

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 5- β -glucosylated hmC (5 β ghmC) but also 5- α -glucosylated hmC (5 α ghmC). In some cases, the enzymes are inhibited by 5 β ghmC, whereas in others they are inhibited by 5 α ghmC. 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, oxoBS-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

*To whom correspondence should be addressed. Tel: +1 978 380 7237; Fax: +1 978 380 7518; Email: zhuz@neb.com

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 PvuRtsII 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 PvuRtsII 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 (T4 β), we also observed differential specificities for alpha-glucosylated T4 DNA (T4 α) among the homologues. For example, AbaSI cutting is greatly inhibited by 5- α -glucosylated hmC (5 α ghmC) when compared with 5- β -glucosylated hmC (5 β ghmC), while PvuRtsII cutting is enhanced on 5 α ghmC when compared with 5 β ghmC.

MATERIALS AND METHODS

Cloning, expression and purification of PvuRtsII homologues

C-terminally intein-tagged PvuRtsII 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 PvuRtsII 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 their specific activities, resulting in a range of 1–400 units (Table 1). One unit of enzyme is defined as the amount to digest 1 μ g of substrate (either T4gt or T4 β , depending on the preference of each enzyme) to completion in NEB buffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9) at 25°C, 20 min.

Analytical size exclusion chromatography

Analytical size exclusion chromatography was performed on a superdex 200 10/300GL column (GE # 17-5175-01), pre-equilibrated in 500 mM potassium acetate, 10 mM Tris-acetate, pH 8.0 buffer. The column was calibrated with blue dextran to measure the void volume (V_o),

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 V_e (elution volume)/ V_o . After calibration, the column was re-equilibrated in the same buffer, and the homologues, varying in concentration from 200 μ g to 3 mg (depending on the stock concentrations) were applied to the column. V_e/V_o for each protein was calculated and the molecular weights were determined from the standard curve.

T4 α -glucosyltransferase

The pAIII17- α -glucosyltransferase (AGT) plasmid containing the coding region for AGT was transformed into a *dcm*⁻ *E. coli* strain T7 Express (NEB # C2566). After selection on solid LB media containing ampicillin (100 μ g/ml), individual colonies were used to inoculate 1 L luria broth (LB) media containing ampicillin (100 μ g/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).

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 lastly, 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

Specificities of enzymes were determined on non-methylated lambda DNA (C), XP12 DNA [5mC (25)], phage T4gt DNA (containing non-glucosylated 5hmC), T4 β (containing 5 β ghmC) and T4 α (containing

Table 1. PvuRtsII homologue information

Enzyme	Specific activity units/mg	Length ^a , similarity (%) and identity (%) ^b	Source	Genbank accession
PvuRtsII	65 000	293, 100, 100	<i>Proteus vulgaris (RtsI)</i>	AP004237.1
AbaSI	16 000	317, 47.8, 35.4	<i>Acinetobacter baumannii SDF</i>	CAP02625.1
AbaHI	17 800	317, 48.9, 36.5	<i>Acinetobacter baumannii AB210</i>	EGK46224.1
AbaAI	20 000	317, 48.2, 35.7	<i>Acinetobacter baumannii AB900</i>	ZP_04663630.1
AbaCI	8 000	317, 48.1, 35.7	<i>Acinetobacter baumannii ATCC 19606</i>	ACQB01000053.1
AbaDI	8 400	317, 47.8, 35.4	<i>Acinetobacter baumannii D1279779</i>	AERZ01000045.1
AbaBGI	11 000	317, 48.2, 35.8	<i>Acinetobacter baumannii ATCC 17978</i>	CP000521.1
AbaTI	5 800	317, 47.8, 35.4	<i>Acinetobacter baumannii MDR-TJ</i>	ABTJ_00232
AbaUI	7 600	317, 48.1, 35.4	<i>Acinetobacter baumannii UMB003</i>	AEPM01000006.1
AcaPI	4 400	317, 47.8, 35.0	<i>Acinetobacter calcoaceticus PHEA-2</i>	ADY83307.1
BbiDI	104 000	311, 43.3, 32.0	<i>Bifidobacterium bifidum NCIMB 41171</i>	ABQP01000008
BmeDI	23 000	300, 40.0, 30.5	<i>Bacillus megaterium strain DSM319</i>	ADF40396.1
EsaMMI	2 000	305, 55.1, 43.1	<i>Marine metagenome</i>	EDB78928.1
EsaNI	2 200	299, 47.8, 39.6	<i>Marine metagenome</i>	EBQ05151.1
Mte37I	3 000	265, 46.1, 36.1	<i>Microbacterium testaceum StLB037</i>	BAJ74504.1
PatTI	14 000	306, 51.4, 36.5	<i>Pseudoalteromonas atlantica T6c ATCC BAA-1087</i>	ABG40021.1
PfrCI	364	292, 41.4, 29.3	<i>Propionibacterium freudenreichii subsp. Shermanii CIRM-BIA1</i>	ADY83307.1
PpeHI	13 000	294, 72.5, 64.6	<i>Proteus penneri ATCC 35198</i>	EEG85915.1
PxyI	93	290, 46.6, 36.0	<i>Paraprevotella xylaniphila YIT 11841</i>	EGG52627.1
YkrI	9 000	299, 49.6, 36.1	<i>Yersinia kristensenii ATCC33638</i>	EFP89557.1

^aamino acid.^bGCG best fit result (Version 11.1, Accelrys Inc., San Diego, CA, USA).

5 α ghmC). Non-methylated lambda DNA was purchased from Sigma (# D3654). XP12, T4wt and T4gt genomic DNA were purified from phage cultures. DNA containing either 5 β ghmC or 5 α ghmC was obtained by further modification of T4gt DNA by the T4 β -glucosyltransferase [(BGT), NEB #M0357] and AGT, respectively. The relative specificities of the PvuRtsII 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 CTT ACT TCT CCT AA(5hmC) GTG GAT GAT AAA GGT AGT TTA TGT GGA-3'FAM and 5'-TCC ACA TAA ACT ACC TTT ATC ATC CAC GTT AGG AGA AGT AAG GGA TAT GTA TGG-3'FAM. Double-stranded oligonucleotides (10 μ M final) were obtained by heating solutions with equal concentrations of top- and bottom-strand oligonucleotide to 95°C followed by a gradual cooling to room temperature. The DNA oligonucleotides with both a top- and bottom-strand 5hmC and 5'FAM labels were synthesized with forward (5'-CCA TAC ATA TCC CTT ACT TCT CCT A) and reverse (5'-TCC ACA TAA ACT ACC TTT ATC ATC CAC G-3') polymerase chain reaction primers and using 5'-CCA TAC ATA TCC CTT ACT TCT CCT AAC

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/ μ l 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 μ M 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 a 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 a 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 2. Oligonucleotides for cytosine modification dependence

Oligo	Sequence
A	5'-GGTTGGACTCAAGACGATA(5hmC)TTACCGGATAAGGCGCAATTATATTACTTAACCT-3' 3'-CCAACCTGAGTTCTGCTATGAATGGCCTATTCCGCGTAAATTAATGAATTGGA-5'
C/C	5'-GGTTGGACTCAAGACGATACTTACCGGATAAGGCGCAATTAGATTACTTAACCT-3' 3'-CCAACCTGAGTTCTGCTATGAATGGCCTATTCCGCGTAAATCTAATGAATTGGA-5'
C	5'-GGTTGGACTCAAGACGATA(5hmC)TTACCGGATAAGGCGCAATTAGATTACTTAACCT-3' 3'-CCAACCTGAGTTCTGCTATGAATGGCCTATTCCGCGTAAATCTAATGAATTGGA-5'
5mC	5'-GGTTGGACTCAAGACGATA(5hmC)TTACCGGATAAGGCGCAATTAGATTACTTAACCT-3' 3'-CCAACCTGAGTTCTGCTATGAATGGCCTATTCCGCGTAAAT(5mC)TAATGAATTGGA-5'
5hmC	5'-GGTTGGACTCAAGACGATA(5hmC)TTACCGGATAAGGCGCAATTAGATTACTTAACCT-3' 3'-CCAACCTGAGTTCTGCTATGAATGGCCTATTCCGCGTAAAT(5hmC)TAATGAATTGGA-5'

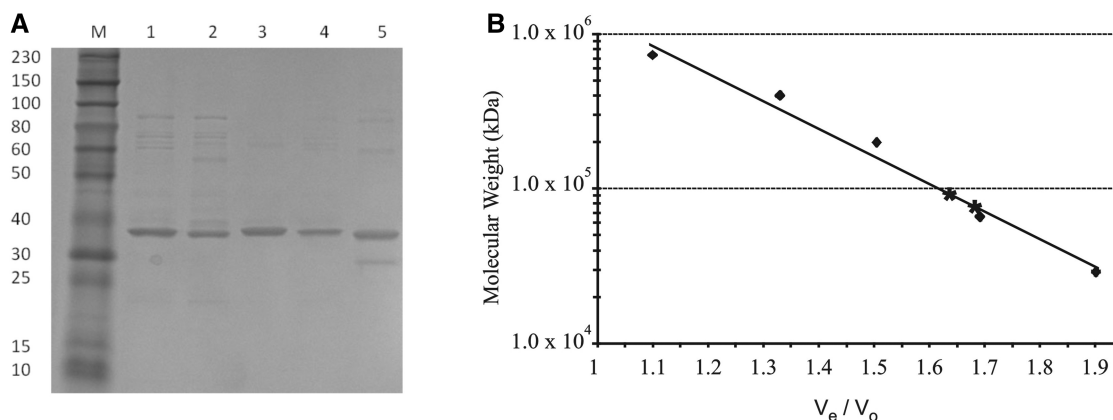


Figure 1. Purified homologues and gel filtration analysis of the homologues. (A) The following homologues were run on a Tris-glycine 4–20% gel as an example of the relative purity of the proteins. Lane M is the ColorPlus protein ladder (NEB # P7710). Lane 1: AbaTI, lane 2: AcaPI, lane 3: AbaUI, lane 4: AbaDI and lane 5: YrkI. (B) Analytical size exclusion chromatography of the homologues. The two asterisks along the standard curve indicate the narrow range of 1.62–1.69 in which all the homologues eluted. The molecular weights were determined relative to their elution volume against that of the molecular weight standards and are summarized in Table 3.

incubated with enzyme for 30 min at room temperature. The reaction products were then resolved using a 5% agarose gel.

RESULTS

A number of PvuRtsII homologues were identified by blasting the PvuRtsII 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 5 α ghmC by AGT of 100% in T4 DNA.

Relative selectivity of the PvuRtsII homologues

AbaSI, a PvuRtsII 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 PvuRtsII homologues. A similar method to that used in the initial substrate selectivity determination of PvuRtsII (22) was used to determine the substrate selectivity of the PvuRtsII 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),

phage T4 β DNA (containing 5 β ghmC) and phage T4 α DNA (containing 5 α ghmC). 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:5 α ghmC:5 β ghmC: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:5 α ghmC:5 β ghmC:5mC:C = 1024:128:16384:16:ND and BbiDI, which exhibits a relative selectivity of 5hmC:5 α ghmC:5 β ghmC: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 5 α ghmC modification had inhibitory effects of

1/500 when compared with 5 β ghmC. In contrast, for PvuRtsII, the 5 α ghmC modification enhanced selectivity by 32-fold when compared with 5 β ghmC. These modifications are important because even though they are not known to exist in the human genome, *in vitro* 5hmC can be converted to 5 α ghmC and 5 β ghmC by BGT and AGT, respectively.

Cleavage properties

To determine the cleavage properties for the PvuRtsII 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)
PvuRtsII	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	2.3 ± 0.2
YkrI	84 ± 8	33.802	2.5 ± 0.2

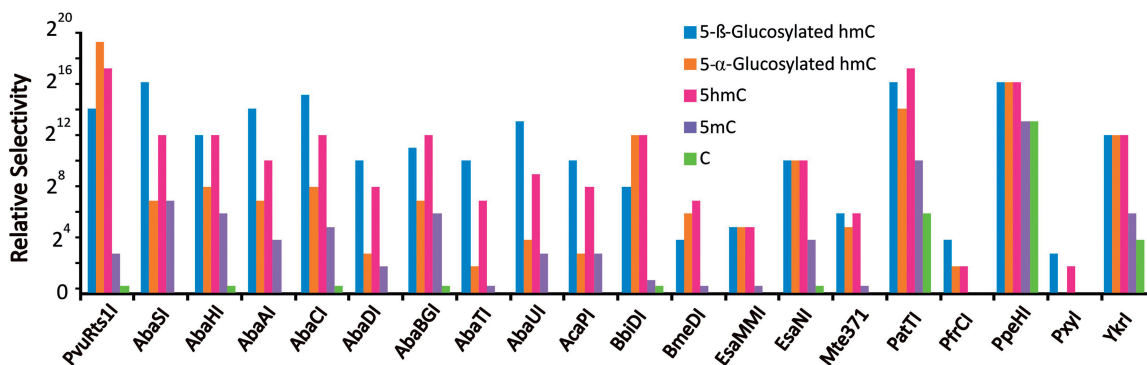


Figure 2. Relative selectivity of PvuRtsII homologues. Selectivity was determined on DNA with different modified cytosines: dcm⁻ (unmodified cytosine), XP12 (methylated cytosines), T4wt (hydroxymethylated cytosines), T4 α (α -glucosylated hydroxymethylated cytosines), T4 β (β -glucosylated hydroxymethylated cytosines).

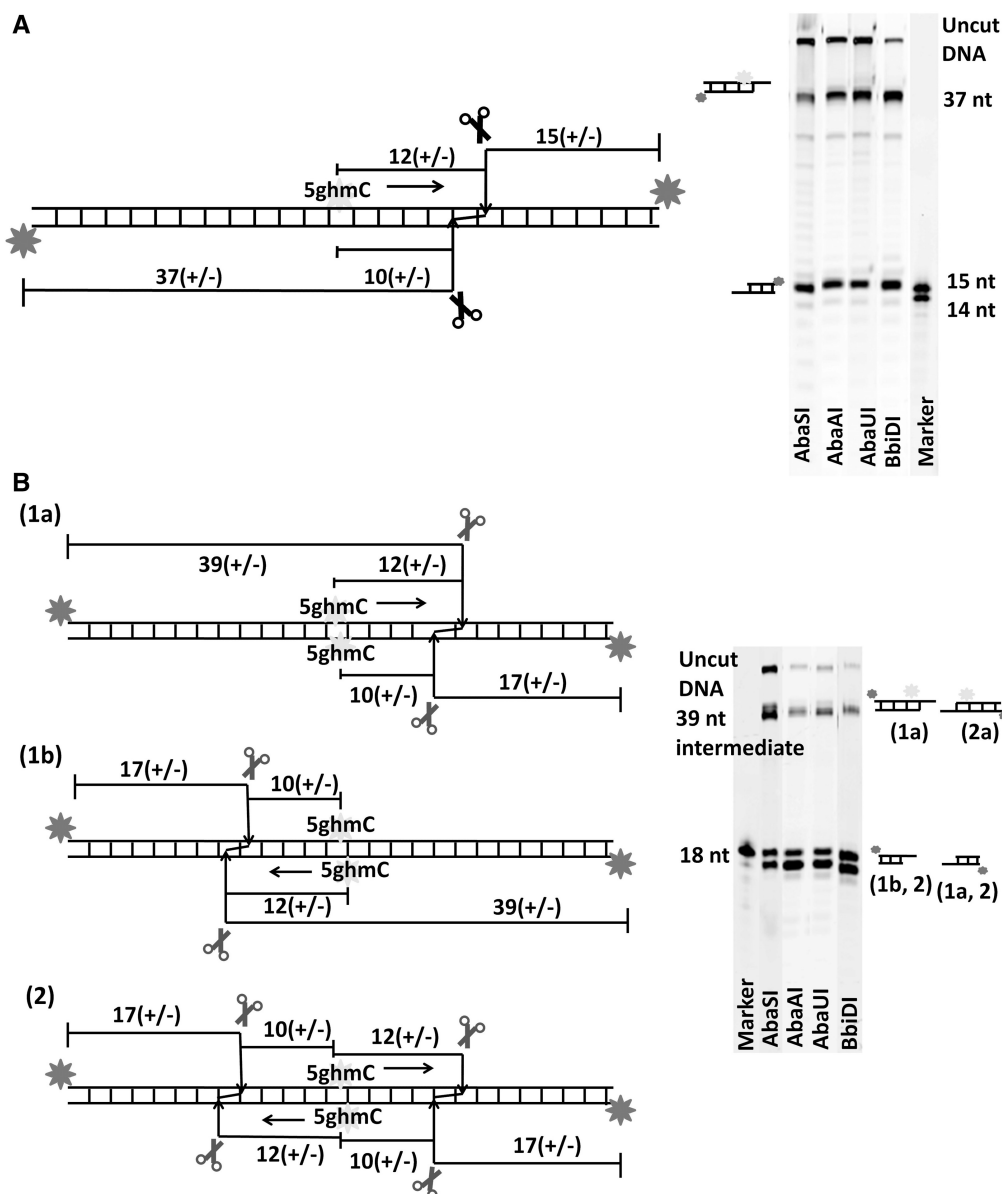


Figure 3. Cleavage site determination of the PvuRtsII homologues. (A) The left side of the figure shows the sequence of the hemi-glucosylhydroxymethylated 3'FAM-labeled 54-bp oligonucleotide used to determine the cleavage site on the same strand of the modification. The position of the modified cytosine along with the expected cut sites on the top and bottom strand is indicated. For the right side of the figure, fragments from AbaAI, AbaUI and BbiDI digestion, along with oligonucleotide markers of 15 and 14 nt were resolved on a denaturing polyacrylamide gel. Recognition of the modification by the homologues resulted in two labeled fragments of 37(+/-) nt, from cleavage on the opposite strand of the modification, and 15(+/-) nt, from cleavage on the same strand of the modification. (B) The left side of the figure shows the sequence of the fully glucosylhydroxymethylated 5'FAM-labeled oligonucleotide used to determine the cleavage site on the opposite strand of the modification. Three expected cut site scenarios are shown: (1a, 1b) recognition of the modification would result in a cleavage to the 3' side of the modification, yielding two labeled fragments of 39(+/-) nt, from cleavage on the same strand of the modification, and 17(+/-) nt, from cleavage on the opposite strand of the modification; (2) recognition of both modifications will result in a right-hand cleavage for the top-strand modification and a left-hand cleavage for the bottom-strand modification yielding two labeled bands of 17(+/-) nt, resulting from cleavage on the opposite strands of the modifications. The gel on the right side of Figure 3B shows the fragments from AbaAI, AbaUI and BbiDI digestion, along with an oligonucleotide marker of 18 nt. Recognition of the substrate by the homologues shows a mixture of labeled fragments of 39(+/-) and 17(+/-), resulting from cleavage scenarios 1a, 1b and 2. The 39(+/-) nt fragment 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.

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

Table 4. Cut sites of the PvuRtsII homologues

Enzyme	Top-strand cut ^a	Bottom-strand cut ^a
PvuRtsII	CN ₁₁₋₁₃	N ₉₋₁₀ G
AbaSI	CN ₁₁₋₁₃	N ₉₋₁₀ G
AbaHI	CN ₁₀₋₁₃	N ₉₋₁₀ G
AbaAI	CN ₁₁₋₁₃	N ₉₋₁₁ G
AbaCI	CN ₁₁₋₁₃	N ₉₋₁₀ G
AbaDI	CN ₁₁₋₁₃	N ₉₋₁₀ G
AbaBGI	CN ₁₁₋₁₃	N ₉₋₁₁ G
AbaTI	CN ₁₁₋₁₃	N ₉₋₁₁ G
AbaUI	CN ₁₁₋₁₃	N ₉₋₁₁ G
AcaPI	CN ₁₁₋₁₃	N ₉₋₁₁ G
BbiDI	CN ₁₀₋₁₃	N ₉₋₁₁ G
BmeDI	CN ₂₋₃	N ₀₋₁ G
EsaMMI	CN ₁₁₋₁₃	N ₉₋₁₁ G
EsaNI	CN ₁₁₋₁₃	N ₉₋₁₁ G
Mte37I	CN ₁₁₋₁₃	N ₉₋₁₀ G
PatTI	CN ₁₁₋₁₃	N ₉₋₁₀ G
PfrCI	CN ₁₁₋₁₃	N ₉₋₁₀ G
PpeHI	CN ₁₁₋₁₃	N ₉₋₁₁ G
PxyI	CN ₁₁₋₁₃	N ₉₋₁₀ G
YkrI	CN ₁₀₋₁₃	N ₉₋₁₀ G

^awith respect to 5ghmC modification.

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 cleave at the same position as AbaSI, N₁₁₋₁₃/N₉₋₁₀ on the 3' side of the modified cytosine (Table 4). However, there are some exceptions such as BmeDI, which cuts to a low degree at N₁₁₋₁₃/N₉₋₁₀ but predominantly at N₂₋₃/N₀₋₁, a cut site close to the 5hmC modification.

Recognition site requirements

For efficient cleavage, PvuRtsII requires both a 5hmC modification and an additional cytosine on the opposite strand. According to Hua *et al.*, 47% of the sequences cut by PvuRtsII homologue AbaSI contain two cytosines 21 nt apart and 45% contain two cytosines 22 nt apart (22). In addition, the cleavage efficiency was determined to be dependent on the modification status where when one of the 5hmCs in the recognition site changes to 5mC or C, the efficiency decreases (22). To determine whether the PvuRtsII homologues also possess specific requirements for site recognition, synthetic oligonucleotides were specifically designed to contain 5hmC on one strand and 5hmC, 5mC, C or no cytosine 22 nt away on the opposite strand (Figure 4A). The extent of digestion was determined by resolving the DNA on either a 10% tris borate EDTA or 5% agarose gel. Similar to PvuRtsII,

while the homologues show modest activity with one 5hmC modification and an additional cytosine 22 nt away and on the opposite strand, the highest activity is exhibited on substrates with two 5hmCs 22 nt apart and on opposite strands (Figure 4B). In addition, the cutting efficiency decreases as the bottom-strand 5hmC modification changes from 5mC to C and there is no detectable cutting in the absence of C. This indicates that all of the homologues have an absolute requirement for a second cytosine on the opposite strand 22 nt away.

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 appear to have evolved as a defense mechanism in the struggle between unicellular organisms and their viruses. PvuRtsII 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 PvuRtsII 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 CN₁₁₋₁₃ on the top strand and N₉₋₁₀G 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 CN₂₋₃ on the top strand and N₀₋₁G on the bottom strand. After performing a sequence alignment with the PvuRtsII 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 PvuRtsII 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

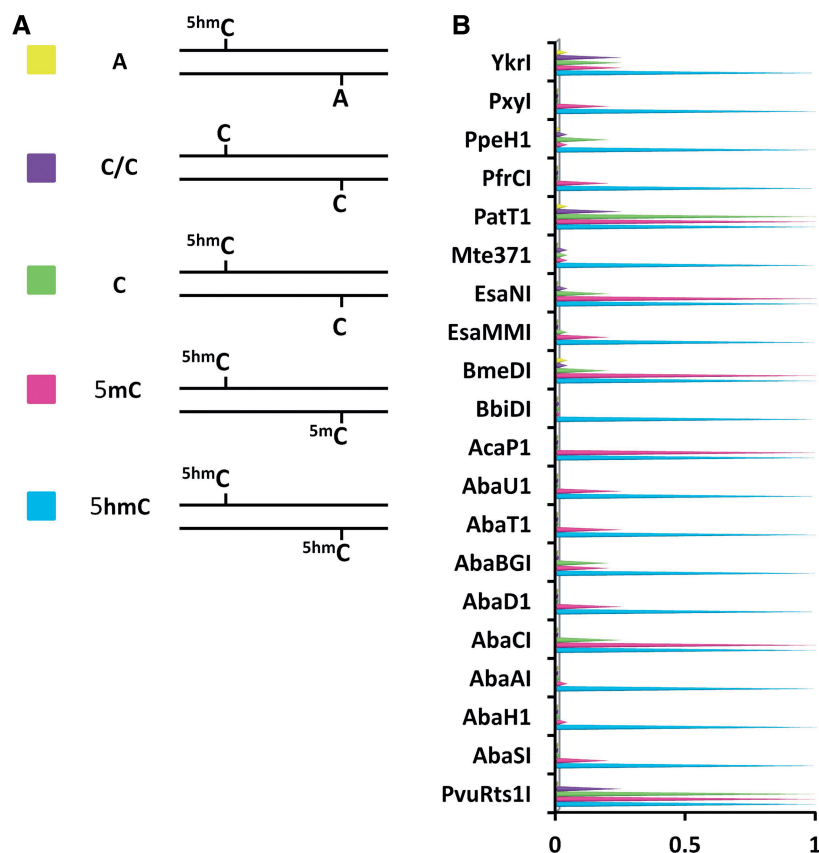


Figure 4. Activity of enzymes on different modified oligonucleotides. The sequence of the oligonucleotides can be found in Table 2. (A) Schematic of the five modified oligonucleotides used for the activity determination; A, C/C, C, 5mC and 5hmC. The two indicated residues on each substrate are 22 nt 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 5 α ghmC 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 α - and β -glucosylated 5hmC DNA. For example, BbiDI, BmeDI and PvuRts1I show high selectivity for 5 α ghmC but are inhibited by 5 β ghmC, while AbaCI, AbaUI, AbaAI and AbaSI (to name a few) show high selectivity for 5 β ghmC but are inhibited by 5 α ghmC (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 PvuRts1I homologues exhibiting either enhanced or inhibitory effects toward 5 α ghmC 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 PvuRts1I 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 5βghmC 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

- Warren, R.A. (1980) Modified bases in bacteriophage DNAs. *Annu. Rev. Microbiol.*, **34**, 137–158.
- 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.
- 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.
- Sutherland, E., Coe, L. and Raleigh, E.A. (1992) McrBC: a multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol.*, **225**, 327–348.
- Xu, S.Y., Corvaglia, A.R., Chan, S.H., Zheng, Y. and Linder, P. (2011) A type IV modification-dependent restriction enzyme SauUSI from *Staphylococcus aureus* subsp. *aureus* USA300. *Nucleic Acids Res.*, **39**, 5597–5610.
- Janosi, L., Yonemitsu, H., Hong, H. and Kaji, A. (1994) Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRtsI) modulated by glucosylation of DNA. *J. Mol. Biol.*, **242**, 45–61.
- Rifat, D., Wright, N.T., Varney, K.M., Weber, D.J. and Black, L.W. (2008) Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target. *J. Mol. Biol.*, **375**, 720–734.
- Lehman, I.R. and Pratt, E.A. (1960) On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. *J. Biol. Chem.*, **235**, 3254–3259.
- Rottach, A., Leonhardt, H. and Spada, F. (2009) DNA methylation-mediated epigenetic control. *J. Cell Biochem.*, **108**, 43–51.
- 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.
- Kriakou, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
- 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.
- Meyer, C., Kowarz, E., Hofmann, J., Renneville, A., Zuna, J., Trka, J., Ben Abdelali, R., Macintyre, E., De Braekeleer, E., De Braekeleer, M. *et al.* (2009) New insights to the MLL recombinome of acute leukemias. *Leukemia*, **23**, 1490–1499.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D. *et al.* (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, **146**, 67–79.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
- Maiti, A. and Drohat, A.C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.*, **286**, 35334–35338.
- Valinluck, V. and Sowers, L.C. (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.*, **67**, 946–950.
- Valinluck, V., Tsai, H.H., Rogstad, D.K., Burdzy, A., Bird, A. and Sowers, L.C. (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.*, **32**, 4100–4108.
- Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W. and Balasubramanian, S. (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science*, **336**, 934–937.
- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B. *et al.* (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, **149**, 1368–1380.
- 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.
- Wang, H., Guan, S., Quimby, A., Cohen-Karni, D., Pradhan, S., Wilson, G., Roberts, R.J., Zhu, Z. and Zheng, Y. (2011) Comparative characterization of the PvuRtsII family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. *Nucleic Acids Res.*, **39**, 9294–9305.
- Sun, Z., Terragni, J., Borgaro, J.G., Liu, Y., Yu, L., Guan, S., Wang, H., Sun, D., Cheng, X., Zhu, Z. *et al.* (2012) High resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep.*, **3**, 1–10.
- Puigbo, P., Guzman, E., Romeu, A. and Garcia-Valle, S. (2007) OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Res.*, **35**, W126–W131.
- Kuo, T.T., Huang, T.C. and Teng, M.H. (1968) 5-Methylcytosine replacing cytosine in the deoxyribonucleic acid of a bacteriophage for *Xanthomonas oryzae*. *J. Mol. Biol.*, **34**, 373–375.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1990) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Georgopoulos, C.P. and Revel, H.R. (1971) Studies with glucosyl transferase mutants of the T-even bacteriophages. *Virology*, **44**, 271–285.
- 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.
- Shi, D., Yu, X., Cabrera-Luque, J., Chen, T.Y., Roth, L., Morizono, H., Allewell, N.M. and Tuchman, M. (2007) A single mutation in the active site swaps the substrate specificity of N-acetyl-L-ornithine transcarbamylase and N-succinyl-L-ornithine transcarbamylase. *Protein Sci.*, **16**, 1689–1699.
- Todd, A.E., Orengo, C.A. and Thornton, J.M. (2001) Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.*, **307**, 1113–1143.
- Gruber, T.D., Borrok, M.J., Westler, W.M., Forest, K.T. and Kiessling, L.L. (2009) Ligand binding and substrate discrimination by UDP-galactopyranose mutase. *J. Mol. Biol.*, **391**, 327–340.