Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation

Hideharu Hashimoto, Emory University
Yiwei Liu, Emory University
Anup K. Upadhyay, Emory University
Yanqi Chang, Emory University
Shelley B. Howerton, Emory University
Paula M Vertino, Emory University
Xing Zhang, Emory University
Xiaodong Cheng, Emory University

Journal Title: Nucleic Acids Research
Volume: Volume 40, Number 11
Publisher: Oxford University Press (OUP): Policy C - Option B | 2012-02-22,
Pages 4841-4849
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1093/nar/gks155
Permanent URL: http://pid.emory.edu/ark:/25593/dwb1c

Final published version: http://nar.oxfordjournals.org/content/40/11/4841

Copyright information:
© 2012 Hashimoto et al. Published by Oxford University Press
This is an Open Access work distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/).

Accessed January 20, 2019 6:39 PM EST
Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation

Hideharu Hashimoto¹, Yiwei Liu¹, Anup K. Upadhyay¹, Yanqi Chang¹, Shelley B. Howerton¹, Paula M. Vertino²,³, Xing Zhang¹ and Xiaodong Cheng¹,*

¹Department of Biochemistry, ²Department of Radiation Oncology, Emory University School of Medicine and ³The Winship Cancer Institute, Emory University, Atlanta, GA 30322, USA

Received October 14, 2011; Revised January 25, 2012; Accepted January 26, 2012

ABSTRACT

Cytosine residues in mammalian DNA occur in at least three forms, cytosine (C), 5-methylcytosine (M; 5mC) and 5-hydroxymethylcytosine (H; 5hmC). During semi-conservative DNA replication, hemi-methylated (M/C) and hemi-hydroxymethylated (H/C) CpG dinucleotides are transiently generated, where only the parental strand is modified and the daughter strand contains native cytosine. Here, we explore the role of DNA methyltransferases (DNMT) and ten eleven translocation (Tet) proteins in perpetuating these states after replication, and the molecular basis of their recognition by methyl-CpG-binding domain (MBD) proteins. Using recombinant proteins and modified double-stranded deoxyoligonucleotides, we show that DNMT1 prefers a hemi-methylated (M/C) substrate (by a factor of >60) over hemi-hydroxymethylated (H/C) and unmodified (C/C) sites, whereas both DNMT3A and DNMT3B have approximately equal activity on all three substrates (C/C, M/C and H/C). Binding of MBD proteins to methylated DNA inhibited Tet1 activity, suggesting that MBD binding may also play a role in regulating the levels of 5hmC. All five MBD proteins generally have reduced binding affinity for 5hmC relative to 5mC in the fully modified context (H/M versus M/M), though their relative abilities to distinguish the two varied considerably. We further show that the deamination product of 5hmC could be excised by thymine DNA glycosylase and MBD4 glycosylases regardless of context.

INTRODUCTION

5-hydroxymethylcytosine (5hmC) is a constituent of nuclear DNA, present in many tissues and cell types (1), but relatively enriched in embryonic stem cells (2) and Purkinje neurons (3). 5hmC in the mammalian genome depends on pre-existing 5-methylcytosine (5mC) (4). There are three mammalian ten eleven translocation (Tet) proteins that convert 5mC to 5hmC (2). In non-CpG context (CpA or CpT), 5mC is asymmetrical and 5hmC occurs largely DNA strand-specific (4). In the context of the palindromic CpG dinucleotide, 5hmC presumably exists as fully hydroxymethylated (H/H) form in cells and persists through cell division (5,6). However, the question remains how these H/H sites are maintained after semi-conservative DNA replication.

5hmC has also been proposed as a potential intermediate in active DNA demethylation via the base excision repair pathway (7,8). Methyl-CpG-binding domain 4 (MBD4) contains both an N-terminal MBD and a C-terminal thymine glycosylase domain that acts on G:T and G:U mismatches (9). Of particular interest is a recent report indicating that, in zebrafish, the activation-induced cytidine deaminase (AID) and MBD4 cooperate to demethylate DNA (10). Consistent with a role in DNA demethylation in mammals, AID is required to demethylate pluripotency genes during reprogramming of the somatic genome in embryonic stem cell fusions (11), and AID-deficient animals are less efficient in erasure of DNA methylation in primordial germ cells (12). It is noteworthy that AID promotes 5mC deamination, resulting in thymine (10,13), as well as 5hmC deamination, which would produce 5-hydroxymethyluracil (5hmU or hU) (8). Here, using in vitro biochemical methods, we investigated whether hydroxymethylation is maintained after DNA replication,
whether methyl-DNA binding proteins inhibit Tet1 activity, and whether the deamination product of hydroxymethylcytosine could be excised by the glycosylase activities of MBD4 and thymine DNA glycosylase (TDG).

MATERIALS AND METHODS

Protein expression and purification

The expression constructs used to generate recombinant proteins in this study are listed in Supplementary Table S1 and detailed purification schemes are provided in Supplementary Methods. All constructs (except hDNMT1 full length) were expressed in *Escherichia coli* BL21 (DE3) Codon-plus RIL (Stratagene) harboring the RIL-Codon plus plasmid. 6×His-SUMO-tagged and non-cleavable 6×His-tagged proteins were expressed using modified pET28a vectors pETHisSumo (14) or pET6H (15), respectively. Glutathione S-transferase-(GST-) tagged proteins were expressed in pGEX-2T (GE healthcare) or a modified pET21d with GST inserted between BamHI and EcoRI restriction sites.

In general, cells were cultured in Luria–Bertani (LB) medium (supplemented with 1 mM MgCl₂ and 1 mM ZnCl₂ and either 50 mg ml⁻¹ kanamycin or 100 mg ml⁻¹ ampicillin) at 37°C until OD₆₀₀ of ~0.5–1.0 before shifting the temperature down to 14–24°C. After 2–3 h, 0.2–0.4 mM isopropyl β-D-thiogalactoside was supplied to induce protein expression and cells were cultured overnight (~12–16 h). For MBD1, the induction time was 2 h. For UHRF1 (residues 124–628), expression cultures were grown overnight at 25°C in autoinduction medium (16). Cells were harvested and lysed by sonication (6 min total, 1 s on for 3 s off) for DNMT3A-C/DNMT3L complex were carried out at 25°C for 15 min. The dried circles were subjected to liquid-scintillation counting with Cytoscint scintillant. Each reaction was performed in duplicate.

To convert from counts-per-minute (cpm) to transferred methyl group concentration, we generated a linear calibration curve using 200 nM of HhaI methyltransferase to completely methylate 125, 250, 500 and 1000 nM of a 12-mer single CpG hemi-methylated DNA. After background subtraction, the resulting cpm’s were plotted against the DNA concentration.

DNA binding assay

Fluorescence polarization measurements were carried out at 25°C on a Synergy 4 Microplate Reader (BioTek). A 10 nM of 6-carboxy-fluorescein (FAM)-labeled double strand DNA [FAM-5'-CCATGCGCTGAC-3'/5'-GTCA GYGCATGG-3'] where X and Y are C, 5mC (M), 5hmC (H), T, U or 5hmU was incubated for 10 min with increasing amounts of proteins in binding buffer (20 mM Tris–HCl, pH 7.5, 150 mM NaCl, 5% glycerol and 0.5 mM TCEP). No change in fluorescence intensity was observed with the addition of protein. Curves were fit individually using GraphPad PRISM 5.0d software (GraphPad Software Inc.). Binding constants (Kᵦ) were estimated to be around 9 mg per 15 ml cell pellets.

*Pichia* expressed His₆-tagged DNMT1 was purified with Ni-column, HiTrap-SP and Sephacryl-300 in the buffer of 20 mM HEPES–NaOH, pH 7.0, 300 mM NaCl, 5% glycerol and 1 mM dithiothreitol (DTT) and further purified from associated nucleic acid through Q column, eluting at ~250 mM NaCl in the same buffer with 0.5 mM TCEP, instead of DTT.

Methyl transfer assays using oligonucleotides

Methyl transfer activity assays were performed in 50 mM Tris–HCl pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA) (for DNMT1) or in 50 mM Tris–HCl pH 7.5, 2.5% glycerol and 0.5 mM TCEP (for DNMT3A2/DNMT3L and DNMT3B2/DNMT3L). The reaction mixture (20 μl total volume) contained 5.5 μM [methyl-³²H] AdoMet (10.0 Ci/mmol; Perkin Elmer) and 1.0 μM oligonucleotides (see below), and enzymes (20 nM DNMT1 N-terminal deletions Δ600, Δ644 and Δ728, or 200 nM DNMT1 FL and Δ350 or 300 nM DNMT3a2/3L and DNMT3B2/3L). All enzymes were pre-incubated with AdoMet for 10 min at 37°C before the addition of DNA.

The reaction times for DNMT1 were 0–15 min, and terminated by the addition of 1.5 mM unlabelled S-adenosyl-l-homocysteine (AdoHcy). The reactions for DNMT3A2/DNMT3L, DNMT3B2/DNMT3L and DNMT3A-C/DNMT3L-C complex were carried out at 37°C for 0–60 min, and terminated by the addition of 1% sodium dodecyl sulfate and 1 mg ml⁻¹ of protease K and heated at 50°C for 15 min. The reaction mixtures were spotted on DE81 paper circles (Whatman), washed twice with 5 ml of cold 0.2 M NH₄HCO₃, twice with 5 ml of deionized water and once with 5 ml of ethanol. The dried circles were subjected to liquid-scintillation counting with Cytoscint scintillant. Each reaction was performed in duplicate.

Overexpression of human DNMT1 in *Pichia pastoris*

Human DNMT1 full-length (residues 1–1616) is overexpressed in *P. pastoris* using the Multicopy *Pichia* Expression pPIC3.5K (Invitrogen) modified to include a His-tag at the N-terminus of expressed proteins. An NdeI site was added so that the *dnmt1* gene can be inserted into the vector. A 6L-scale fermentation resulted in ~500 ml of the induced cell pellets, which were divided into 32 tubes. The amount of soluble full-length DNMT1 was estimated to be around 9 mg per 15 ml cell pellets.
calculated as $[mP] = \frac{[\text{maximum } mP]}{(K_D + [C]) + [\text{baseline } mP]}$, and saturated $[mP]$ was calculated as saturation = $(mP) - \frac{[\text{baseline } mP]}{(\text{maximum } mP) - [\text{baseline } mP]}$, where $[mP]$ is milli-Polarization and $[C]$ is protein concentration. Averaged $K_D$ and its standard error were reported.

DNA glycosylase activity assay

DNA glycosylase activity assay was performed similar to previously described except FAM labeled DNA oligonucleotides were used instead of radio-labeled DNAs (15). Purified MBD4 or TDG protein (0.5 μM) and 0.5 μM of double strand FAM labeled 32-mer annealed DNA (see below) were mixed in 20 μl nick buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, 0.1% BSA) and incubated at 37°C for 1 h. Reactions were stopped by adding 2 μl of 1 N NaOH, and boiled for 10 min before 20 μl of loading buffer (98% formamide, 1 M EDTA at 1 mg/ml of bromophenol blue and xylene cyanole) were added and boiled for another 10 min. Samples were immediately put into ice water to cool down and loaded on a 10 × 10 cm$^2$ denaturing PAGE gel containing 15% acrylamide, 7 M urea and 24% formamide in 1 × TBE buffer. The gels were run at 200 V for 60 min. FAM-labeled single strand DNA was visualized by UV exposure. The following oligonucleotides were synthesized at the New England Biolabs:

(FAM) -5′-TCGGATGGTGGGTCAGXXGCATGATAG
TGTA-3′
3′-AGCCCTACAACACCGAGTCGYCTACTATCACAT-5′

where X, Y = C, 5mC (M), 5hmC (H), U, T or 5hmU (hU).

RESULTS

DNMT3a and DNMT3b have comparable activities on C/C, M/C and H/C substrates

In mammals, DNA methyltransferases (DNMTs) include three members, in two families that are structurally and functionally distinct (17). The DNMT3A and DNMT3B (18,19), coupled with regulatory factor DNMT3-Like (DNMT3L) protein (20,21), establish the initial methylation pattern de novo, while DNMT1 and its accessory protein UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1) (22,23) maintain this pattern during chromosome replication. We first asked which Dnmts methylate the newly synthesized cytosine in the context of hemi-hydroxymethylated CpG site (H/C) (Figure 1). Using a 32-bp DNA oligonucleotide containing a single CpG site, either unmodified (C/C), hemi-methylated (M/C) or hemi-hydroxymethylated (H/C), we find that whereas the known maintenance methyltransferase DNMT1 (24) has high intrinsic activity for the M/C substrate ($K_{cat} = 6.6 \text{ h}^{-1}$; Figure 1b and Supplementary Figure S1), it has measurable but greatly reduced activity for the H/C substrate ($K_{cat} \approx 0.1 \text{ h}^{-1}$) and no detectable activity on C/C (Figure 1b). Thus, unlike M/C, H/C is not a preferred substrate of DNMT1 and is unlikely to be methylated by DNMT1 after replication, in agreement with previous findings (25). Consistent with this notion, UHRF1, which is essential for DNMT1 function and selectivity for hemi-methylated CpG (M/C) sites in vivo (22,23), loses its intrinsic preference for hemi-methylated DNA when 5mC is replaced by 5hmC. UHRF1 (residues 124–628) shows a >10-fold reduced binding affinity for H/C DNA as compared to M/C DNA, that was similar in magnitude to its affinity for fully methylated (M/M), fully hydroxymethylated (H/H) and H/M DNA (Figure 2a and Supplementary Discussion). Therefore, neither DNMT1 nor UHRF1 are likely to be involved in post-replicative maintenance of H/C methylation, which, without the involvement of DNMT3 (see below), would lead to ‘passive demethylation’ of 5hmC.

De novo DNMT3 family includes two active enzymes, DNMT3A and DNMT3B (18,19), and one regulatory factor, DNMT3L protein (20,21). DNMT3L enhances methylation by both DNMT3A and DNMT3B (26–30) via direct interaction with the catalytic domain of DNMT3A and DNMT3B and stabilizing their binding of AdoMet (31). We purified DNMT3A2 [a shorter isoform of DNMT3A (32)] in complex with DNMT3L as well as the complex between DNMT3BΔ218 (deletion of N-terminal 218 residues, termed 3B2, approximately equivalent to DNMT3A2 in size) and DNMT3L (Supplementary Figure S2). Unlike DNMT1, DNMT3A2/3L and DNMT3B2/3L have approximately equal activities on all three (M/C, H/C, C/C) substrates, with $K_{cat}$ values of 0.2–0.4 and ~0.2 h$^{-1}$, respectively (Figure 1c and d), suggesting that DNMT3A and DNMT3B do not distinguish various modifications in the hemi-modified context. The very slow catalytic turnover of DNMT3A and DNMT3B is probably due to the product binding of methylated DNA, considering that almost all of the cellular complement of DNMT3A and DNMT3B, but not DNMT1, is strongly anchored to nucleosomes containing methylated DNA (33,34).

Recognition of 5hmC by known MBDS

Tet proteins use 5mC as substrate to generate 5hmC (2). The amount of 5hmC, which inversely correlates with the amount of 5mC in Purkinje cell DNA and in granule cell DNA (3), could potentially be regulated by the combined cellular activities of Tet1-3 enzymes, DNMTs (through the regulation of the levels of 5mC substrate) and the methyl-specific DNA binding proteins (by binding and potentially masking 5mC), such as MeCP2, which is found in high concentrations in the brain (35). It was recently found that 5hmC levels in mouse cerebellum are negatively correlated with MeCP2 gene dosage (36). Consistent with this idea, we found that MeCP2 binding to methylated DNA inhibits Tet1 activity in vitro (Supplementary Figure S3).

To explore the effect of DNA hydroxylation on the function of MBD proteins, we measured the dissociation constants ($K_D$) between five MBD proteins and double stranded oligonucleotides containing a single CpG
dinucleotide with six different modification states (C/C, M/C, H/C, M/M, H/M and H/H as shown in Figure 1a) using fluorescence polarization analysis. MeCP2 (residues 77–205) binds fully methylated DNA (M/M) with a $K_D$ of 10 nM; changing a single methyl group to hydroxymethyl on one strand (H/M) increased the $K_D$ to 46 nM (an $C_{24}^5$-fold weaker binding) (Figure 2b). Full hydroxymethylation (H/H) weakened the binding by another factor of 5 with $K_D$ of 0.26 mM, a value similar to that of H/C (0.19 $\mu$M) and about half of C/C (0.5 $\mu$M) (Figure 2b), suggesting that the hydroxyl group of 5hmC disrupted the specific interaction between MeCP2 and 5mC, reducing the binding affinity to that of unmodified cytosine. In comparison, there was little apparent discrimination in the context of hemi-modification (0.19 $\mu$M for H/C versus 0.13 $\mu$M for M/C).

A similar trend was observed for the MBD domain of MBD1 (Figure 2c) and MBD2 (Figure 2d): each methyl group hydroxylation from M/M to H/M to H/H results in weaker binding (by a factor of 18 and 11 for MBD1 and 10 and 5 for MBD2, respectively). These results are in general agreement with the finding that MBD domains have a lower affinity toward sequences containing 5hmC (37,38). Nevertheless, it is noteworthy that the ability of the different MBDs to discriminate between these substrates (M/M versus H/M) varies considerably (from 2- to 18-fold), with MBD3 and MBD4 showing little discrimination (Figure 2e and f) (summarized in Figure 2g). Furthermore, although MeCP2 and MBD1 have significantly lower affinity for hemi-methylated and hemi-hydroxymethylated CpG dinucleotides (H/M) relative to M/M DNA (5- and 18-fold decreased affinity,
respectively), this substrate is still preferred over hemi-methylated DNA (M/C) by a factor of \( \frac{1}{24} \), and is bound by MeCP2 and MBD1 with an affinity similar to that which MBD2 binds fully methylated (M/M) DNA, and a greater affinity than MBD3 and MBD4, bind to fully methylated DNA (M/M) (summarized in Figure 2g). These data suggest that in vivo there may be additional factors (e.g. relative local concentration, tissue-specific expression) that determine whether, and to what degree, 5hmC DNA is bound by MBDS.

Excision of 5hmU by MBD4 and TDG

In the pathway shown in Figure 3a, Tet-mediated production of 5hmC becomes the substrate for AID, which converts 5hmC to 5hmU. We thus tested whether 5hmU can be excised by MBD4 and TDG glycosylase domains (15,39). We used the same 32-bp DNA duplexes containing a G:X mismatch within the CpG sequence context (where X = 5hmU, T or U) as the substrate. The X-containing strand was FAM labeled, and the excision...
of the mismatched base was monitored by denaturing gel electrophoresis following NaOH hydrolysis (Supplementary Figure S4). As expected, no glycosylase activity was observed on oligonucleotides bearing the ‘natural’ G:C and G:M base pairs, and there was efficient cleavage of substrates bearing G:T and G:U mismatches (Figure 3b and c). We extend these findings to show here that the TDG and MBD4 glycosylases are also inactive on G:H (which preserves Watson–Crick base-pair hydrogen bonds; see Figure 3a), but act on G:5hmU mismatched substrates, and further, that the modification status (methyl or hydroxymethyl) of the C in the opposite strand of neighboring G had no impact on the ability to remove G:T, G:U or G:5hmU mismatches (Figure 3b and c). Therefore, TDG and MBD4 are capable of acting on AID-generated 5hmU and completing the ‘demethylation’ of 5hmC.

**DISCUSSION**

Taken together, we suggest that DNMT3A and DNMT3B are capable of acting on the hemi-hydroxymethylated (H/C) CpG sites generated during DNA replication, to yield hemi-hydroxylated and hemi-methylated (H/M) CpG sites. Recently, a maintenance methylation function has been proposed for DNMT3A and DNMT3B (40), which have been suggested to complete the methylation of sites missed by or resistant to DNMT1 activity during DNA replication (e.g. repetitive DNA). Assuming H/M sites are substrates for modification by Tet proteins (Supplementary Figure S5), there is the potential for regenerating a fully modified hydroxymethylated H/H site, recapitulating the parental DNA state after replication (Figure 1a). It is interesting to note that while Dnmt3a (but not Dnmt3b or Dnmt1) expressed in both paternal and maternal pronuclei of mouse zygotes, Tet3 (but not Tet1 or Tet2) expressed specifically in the male
pronucleus (41). The expression patterns of Dnmt3a and Tet3 coincide with the loss of methylation and occurrence of 5hmC in the early mouse embryo (5,6).

We also showed that five MBD proteins have markedly varied abilities to discriminate M/M versus H/M substrates. MeCP2 and MBD1 bind H/M with an affinity similar to that which MBD2 binds fully methylated (M/M) DNA (Figure 2). These data suggest that MBD proteins (particularly MeCP2 and MBD1) may play a role in regulating the levels of 5hmC in vivo, by inhibiting Tet activities through their binding to M/M and H/M substrates. In mammals, all three Tet proteins, Tet1, Tet2 and Tet3 catalyze similar reactions, converting 5mC to 5hmC (42). One important future question will be to determine biochemically whether Tet proteins, like DNMTs, divide the labor among the family members, with one (or more) establishing the initial hydroxymethylation pattern de novo (i.e. using M/M as the substrate), and other(s) maintaining this pattern during chromosome replication (using H/M as the substrate). Finally, we demonstrated that MBD4 and TDG glycosylase domains are both active on AID-generated 5hmU and are thus capable completing the ‘active demethylation’ of 5hmC. An important future question will be to determine whether the two have largely complementary but context-dependent functions in 5hmC turnover. It would also be interesting to know whether Tet1-3 proteins interact with AID complex (including the damage response factor Gadd45 (10)) and TDG or MBD4 by some means.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online: Supplementary Table 1, Supplementary Figures 1–5, Supplementary Methods, Supplementary Discussion and Supplementary References [43–67].

ACKNOWLEDGEMENTS
We thank sincerely Brenda Baker of New England Biolabs for DNA oligo synthesis, Cris A. Lee of the Biochemistry, Cell and Developmental Biology program of Emory University for expression and purification of MeCP2 fragment, Dr Yi Zhang of University of North Carolina at Chapel Hill for providing initial Tet1 construct, Dr Robert M. Blumenthal of the University of Toledo College of Medicine for critical comments and editing the manuscript. H.H. performed purifications of DNMT1 N-terminal deletion fragments, purifications of MBD1 and MBD4, methylation assays, binding assays and glycosylation assays. Y.L. performed DNMT3 purifications and methylation assays. A.K.U. purified full-length DNMT1 and performed MBD inhibition of Tet1 activity. Y.C. purified MBD2 and MBD3; S.H.B. expressed full-length DNMT1 in P. pastoris; P.M.V. provided initial full-length human DNMT1 construct. X.Z. and X.C. organized and designed the scope of the study, and all were involved in analyzing data and helped in writing and revising the manuscript.

FUNDING
U.S. National Institutes of Health (GM049245-18 to X.C.) and NIH (grants CA077337 and CA132065 to P.M.V.). X.C. is a Georgia Research Alliance Eminent Scholar and P.M.V. is a Georgia Cancer Coalition Distinguished Cancer Scholar. The open access publication charge for this paper has been waived by Oxford University Press - NAR Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal.

Conflict of interest statement. None declared.

REFERENCES


