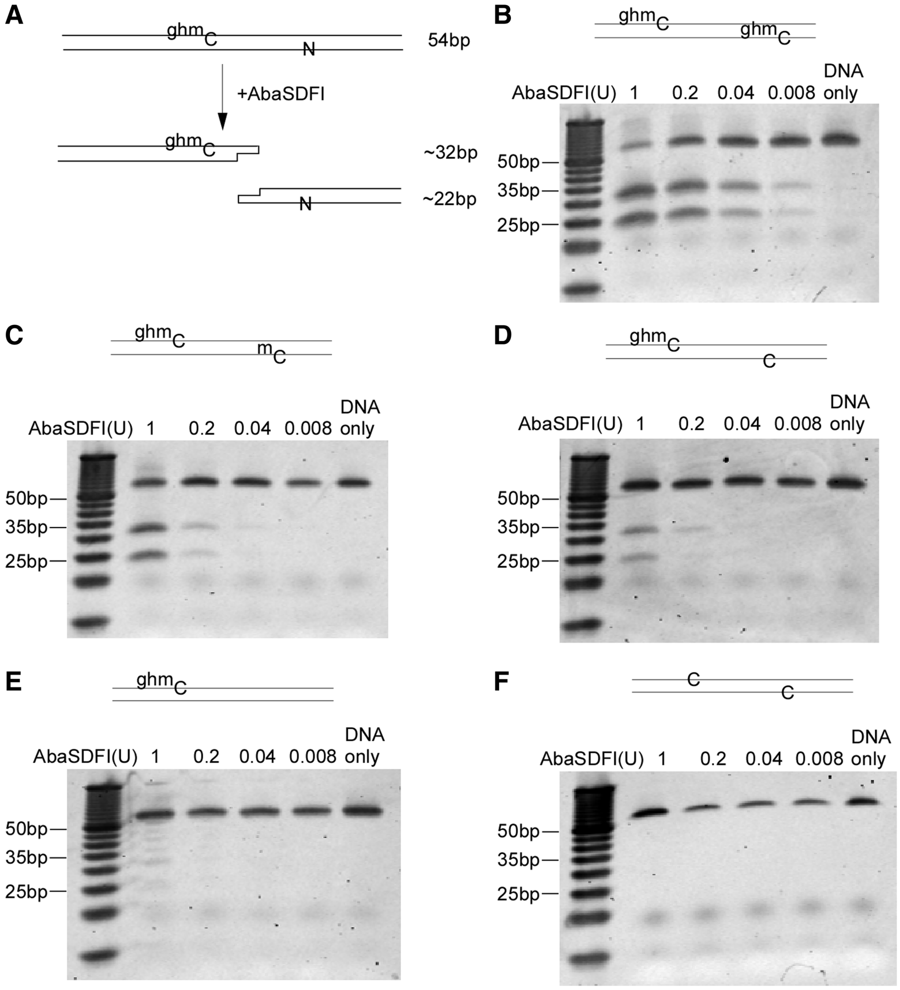

Figure 5. Activity of AbaSDFI on synthetic oligonucleotides with different modified recognition sites. (**A**) Expected digested fragments from AbaSDFI digestion. Sequences of the oligonucleotide can be found in Supplementary Table S4; (**B**) Activity of AbaSDFI on the synthetic oligonucleotide with two 5ghmC, separated by 21 nt; 5 pmol of DNA substrate was digested using a titration of AbaSDFI and resolved on a 20% polyacrylamide PAGE. The gel was stained with SYBR Gold. (**C**) Activity of AbaSDFI on the synthetic oligonucleotide with one 5ghmC and one 5mC; (**D**) activity of AbaSDFI on the synthetic oligonucleotide with one 5ghmC and one C; (**E**) activity of AbaSDFI on the synthetic oligonucleotide with only one 5ghmC and no cytosine in the region 20–25 nt away (Supplementary Table S4); (**F**) activity of AbaSDFI on the synthetic oligonucleotide with two unmodified C [compare with substrate in (D)].

that it preserves the 2-base 3′ extension on the digested DNA fragments to allow precise determination of the cleavage sites in both strands from the sequencing data. By considering the variable cleavage distance of the enzyme, i.e. either 12/10 (denoted as $C_{12/10}$, C is the cytosine being recognized) or 11/9 (denoted as $C_{11/9}$) for the 2N library, the sequences around the cleavage sites can be grouped based on a few different configurations (Figure 6AB), for example, sequences with two symmetrical $C_{12/10}$ cleavages, sequences with two symmetrical $C_{11/9}$ cleavages, or, sequences with 1 $C_{12/10}$ and 1 $C_{11/9}$ cleavages, etc.(Figure 6B). Figure 6B shows the comparison on the frequency of occurrences of these sites between the cloned library and those expected by chance. For example, ~20% of the cleavage sites have two symmetrical $C_{12/10}$ (category 2) and ~20% have two symmetrical $C_{11/9}$ (category 1), which are significantly higher than 3.5% expected by chance (Figure 6B). The same significant overrepresentation is seen for sites with 1 $C_{12/10}$ and 1 $C_{11/9}$ (category 3 in Figure 6B). While these configurations are significantly overrepresented in the cloned library, sites with C in only one side of the cleavage sites (category 5 in Figure 6B), or, sites with no suitable C in the vicinity of the cleavage sites (category 6 in Figure 6B) only constitute 11% and 3.3% of the cloned library respectively, much lower than the 49% and 32% expected by chance. Thus, it appears that sites that are not recognized by AbaSDFI are significantly underrepresented in the cloned library. There are ~20% of the sequences which contain '$C_{12/10}C_{11/9}$' as a half site and a $C_{12/10}$ or $C_{11/9}$ as the other half site (category 4 in Figure 6B). For these, we could not determine which C in the 'CC' is recognized by the enzyme. Nevertheless, they are significantly overrepresented as well and may be 5hmC-containing sites. Overall, a high percentage (86%) of all the cleavage sites appears to be true cleavage sites catalyzed by the enzyme.

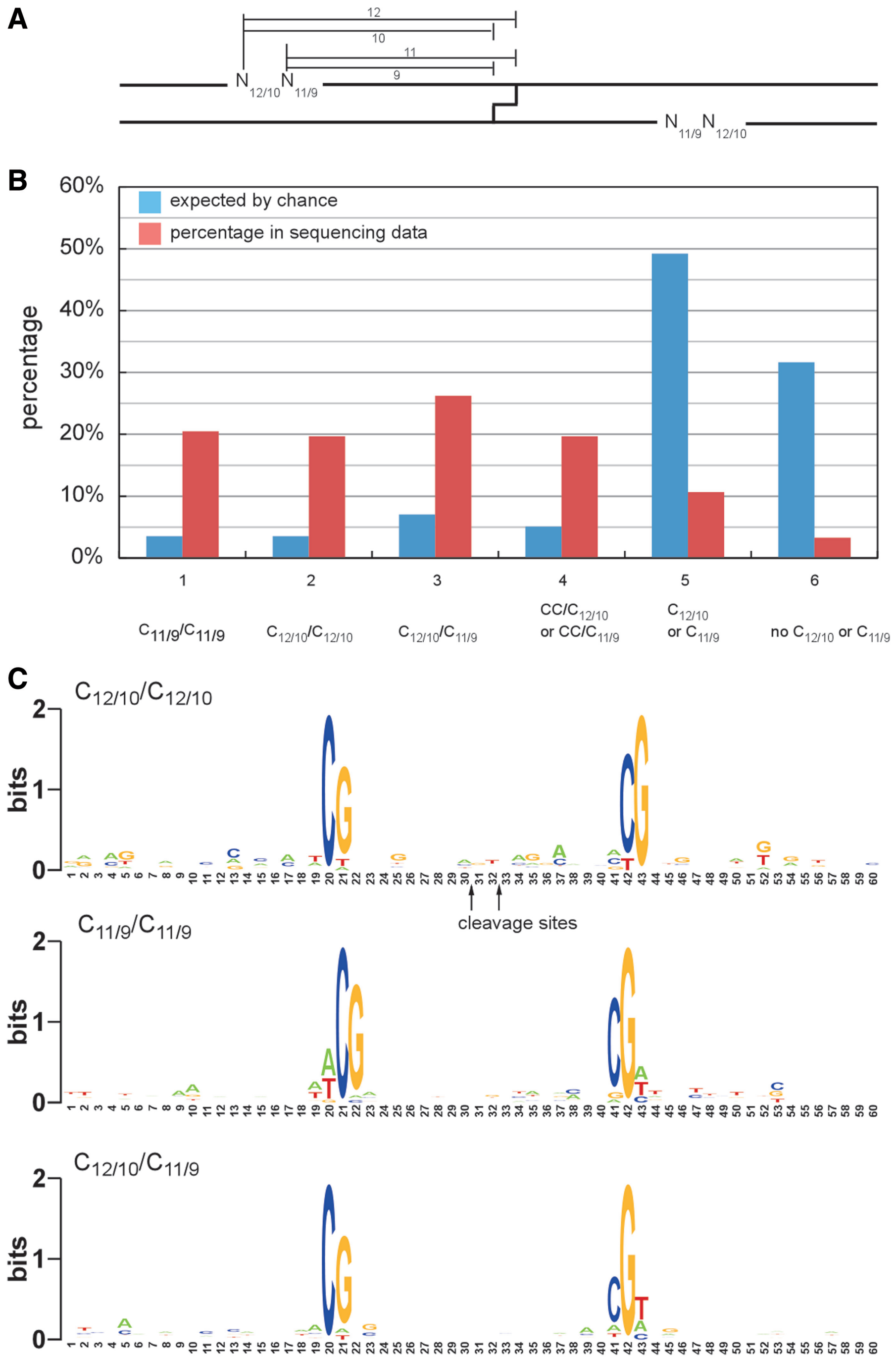

Figure 6. Analysis of the mapped cleavage sites of AbaSDFI in the 2N library from rat brain genomic DNA. (**A**) Schematic drawing of the recognition sites. 'N' indicates the positions analyzed; (**B**) comparison of site distribution in sequenced library and expected by chance. A total of the 122 mapped cleavage sites in the 2N library were analyzed. Each category (1–6) represents a particular site configuration. Blue bars show the percentages expected by chance. Red bars show the percentage in the analyzed data set. 1: symmetrical 11/9 cuts, denoted as $\underline{C_{12/10}}C_{11/9}/C_{11/9}\underline{C_{12/10}}$, which means both left and right $N_{11/9}$ position is C, both left and right $N_{12/10}$ position is NOT C (underlined). The expected percentage is calculated as 0.75*0.25*0.25*0.75 = 3.5%; 2: symmetrical 12/10 cuts, $C_{12/10}\underline{C_{11/9}}/\underline{C_{11/9}}C_{12/10}$; 3: one side 11/9 cut, the other side 12/10 cut, $C_{12/10}\underline{C_{11/9}}/C_{11/9}\underline{C_{12/10}}$ or $\underline{C_{12/10}}C_{11/9}/\underline{C_{11/9}}C_{12/10}$; 4: sites with $C_{12/10}C_{11/9}$ as one side, the other side is either $C_{12/10}$ or $C_{11/9}$; 5: site with $C_{12/10}$ or $C_{11/9}$ as one side, the other side does not have C; 6: both sides do not have C. (**C**) Sequence logos in categories with two symmetrical $C_{12/10}$ cuts, two symmetrical $C_{11/9}$ cuts, and 1 $C_{12/10}$ 1 $C_{11/9}$ cuts. The sequence logos are generated by Weblogo (23).

Figure 6C shows the sequence logo representation of the sites in categories 1 (symmetrical $C_{11/9}$), 2 (symmetrical $C_{12/10}$) and 3 (1 $C_{12/10}$ 1 $C_{11/9}$) in Figure 6B (23). Supplementary Figures S6 and S7 list the aligned sequences in categories 1 and 2. It is important to realize that due to the symmetrical nature of the AbaSDFI recognition sites, the motifs presented in Figure 6C do not distinguish which of the two cytosines around the cleavage sites is the real 5ghmC site. Based on the results in Figure 5, it is possible that both cytosines are 5hmC, or, one is 5hmC and the other is 5mC or unmodified cytosine. Indeed, this may be reflected by the appearance of the enriched CG dinucleotide encompassing the recognized cytosines in Figure 6C. On the one hand, it is expected since the 5hmCs most likely arise from the methylated CpG sites from the action of the TET enzymes in the brain DNA (5); on the other hand, the methylated CG sites may also constitute half of the recognition sites for AbaSDFI. Interestingly, the flanking position on the 5′-side of the recognized C shows an overrepresentation of A or T in the symmetrical $C_{11/9}$ cleavages, whereas it is absent in the symmetrical $C_{12/10}$ cleavages (Figure 6C). This suggests that the cleavage distance may be affected by the nucleotide flanking the recognized cytosine. Further experiments are needed to test this hypothesis. In addition, Supplementary Figure S5 summarizes the analysis of 188 sequenced cleavage sites in the 3N library from which similar observations can be made.

DISCUSSION

In this article, we compared the *in vitro* biochemical properties of a few members in the PvuRts1I family. The first example of this family, PvuRts1I, was known to restrict T-even bacteriophages with 5hmC or 5ghmC in their genomic DNA (3). Using a relatively mild purification procedure, we were able to obtain pure enzymes in highly active form; all exhibit DNA modification-dependent endonuclease activity with similar cleavage properties. In addition, our results suggest that these enzymes differ from each other in their relative selectivity toward various forms of modified cytosine. The relative selectivity provides a quantitative index of their ‘fidelity’ towards each desired forms of modified cytosine. From the application perspective, this is important due to the extremely low abundance of 5hmC compared with the 5mC and C in the genome. As a result, we find that AbaSDFI, a homolog of PvuRts1I, has the highest (8000:1) relative selectivity between 5ghmC and 5mC. Furthermore, it does not have any detectable activity on unmodified C. These properties allow reliable mapping of genomic 5hmC sites in large mammalian genomes. This high selectivity may also reflect the enzyme's inherent ability to distinguish the major structural difference between 5ghmC and 5mC.

The PvuRts1I family differs from many other well-studied restriction endonucleases in several aspects. First, they display no identifiable previously known motifs for metal-ion chelation or catalysis, suggesting a novel enzymatic DNA cleavage chemistry. Multiple sequence alignment reveals that the N-terminal region is more conserved than the C-terminal region both in amino acid sequence and in the secondary structure elements (Figure 1A). There are strings of conserved positions in both the N-terminal and the C-terminal regions (Figure 1A). We surmise that a cluster of conserved acidic residues in the N-terminal region may be responsible for chelating Mg^{2+} and could form part of the active center (Figure 1B). Furthermore, it is tempting to assume the nearby conserved histidines may act as the general base in the cleavage process (16). If true, this implies a domain organization with the N-terminus responsible for cleavage activity and the C-terminus responsible for binding. This is different from the domain organization of other type IIS restriction endonucleases such as FokI (24) or MspJI (8), which have an N-terminal domain for binding and a C-terminal domain for cleavage, but is similar to that of MmeI (25). Site-directed mutagenesis and structure determination will be needed for further elucidation. Second, the requirement of two cytosines within a defined distance range on separate DNA strands suggests a likely dimerization step in the cleavage process. Intriguingly, our results suggest that as long as one binding site contains the recognized modified cytosine, e.g. 5hmC or 5ghmC, the other site can be 5mC, or even unmodified cytosine, with moderate decrease of the cleavage efficiency (Figure 5). It is possible that PvuRts1I-like enzymes recognize not only the 5-modification on the cytosine, but also other structural elements of cytosine; this may explain why two cytosines are required for cleavage. Further biochemical studies are needed to clarify the role of the second cytosine.

Theoretically, the coverage of 5hmC sites using the PvuRts1I-like enzymes could be quite high, due to its flexible requirement for the binding sites. Based on our data, for each 5hmC site, as long as there is another cytosine, modified or not, at its 3′-side in the opposite strand and 20–22 nt away, it should elicit enzymatic cleavage. This translates to a theoretical coverage of ~58% (1−0.75*0.75*0.75), assuming there is no severe bias of base composition in the genome. In our proof-of-principle experiment in Figure 6, all the mapped sites contain at least one 5hmC. Although there is still ambiguity in the results as to which cytosine is the real 5hmC, this provides a much higher resolution than the hMeDIP-like approaches and may offer better insights into the existence of 5hmC in the genome. Future experiments will see the application of these enzymes in the genomic mapping of different cell types using the latest high-throughput sequencing technologies.

The *in vitro* biochemical properties of the PvuRts1I enzymes dictate the experimental approaches to mapping genomic 5hmC sites along with the computational interpretation of the sequencing data. Because these enzymes can generate a mixture of ends from a single recognition site, we used double-stranded DNA adaptors with randomized 2- or 3-base 3′-overhangs to separate the population. This provides the advantage of precisely locating the cleavage sites in both DNA strands in the sequencing data, which in turn reduces the uncertainty in searching for the nearby recognition sites. Given the variable cleavage distance property of the PvuRts1I enzymes, this may be a crucial step in library construction.

In summary, the PvuRts1I family of enzymes defines a unique group of DNA modification-dependent restriction endonucleases. Having them and combined with the high-throughput sequencing platforms, it should be possible to improve the resolution of the current hydroxymethylomes in mammalian cells.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Drs Elisabeth Raleigh, William Jack and Zhiyi Sun for critical reading and suggestions on the article. We thank in-house organic synthesis for synthesizing 5hmC oligos. The rat brain DNA is a kind gift from Dr Alan Herbert (Boston University School of Medicine).

FUNDING

Funding for open access charge: National Institutes of General Medical Sciences, SBIR (grant #4R44GM095209-02 to Y.Z., in part); New England Biolabs Inc.

Conflict of interest statement. None declared.

REFERENCES

1. Warren,R.A. (1980) Modified bases in bacteriophage DNAs. *Annu. Rev. Microbiol.*, **34**, 137–158.
2. Ishaq,M. and Kaji,A. (1980) Mechanism of T4 phage restriction by plasmid Rts 1. Cleavage of T4 phage DNA by Rts 1-specific enzyme. *J. Biol. Chem.*, **255**, 4040–4047.
3. Janosi,L., Yonemitsu,H., Hong,H. and Kaji,A. (1994) Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRts1I) modulated by glucosylation of DNA. *J. Mol. Biol.*, **242**, 45–61.
4. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
5. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
6. Kriaucionis,S. and Heintz,N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
7. Wyatt,G.R. and Cohen,S.S. (1953) The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochem. J.*, **55**, 774–782.
8. Zheng,Y., Cohen-Karni,D., Xu,D., Chin,H.G., Wilson,G., Pradhan,S. and Roberts,R.J. (2010) A unique family of Mrr-like modification-dependent restriction endonucleases. *Nucleic Acids Res.*, **38**, 5527–5534.
9. Raleigh,E.A. (1992) Organization and function of the mcrBC genes of Escherichia coli K-12. *Mol. Microbiol.*, **6**, 1079–1086.
10. Huang,Y., Pastor,W.A., Shen,Y., Tahiliani,M., Liu,D.R. and Rao,A. (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE*, **5**, e8888.
11. Jin,S.G., Kadam,S. and Pfeifer,G.P. (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **38**, e125.
12. Ficz,G., Branco,M.R., Seisenberger,S., Santos,F., Krueger,F., Hore,T.A., Marques,C.J., Andrews,S. and Reik,W. (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, **473**, 398–402.
13. Szwagierczak,A., Brachmann,A., Schmidt,C.S., Bultmann,S., Leonhardt,H. and Spada,F. (2011) Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 5149–5156.
14. Song,C.X., Szulwach,K.E., Fu,Y., Dai,Q., Yi,C., Li,X., Li,Y., Chen,C.H., Zhang,W., Jian,X. *et al.* (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.*, **29**, 68–72.
15. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
16. Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
17. Pei,J. and Grishin,N.V. (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, **17**, 700–712.
18. Pei,J., Kim,B.H., Tang,M. and Grishin,N.V. (2007) PROMALS web server for accurate multiple protein sequence alignments. *Nucleic Acids Res.*, **35**, W649–W652.
19. Chong,S., Mersha,F.B., Comb,D.G., Scott,M.E., Landry,D., Vence,L.M., Perler,F.B., Benner,J., Kucera,R.B., Hirvonen,C.A. *et al.* (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*, **192**, 271–281.
20. Wei,H., Therrien,C., Blanchard,A., Guan,S. and Zhu,Z. (2008) The Fidelity Index provides a systematic quantitation of star activity of DNA restriction endonucleases. *Nucleic Acids Res.*, **36**, e50.
21. Liu,G., Ou,H.Y., Wang,T., Li,L., Tan,H., Zhou,X., Rajakumar,K., Deng,Z. and He,X. (2010) Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS Genet.*, **6**, e1001253.
22. 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.
23. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
24. Li,L., Wu,L.P. and Chandrasegaran,S. (1992) Functional domains in Fok I restriction endonuclease. *Proc. Natl Acad. Sci. USA*, **89**, 4275–4279.
25. Nakonieczna,J., Kaczorowski,T., Obarska-Kosinska,A. and Bujnicki,J.M. (2009) Functional analysis of MmeI from methanol utilizer Methylophilus methylotrophus, a subtype IIC restriction-modification enzyme related to type I enzymes. *Appl. Environ. Microbiol.*, **75**, 212–223.