

Characterization of PvuRts1I endonuclease as a tool to investigate genomic 5-hydroxymethylcytosine

Aleksandra Szwagierczak, Andreas Brachmann, Christine S. Schmidt, Sebastian Bultmann, Heinrich Leonhardt* and Fabio Spada*

Ludwig Maximilians University Munich, Department of Biology and Center for Integrated Protein Science Munich (CIPS^M), 82152 Planegg-Martinsried, Germany

Received October 18, 2010; Revised February 14, 2011; Accepted February 15, 2011

ABSTRACT

In mammalian genomes a sixth base, 5-hydroxymethylcytosine (^{hm}C), is generated by enzymatic oxidation of 5-methylcytosine (^mC). This discovery has raised fundamental questions about the functional relevance of ^{hm}C in mammalian genomes. Due to their very similar chemical structure, discrimination of the rare ^{hm}C against the far more abundant ^mC is technically challenging and to date no methods for direct sequencing of ^{hm}C have been reported. Here, we report on a purified recombinant endonuclease, PvuRts1I, which selectively cleaves ^{hm}C-containing sequences. We determined the consensus cleavage site of PvuRts1I as ^{hm}CN_{11–12}/N_{9–10}G and show first data on its potential to interrogate ^{hm}C patterns in mammalian genomes.

INTRODUCTION

In higher eukaryotes, only the C5 position of genomic cytosine is subject to enzymatically catalyzed post-replicative modification. Methylation at this position has long been known to play major roles in epigenetic control of transcriptional activity and, as a consequence, to affect fundamental processes such as development (including natural reprogramming of cell fate), imprinting, X chromosome inactivation, genome stability and predisposition to neoplastic transformation (1,2). The recent discovery of the further modification of 5-methylcytosine (^mC) to 5-hydroxymethylcytosine (^{hm}C) by the family of Tet dioxygenases has raised major questions on the functional relevance of this sixth base in mammalian genomes (3,4). While recent evidence supports a role for ^{hm}C as an intermediate in the erasure of cytosine methylation (5), other roles in controlling

genomic functions cannot be excluded. The definition of these roles will require profiling of genomic ^{hm}C patterns, which presents a major technical challenge as ^{hm}C is structurally and chemically very similar to ^mC but in general far less abundant in mammalian genomes (3,4,6–9). The gold standard methodology for profiling of genomic ^mC sites, bisulfite conversion, cannot discriminate ^{hm}C from ^mC and all available restriction endonucleases are either equally sensitive to ^mC and ^{hm}C or not sensitive to either (10–12). While antibodies raised against ^{hm}C are commercially available, their use to probe ^{hm}C frequency by DNA immunoprecipitation has yet to be reported and the accuracy of this method will depend on the relative affinity of these antibodies for ^{hm}C versus ^mC as the latter is present in large excess in mammalian genomes. Very recently enzymatic methods for selective labeling and identification of ^{hm}C have been reported (7,13).

Interestingly, ^{hm}C is also present in the genomes of viruses that infect bacteria and unicellular algae, where it serves as protection against the restriction systems of the host. In particular, ^{hm}C accounts for up to 100% of the cytosine residues in the genomes of T-even coliphages. In these phages the hydroxymethyl group is added at the level of the dCMP precursor and further linked to glucose (in both α - and β -configurations) or gentiobiose after incorporation of the nucleotide in the genome (14–16). We sought to exploit enzymatic activities that evolved as part of the struggle between bacteria and these viruses to selectively detect ^{hm}C in mammalian genomes. Recently, we described an assay for quantification of global genomic ^{hm}C levels based on the transfer of tritiated glucose to ^{hm}C by T4 β -glucosyltransferase (7). Interestingly, restriction systems have evolved in bacteria that address the phage counter defense measures by specifically recognizing modified cytosine. Among these the McrBC system and the recently described MspJI endonuclease recognize sequences containing both ^mC and ^{hm}C (17,18) and

*To whom correspondence should be addressed. Tel: +49 89 2180 74230; Fax: +49 89 2180 74236; Email: f.spada@lmu.de
Correspondence may also be addressed to Heinrich Leonhardt. Tel: +49 89 2180 74232; Fax: +49 89 2180 74236; Email: h.leonhardt@lmu.de

therefore *per se* are not useful to discriminate these modified cytosines. At least two endonucleases, PvuRtsII and GmrSD, were shown to restrict DNA containing glucosylated ^{hm}C (19,20). However, GmrSD does not cleave non-glucosylated (^{hm}C-containing) T4 DNA, has the additional disadvantages of being a heterodimer and of co-purifying with the GroEL chaperonin (19). PvuRtsII is encoded by a single gene present on the kanamycin resistance plasmid RtsI originally isolated from *Proteus vulgaris* and its restriction activity *in vivo* was shown to be modulated by ^{hm}C glucosylation in a complex fashion (20). However, as PvuRtsII was not purified, its activity has not been characterized *in vitro*.

Here, we show that purified recombinant PvuRtsII selectively cleaves ^{hm}C-containing DNA and determine its cleavage site. In addition, we present initial data on the use of ^{hm}C as a tool to investigate ^{hm}C patterns in mammalian genomes.

MATERIALS AND METHODS

Cloning and purification of PvuRtsII

The sequence encoding PvuRtsII was synthesized at Mr Gene GmbH (Regensburg) and cloned into the pET28a vector (Novagen). BL21(DE3) *Escherichia coli* cells carrying the expression vector were grown in LB medium at 37°C until $A_{600} = 0.6-0.7$ and induced with 1 mM isopropyl β -D-thiogalactopyranoside for 16 h at 18°C. Lysates were prepared by sonication in 300 mM NaCl, 50 mM Na₂HPO₄ pH 8.0, 10 mM imidazole, 10% glycerol and 1 mM β -mercaptoethanol, cleared by centrifugation and applied to a nickel-nitrilotriacetic acid column (QIAGEN) pre-equilibrated with lysis buffer. Washing and elution were performed with lysis buffer containing 20 and 250 mM imidazole, respectively. Eluted proteins were applied to a Superdex S-200 preparative gel filtration column (GE Healthcare) in 150 mM NaCl, 20 mM Tris pH 8.0, 10% glycerol, 1 mM DTT and peak fractions were pooled. The stability of PvuRtsII upon storage was improved by supplementation with 10% glycerol.

Preparation of DNA substrates

In vivo α/β -glucosylated and non-glucosylated T4 phage DNA was isolated essentially as described (4). Briefly, T4 stocks were propagated on *E. coli* strain CR63, which was also used for the isolation of glucosylated T4 DNA. To isolate non-glucosylated T4 DNA, wild-type T4 phage was amplified on an ER1565 *galU* mutant strain. β -glucosylated T4 DNA was generated *in vitro* by treatment of non-glucosylated T4 DNA with purified T4 β -glucosyltransferase (7). Genomic DNA was isolated from mouse cerebellum and triple knockout (TKO) embryonic stem cells (ESCs) (21) as described (7).

Reference DNA fragments containing exclusively ^{hm}C, ^mC or unmodified cytosine residues were prepared by PCR using 5-hydroxymethyl-dCTP (Bioline GmbH), 5-methyl-dCTP (Jena Bioscience GmbH) and dCTP, respectively. T4 phage DNA template, Phusion HF DNA

Polymerase (Finnzymes) and primer 5'-GTG AAG TAA GTA ATA AAT GGA TTG-3', which does not contain cytosine residues, were used for amplification of all reference DNA fragments. To generate the reference 1139 bp fragment with 100% ^{hm}C for restriction with PvuRtsII the second primer was 5'-TGG AGA AGG AGA ATG AAG AAT AAT-3', which also does not contain cytosine residues. To generate the 800 and 500 bp control substrates containing only ^mC and only unmodified cytosine for restriction with PvuRtsII the second primer was 5'-GCC ATA TTG ATA ATG AAA TTA AAT GTA-3' and 5'-TCA GCA ATT TTA ATA TTT CCA TCT TC-3', respectively. PCR products were purified by gel electrophoresis followed by silica column purification (Nucleospin, Macherey-Nagel). The 140 bp fragment used to determine the orientation of the PvuRtsII cleavage overhang was amplified with primers 5'-TAT ACT GAA GTA CTT CAT CA-3' and 5'-CTT TGC GTG ATT TAT ATG TA-3'.

For the preparation of substrates with a single PvuRtsII consensus containing ^{hm}C or ^mC in symmetrical or asymmetrical configuration a 94 bp fragment was amplified from the T4 genome with primers 5'-CTC GTA GAC TGC GTA CCA ATC TAA CTC AGG ATA GTT GAT-3' and 5'-TAT GAT AAG TAT GTA GGT TAT T-3'. This fragment contains a single site corresponding to the identified PvuRtsII consensus ^{hm}CN₁₁₋₁₂/N₉₋₁₀G and was used as a template according to the strategy depicted in Figure 3. To generate substrates with symmetric cytosine modifications or unmodified cytosine the fragment was amplified with forward primer 5'-CTC GTA GAC TGC GTA CCA-3' and reverse primer 1 5'-TAT GAT AAG TAT GTA GGT TAT T-3' in the presence of the respective modified or unmodified dCTP. To generate substrates with asymmetric cytosine modifications the same forward primer was paired with reverse primer 2 5'-TAT GAT AAG TAT GTA GGT TAT TCA A-3'.

DNA restriction with PvuRtsII and identification of cleavage and recognition site

Unless otherwise stated the reaction conditions contained 150 mM NaCl, 20 mM Tris pH 8.0, 5 mM MgCl₂, 1 mM DTT. One unit of PvuRtsII was defined as amount of enzyme required to digest 1 μ g of ^{hm}C-containing T4 DNA in 15 min at 22°C. For assessment of enzyme specificity, 100 ng of each control fragment were digested separately or together with 200 ng of genomic DNA in 30 μ l reactions containing standard buffer and 1 U of purified PvuRtsII at 22°C for 15 min.

For identification of the cleavage and recognition site, the 1139 bp fully hydroxymethylated fragment amplified from the T4 genome or whole non-glucosylated T4 DNA were digested under standard conditions. Fragment ends were blunted with Klenow polymerase (NEB) and cloned using the Zero Blunt[®] PCR Cloning Kit (Invitrogen). Randomly selected clones were sequenced and the data were analyzed using WebLogo (22).

RESULTS

^{hm}C-specific endonuclease activity of PvuRtsII

His-tagged PvuRtsII was expressed in *E. coli* and purified to homogeneity by sequential Ni²⁺ affinity and size exclusion chromatography (Figure 1A). As bacteria carrying the RtsI plasmid were shown to restrict the ^{hm}C-containing T-even phages, but not ^mC-containing T-odd phages or λ phage, which does not contain modified cytosine (20), we initially used T4 genomic DNA as a substrate to test the activity of purified PvuRtsII. T4 genomic DNA was isolated from both *galU*⁺ and *galU*⁻ strains, the latter being UDP-glucose deficient and thus containing only non-glucosylated ^{hm}C. Under the same digestion conditions non-glucosylated T4 DNA was digested more efficiently than both naturally α- and β-glucosylated and *in vitro* β-glucosylated counterparts (Figure 1B). Non-glucosylated T4 DNA was cleaved into fragments with an apparent size of about 200 bp, indicating that PvuRtsII recognizes a frequently occurring sequence (Figure 1B and Supplementary Figures S1 and S2). We then used non-glucosylated T4 DNA to test the activity of the enzyme under various conditions. PvuRtsII was strictly dependant on Mg²⁺ ions, which could not be substituted with Ca²⁺, and endonuclease activity was maximal in the presence of 100–200 mM NaCl (Supplementary Figure S1A and B). However, during purification we observed that the enzyme is unstable in solutions of ionic strength lower than 150 mM NaCl. The activity of PvuRtsII was found highest at pH 7.5–8.0 and was unaffected by the presence of Tween 20 or Triton X-100 (Supplementary Figure S2A and B). We also observed that after prolonged incubation PvuRtsII precipitates even at room temperature, consistent with the reported temperature sensitivity of the phage restriction activity in cells carrying the RtsI plasmid (20). Upon short incubation times maximal activity was observed at 22°C (Supplementary Figure 2C). Thus, the relative amounts of enzyme and DNA substrate were standardized so that digestion was complete in 15 min at 22°C in the presence of 150 mM NaCl (Supplementary Figures S1C and S2C).

The specificity of PvuRtsII with respect to cytosine modification was further tested by digesting reference fragments containing exclusively unmodified cytosine

(500 bp), ^mC (800 bp) or ^{hm}C (1139 bp; Figure 1C). Under standard digestion conditions purified PvuRtsII selectively cleaved the ^{hm}C-containing fragment, consistent with the relative restriction efficiency of bacteriophages with distinct cytosine modifications by bacteria carrying the RtsI plasmid (20).

Determination of PvuRtsII cleavage sites

To identify the cleavage pattern of PvuRtsII we generated libraries of restriction fragments from either the whole T4 genome (Supplementary Figure S3) or an 1139 bp fragment amplified from the same genome containing exclusively hydroxymethylated cytosines (Figure 2). Random sequencing of 161 and 133 fragment ends from the whole T4 genome and 1139 bp fragment libraries revealed that 85 and 89%, respectively, matched the consensus sequence ^{hm}CN_{11–12}/N_{9–10}G. Among these 78 and 87%, respectively, showed one of three similar sequence patterns, ^{hm}CN₁₂/N₁₀G, ^{hm}CN₁₂/N₉G and ^{hm}CN₁₁/N₉G, while for the remaining fragment ends the exact number of nucleotides between the modified cytosine and the cleavage site could not be determined unambiguously due to the occurrence of multiple ^{hm}C residues upstream of the cleavage site. Of the sequenced fragment ends, 14 and 11% from the whole T4 genome and 1139 bp fragment libraries, respectively, did not match the ^{hm}CN_{11–12}/N_{9–10}G consensus. However, 100 and 80% of these ends, respectively, contained at least one ^{hm}C residue 10–13 nt upstream of the cleavage site, while no guanine was present in the T4 genomic sequence 10–11 nt downstream the cleavage site (Supplementary Figure S4). The sequenced clones from the 1139 bp T4 genomic fragment library corresponded to an 81% coverage of the fragment, with some PvuRtsII fragments occurring multiple times, while other fragments that were predicted on the basis of the ^{hm}CN_{11–12}/N_{9–10}G consensus were not found (Figure 2 and Supplementary Figure S5). Examination of the missing fragments did not show any common sequence feature beyond the ^{hm}CN_{11–12}/N_{9–10}G consensus (Supplementary Figure S6), suggesting that their absence from the sequenced fragments was due to limited sampling. Alignment of sequenced fragment ends from the T4 genomic fragment library showed that 2 nt around the cleavage site were missing from all clones, suggesting a 2 nt 3'-overhang cleavage pattern

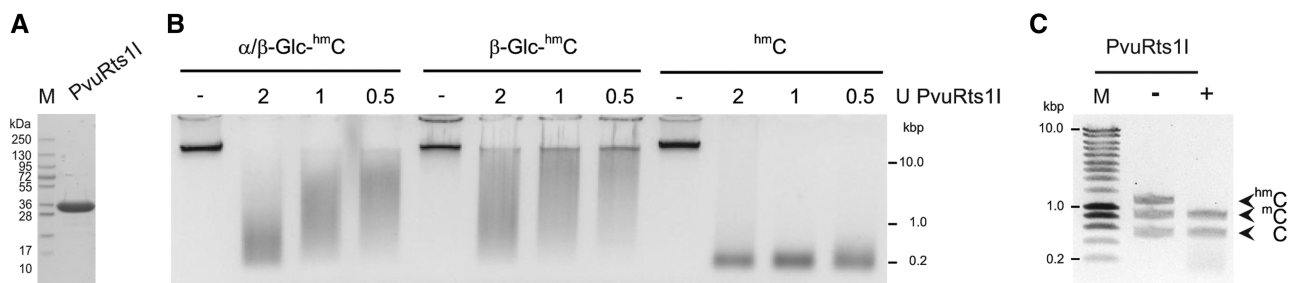


Figure 1. Selective restriction of ^{hm}C-containing DNA by PvuRtsII. (A) Purified PvuRtsII was resolved on a SDS-polyacrylamide gel and stained with coomassie blue. (B) T4 genomic DNA with the naturally occurring pattern of α- and β-glucosylated ^{hm}C, only β-glucosylated ^{hm}C or non-glucosylated ^{hm}C was incubated without or with decreasing amounts of PvuRtsII as indicated. (C) Reference PCR fragments of 1139, 800 and 500 bp containing ^{hm}C, ^mC and unmodified cytosine at all cytosine residues, respectively, were incubated with or without PvuRtsII as indicated.

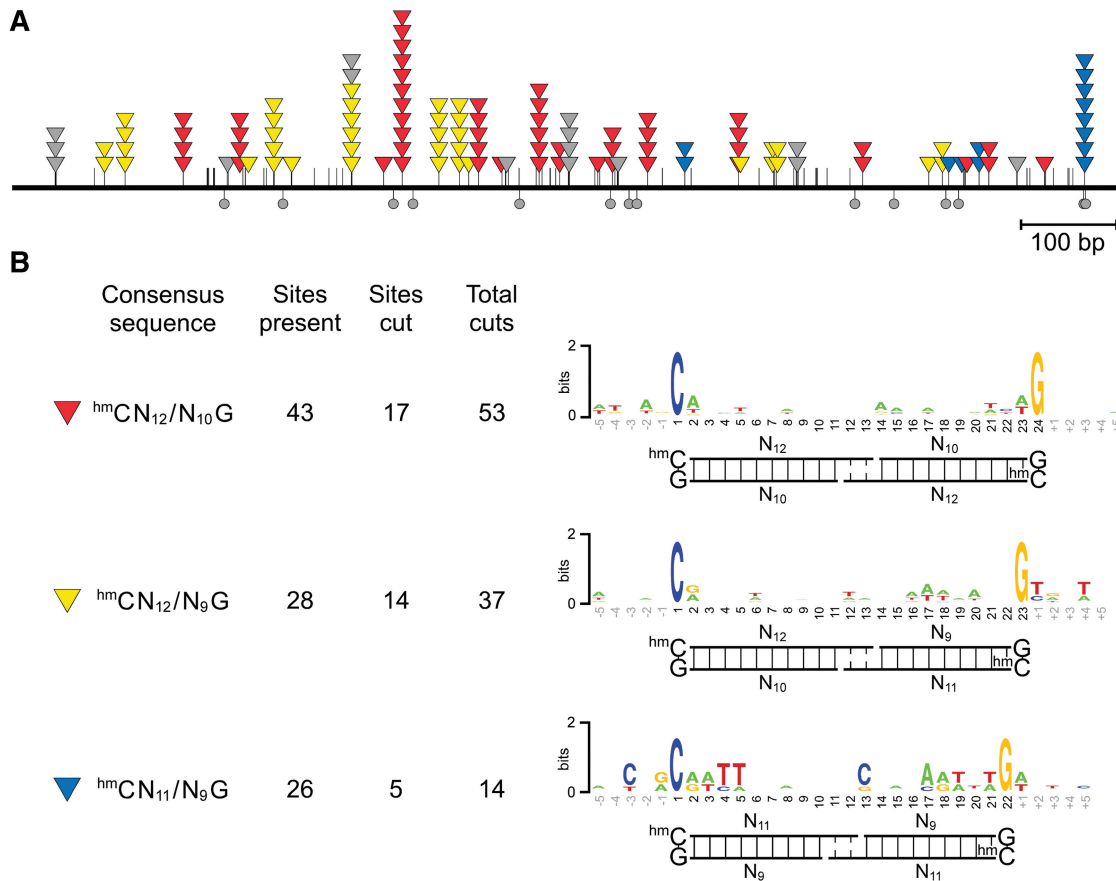


Figure 2. Cleavage site of PvuRtsII. A library of PvuRtsII restriction fragments was generated from an 1139 bp PCR fragment containing only hydroxymethylated cytosine residues and the sequence of 133 restriction fragment ends from randomly chosen clones was determined. (A) Graphical map of the fragment ends. A total of 119 analyzed fragment ends (triangles) matched the consensus sequence ^{hm}CN_{11–12}/N_{9–10}G, which was present at 97 sites (thin vertical lines) in the 1139 bp PCR fragment (thick horizontal line). Fifty three fragment ends related to the sequence motif ^{hm}CN₁₂/N₁₀G (dark green triangles), 37 to ^{hm}CN₁₁/N₁₀G (bright green triangles) and 14 to ^{hm}CN₁₁/N₉G (light green triangles), while 15 fragment ends matching the consensus sequence ^{hm}CN_{11–12}/N_{9–10}G could not be assigned unambiguously to any of these subsets (gray triangles). Fourteen fragment ends did not match the prevalent consensus sequence (gray circles, see Supplementary Figure S3). (B) Occurrence of the three subsets of cleavage sites and LOGO representation of the corresponding consensus sequence. The absolute height of each position reflects its overall conservation, while the relative height of nucleotide letters represents their relative frequency. The slash in the three cleavage sequence subtypes indicates the exact cleavage site.

(Supplementary Figure S5). This was confirmed by direct sequencing of the two fragments generated by digestion of a 140 bp amplicon containing a single PvuRtsII site (Supplementary Figure S7).

The results above reveal a symmetric nature of the preferred cleavage sites and raise the issue of PvuRtsII activity on sites with modified cytosine in symmetric and asymmetric configuration. To clarify this issue, we used a PCR strategy to generate DNA substrates with identical sequence and containing a single PvuRtsII consensus site with ^{hm}C or ^mC in symmetrical and asymmetrical configurations or no modified cytosine (Figure 3A). In the presence of enzyme amounts that did not cleave substrates with unmodified and ^mC sites, digestion of substrates with asymmetric ^{hm}C at the PvuRtsII site was reduced with respect to substrates with symmetric ^{hm}C, but still appreciable. Residual undigested substrate with symmetric ^{hm}C at the PvuRtsII site in these reaction conditions was typically observed with such short substrates, but not with longer ones.

Digestion of mammalian genomic DNA with PvuRtsII

To investigate cleavage site preference and efficiency of PvuRtsII digestion for mammalian genomic DNA, we initially selected the upstream regulatory region III of the mouse *nanog* gene (23). As this region was shown to be bound by Tet1 and to acquire CpG methylation upon knockdown of Tet1 in ESCs (5), it represents a potential candidate as a mammalian genomic sequence containing ^{hm}C. Real time amplification of this region from ESC genomic DNA did not show a significant decrease of product after PvuRtsII digestion (data not shown). We then devised a strategy to positively identify rare PvuRtsII digestion products. After PvuRtsII digestion genomic fragments were ligated to a linker with a random 2 nt 3'-overhang. Ligation products were then amplified using *nanog* specific primers paired with a linker specific primer, but no amplification product could be obtained (data not shown). This result may be explained by an extremely seldom occurrence of ^{hm}C at

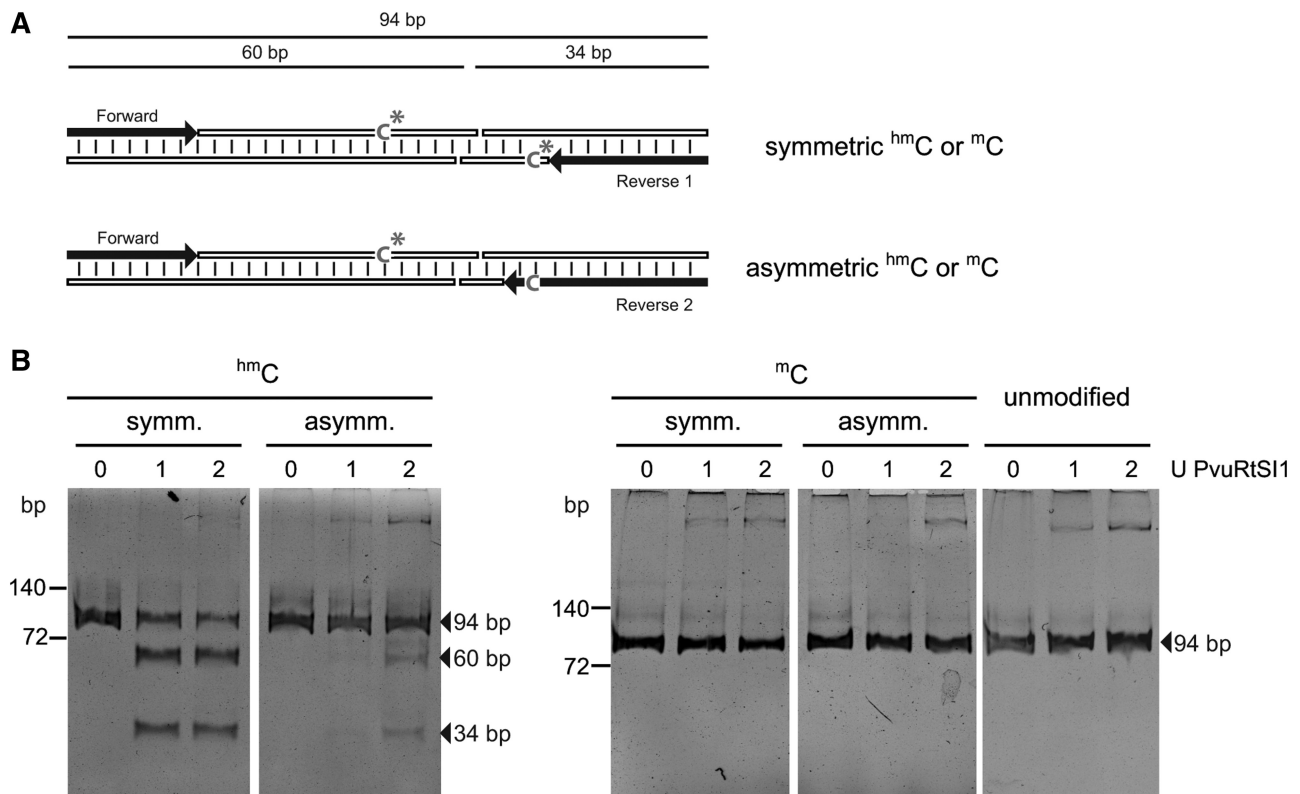


Figure 3. Differential activity of PvuRtsII on sites with symmetric and asymmetric ^{hm}C. Ninety-four bp long substrates with identical sequence were generated that contain a single PvuRtsII consensus site (CN₁₂/N₁₀G) with ^{hm}C or ^mC in symmetrical and asymmetrical configurations or no modified cytosine. **(A)** Strategy for generation of the substrates by PCR amplification in the presence of modified nucleotides. The size of the PvuRtsII digestion products is indicated. **(B)** The variously modified substrates were digested with the indicated amounts of PvuRtsII and digestion products were resolved on polyacrylamide gels. Note the reduced but tangible digestion of the substrate containing asymmetric ^{hm}C.

cleavage sites of this locus (especially in symmetric configuration), inefficiency of PvuRtsII digestion or both. In this regard, it is important to consider that positive identification of ^{hm}C sites in this region of the *nanog* locus has actually not been reported for ESCs. In addition, during the revision of the present work a manuscript was published (24) that could not confirm the reduced *nanog* expression and ESC differentiation previously reported upon Tet1 knockdown (5), raising uncertainty about the actual occurrence of ^{hm}C at the *nanog* promoter in ESCs.

As there are no clear and quantitative data on the levels and density of ^{hm}C at specific genomic sites available yet we generated defined substrates to validate the PvuRtsII cut-ligation amplification protocol for the identification of ^{hm}C sites. We PCR amplified region III of the *nanog* promoter in the presence of increasing concentrations of 5-hydroxymethyl-dCTP and confirmed the incorporation of proportional levels of ^{hm}C using the recently reported β -glucosylation assay (7) (data not shown). Fragment samples with increasing ^{hm}C content were then digested with PvuRtsII and the same ligation/PCR strategy for the identification of digestion products was applied as described above (Supplementary Figure S8A). Detection of fragments with ends corresponding to the PvuRtsII cleavage pattern raised with increasing ^{hm}C content.

We previously quantified global ^{hm}C levels in genomic DNA from ESCs and adult somatic tissues using *in vitro*

^{hm}C glucosylation (7). Consistent with other studies (3,6,8,9), this analysis revealed that genomic DNA from adult brain regions has a high ^{hm}C content. In addition, we showed that in ESCs that are TKO for all three major DNA methyltransferases Dnmt1, 3a and 3b (21) genomic ^{hm}C levels were around the estimated limit of detection, although reproducibly above background. Therefore, we compared the PvuRtsII restriction pattern of genomic DNA from cerebellum and TKO ESCs as representative of samples with high and very low ^{hm}C levels, respectively. As internal controls, we co-digested each of the two genomic DNA samples with the same reference fragments as used to test the specificity of PvuRtsII with respect to cytosine modification (Figure 1C). As expected from the relative low abundance of ^{hm}C in mammalian genomic DNA, there was a limited reduction of high molecular weight fragments and appearance of lower molecular weight smear (Figure 4). However, DNA from cerebellum was clearly digested to a higher extent than DNA from TKO ESCs as evident from the line scans across the respective gel lanes (Figure 4). The low but appreciable degree of digestion observed for genomic DNA from TKO ESCs does not seem to result from relaxed specificity or contaminating nuclease activities, as only control substrates containing ^{hm}C, but not ^mC or unmodified cytosine, were digested when incubated either separately or together with genomic DNA (Figure 1C and Figure 4).

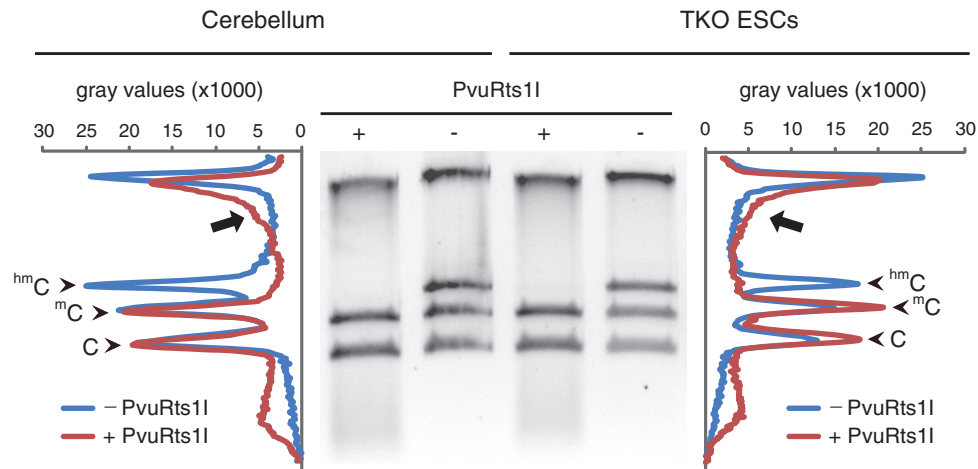


Figure 4. Restriction of mouse genomic DNA by PvuRts1I reflects ^{hm}C content. Genomic DNA from mouse cerebellum or TKO ESCs was mixed with three reference PCR fragments of 1139, 800 and 500 bp containing ^{hm}C , ^{m}C and unmodified cytosine at all cytosine residues, respectively, and incubated with or without PvuRts1I as indicated. Digests were resolved on a 0.8% agarose gel stained with ethidium bromide. Line scans of the gel lanes are aligned to the image of the gel. Red and blue lines correspond to samples incubated with and without enzyme, respectively. Arrows point to the main difference in the profiles from cerebellum and TKO ESC DNA digested with PvuRts1I (red lines).

Absence of digestion of control substrates containing ^{m}C and unmodified cytosine was evident from the unaltered ratio of their respective signals in the presence and absence of enzyme. This result shows that the extent of digestion by PvuRts1I reflects the relative ^{hm}C content in mammalian genomic DNA.

DISCUSSION

Several modification and restriction systems have evolved as defense and counter defense strategies in the struggle between unicellular microorganisms and their viruses. Here, we show that, in contrast to previously characterized endonucleases which cleave ^{hm}C -containing sequences, PvuRts1I has a preference for the non-glucosylated form of this base and discriminates against ^{m}C . This specificity makes PvuRts1I an attractive tool to investigate genomic ^{hm}C patterns in higher eukaryotes and complements the very recently published methods for enzymatic labeling of this sixth base (7,13).

Importantly, we show that the extent of PvuRts1I digestion reflects the relative abundance of ^{hm}C in genomic DNA from cerebellum and TKO ESCs. The limited extent of digestion even for samples with relatively high ^{hm}C content is in line with the cleavage site preference and dependence on cytosine modification that we determined. We calculate that the statistical probability of the PvuRts1I consensus site $CN_{11-12}/N_{9-10}G$ in the mouse genome is 0.126. Combined with the global ^{hm}C occurrence in mouse tissues (up to 0.13% of all bases or 0.65% of Cs) (3,7-9) this translates into a PvuRts1I cleavage site every 1.9×10^5 bases. As this is in the size range of fragments typically obtained with standard procedures for isolation of genomic DNA, more careful isolation methods should be used and/or PvuRts1I specific ends could be enriched by ligating biotinylated PvuRts1I compatible linkers. Alternatively, digestion conditions could be optimized or DNA could be denatured and a

second strand synthesized with ^{hm}C nucleotides to cut and reveal the likely more abundant hemimodified PvuRts1I sites.

Notably, while cerebellum has been previously reported among the tissues with the highest levels of genomic ^{hm}C (3,7,9), complete absence of ^{m}C and therefore ^{hm}C would be expected in TKO ESCs due to the lack of all three major Dnmts (21). However, we previously detected ^{hm}C levels slightly above background in TKO ESCs (7) and here we show minimal but appreciable digestion by PvuRts1I. In this context, it is interesting to note that ESCs express the highly conserved Dnmt2 (25,26), the only Dnmt family member with an intact catalytic domain that has not been genetically inactivated in TKO ESCs. Although Dnmt2 has a major role as a tRNA methyltransferase and its function as a DNA methyltransferase is still debated (27-32), it was recently shown to methylate genomic sequences in *Drosophila* (32,33). Future work should clarify whether the genome of TKO ESCs harbors any residual ^{m}C and ^{hm}C .

Restriction of genomic DNA with PvuRts1I may be combined with PCR amplification for analysis of specific loci or with massive parallel sequencing or microarray hybridization for genome-wide mapping. The calculations reported above for the frequency of PvuRts1I cleavage sites based on a random ^{hm}C distribution bring up the argument that the extent of random breaks in genomic DNA preparations would contribute very significant noise in deep sequencing and microarray applications. This drawback may at least be partially overcome if specific PvuRts1I ends are enriched by ligating linkers with a random 2 nt 3'-overhang as described here and discussed above, a strategy that can be integrated with procedures for generation of sequencing libraries. Also, our simulation of genomic fragments containing known levels of randomly distributed ^{hm}C clearly shows that relatively high local concentrations of ^{hm}C sites are required for efficient detection by PvuRts1I. The first genome-wide ^{hm}C

profiles from mammalian tissues have just been reported (13). From these first data sets, it is apparent that genomic ^{hm}C is not randomly distributed and that its accumulation in gene bodies is proportional to transcriptional activity. Thus, PvuRtsII may prove a valuable tool to probe ^{hm}C accumulation at defined genomic regions. In addition, the selectivity of PvuRtsII for ^{hm}C-containing sites may constitute an advantage with respect to endonucleases such as McrBC and MspJ1 as these enzymes do not discriminate between ^mC and ^{hm}C and require *in vitro* enzymatic ^{hm}C glucosylation to specifically protect ^{hm}C-containing sites from digestion and thus distinguish them from ^mC sites.

In conclusion, we show that PvuRtsII is an ^{hm}C specific endonuclease and provide a biochemical characterization of its enzymatic properties for future applications as diagnostic tool in the analysis of ^{hm}C distribution at genomic loci in development and disease.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Ken Kreutzer (Duke University, Durham, NC) for providing the T4 phage stock and *E. coli* strains CR63 and ER1565 *galU* and Masaki Okano (RIKEN Center for Developmental Biology, Kobe, Japan) for the TKO ESCs.

FUNDING

Deutsche Forschungsgemeinschaft (DFG SPP 1356, SFB 646 and 684 to H.L.); Life Science Munich (LSM to S.B.). Financial support by the Elite Network of Bavaria (International Doctorate Program NanoBioTechnology) and International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS to C.S.S.). Funding for open access charge: Deutsche Forschungsgemeinschaft.

Conflict of interest statement. None declared.

REFERENCES

- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Rottach, A., Leonhardt, H. and Spada, F. (2009) DNA methylation-mediated epigenetic control. *J. Cell. Biochem.*, **108**, 43–51.
- Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, **324**, 929–930.
- 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.
- 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.
- Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J. and Fan, G. (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.*, **13**, 423–430.
- Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. and Leonhardt, H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **38**, e181.
- Münzel, M., Globisch, D., Bruckl, T., Wagner, M., Welzmler, V., Michalakis, S., Müller, M., Biel, M. and Carell, T. (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.*, **49**, 5375–5377.
- Globisch, D., Munzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Bruckl, T., Biel, M. and Carell, T. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*, **5**, e15367.
- 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.
- 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.
- Nestor, C., Ruzov, A., Meehan, R. and Dunican, D. (2010) Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. *Biotechniques*, **48**, 317–319.
- 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.* (2010) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.*, **29**, 68–72.
- Flaks, J.G. and Cohen, S.S. (1957) The enzymic synthesis of 5-hydroxymethyldeoxycytidylic acid. *Biochim. Biophys. Acta*, **25**, 667–668.
- Wiberg, J.S. and Buchanan, J.M. (1964) Studies on labile deoxycytidylate hydroxymethylases from *Escherichia coli* B infected with temperature-sensitive mutants of bacteriophage T4. *Proc. Natl Acad. Sci. USA*, **51**, 421–428.
- 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.
- Raleigh, E.A. (1992) Organization and function of the *mcrBC* genes of *Escherichia coli* K-12. *Mol. Microbiol.*, **6**, 1079–1086.
- 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.
- Bair, C.L. and Black, L.W. (2007) A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J. Mol. Biol.*, **366**, 768–778.
- Janosi, L., Yonemitsu, H., Hong, H. and Kaji, A. (1994) Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRtsII) modulated by glucosylation of DNA. *J. Mol. Biol.*, **242**, 45–61.
- Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R. *et al.* (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells*, **11**, 805–814.
- Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
- Hattori, N., Imao, Y., Nishino, K., Hattori, N., Ohgane, J., Yagi, S., Tanaka, S. and Shiota, K. (2007) Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes Cells*, **12**, 387–396.
- Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G. *et al.* (2011) Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, **8**, 200–213.
- Okano, M., Xie, S. and Li, E. (1998) Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.*, **26**, 2536–2540.

26. Yoder, J.A. and Bestor, T.H. (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum. Mol. Genet.*, **7**, 279–284.
27. Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E. and Bestor, T.H. (2006) Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science*, **311**, 395–398.
28. Rai, K., Chidester, S., Zavala, C.V., Manos, E.J., James, S.R., Karpf, A.R., Jones, D.A. and Cairns, B.R. (2007) Dnmt2 functions in the cytoplasm to promote liver, brain and retina development in zebrafish. *Genes Dev.*, **21**, 261–266.
29. Schaefer, M. and Lyko, F. (2010) Lack of evidence for DNA methylation of Invader4 retroelements in *Drosophila* and implications for Dnmt2-mediated epigenetic regulation. *Nat. Genet.*, **42**, 920–921.
30. Schaefer, M. and Lyko, F. (2010) Solving the Dnmt2 enigma. *Chromosoma*, **119**, 35–40.
31. Zemach, A., McDaniel, I.E., Silva, P. and Zilberman, D. (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, **328**, 916–919.
32. Gou, D., Rubalcava, M., Sauer, S., Mora-Bermúdez, F., Erdjument-Bromage, H., Tempst, P., Kremmer, E. and Sauer, F. (2010) SETDB1 is involved in postembryonic DNA methylation and gene silencing in *Drosophila*. *PLoS ONE*, **5**, e10581.
33. Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M.C. and Reuter, G. (2009) Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nat. Genet.*, **41**, 696–702.