

Cytosine-5-methyltransferases add aldehydes to DNA

Zita Liutkevičiūtė¹, Gražvydas Lukinavičius^{1,3},
Viktoras Masevičius^{1,2}, Dalia Daujotytė^{1,3} &
Saulius Klimasauskas¹

Targeted methylation of cytosine residues by S-adenosylmethionine-dependent DNA methyltransferases modulates gene expression in vertebrates. Here we show that cytosine-5-methyltransferases catalyze reversible covalent addition of exogenous aliphatic aldehydes to their target residues in DNA, thus yielding corresponding 5- α -hydroxyalkylcytosines. Such atypical enzymatic reactions with non-cofactor-like substrates open new ways for sequence-specific derivatization of DNA and demonstrate enzymatic exchange of 5-hydroxymethyl groups on cytosine in support of an oxidative mechanism of DNA demethylation.

DNA methylation is an important biological mechanism that regulates gene expression in vertebrates and serves as a species 'self-code' in bacteria. Genomic methylation patterns are established by DNA methyltransferases (MTases), which catalyze targeted transfers of methyl groups from S-adenosyl-L-methionine (SAM, **1**) to adenine or cytosine residues. In higher eukaryotes, the cytosine-5 methylation leads to strong and heritable gene silencing¹. Along with DNA demethylation events, whereby 5-methylcytosines (mC, **2**) are converted back to cytosines², these covalent modifications at the 5 position of cytosine have vital roles in cellular differentiation, parental imprinting and silencing of endogenous retroviruses.

Enzymatic transmethylation generally proceed via a direct nucleophilic S_N2 attack of a target atom onto the sulfonium-bound methyl group of SAM^{3,4}. DNA cytosine-5-methyltransferases (C5-MTases)⁵ (Scheme 1a), RNA U5-MTases⁶, thymidylate synthases and 5-hydroxymethylcytidylate synthases⁷ all use a covalent mechanism for nucleophilic activation of their target pyrimidine residues. A model system for mechanistic studies of C5 methylation is the HhaI methyltransferase (M.HhaI), which methylates the GCGC target sites in DNA⁸. In M.HhaI and other C5-MTases, the activated cytosine intermediate (ACI) is not only active toward certain synthetic SAM analogs (reviewed in ref. 3) but also undergoes fast exchange of the 5-H proton with water⁵ and a slow hydrolytic deamination to afford uracil⁹ when cofactor is not present.

To explore the chemical reactivity of ACI, binary MTase-DNA complexes were prepared and screened against a series of electrophilic compounds such as aldehydes, ketones and electronegatively substituted vinyl derivatives. The DNA substrates contained one or two

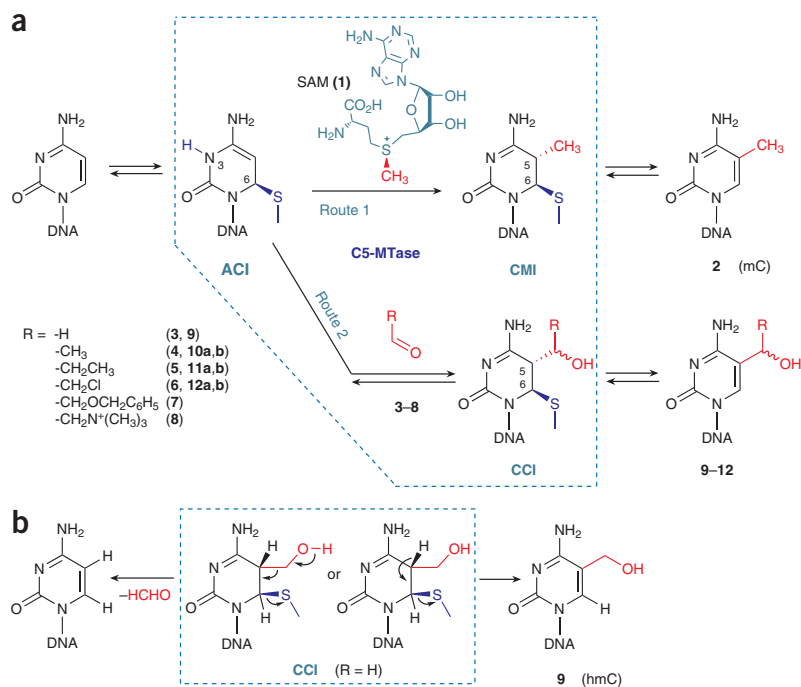
³³P-labeled cytosine nucleotides (see Supplementary Table 1 online) such that subsequent TLC analysis of labeled 5'-mononucleotides permitted a sensitive and selective detection of modifications in the target residue. In parallel, a total nucleoside composition analysis of similarly treated unlabeled DNA duplexes was performed by HPLC-MS. Notably, analyses of binary M.HhaI-DNA complexes treated with formaldehyde (**3**) showed a conversion of the target cytosine into a new compound (Fig. 1a,b). The structure of the new compound was established by analysis of UV and MS spectra and by direct chromatographic comparison with authentic 5-hydroxymethyl-2'-deoxycytidine (hmC, **9**) (Supplementary Table 2 and Supplementary Figs. 1 and 2 online). Additional support for the C5 coupling was obtained from (i) the absence of 5-H atoms⁵ in reaction products (Supplementary Fig. 3 online) and (ii) complete immunity of 5-methylated cytosine to the formaldehyde treatment (Fig. 1a). Therefore, we conclude that the M.HhaI-activated cytosine is converted to hmC. Notably, our control experiments showed no detectable formation of N-hydroxymethylated nucleosides (Supplementary Fig. 4 online), which are known to occur upon treatment of DNA with formaldehyde under harsher conditions¹⁰.

Similar experiments with acetaldehyde (**4**, Fig. 1a) and propionaldehyde (**5**, Supplementary Fig. 5 online) yielded C5 adducts **10** and **11**. Two isomers for each coupling product were apparent in HPLC-MS traces (Fig. 1b and Supplementary Fig. 6a online), which is consistent with the formation of an asymmetric center at the α -carbon. TLC analyses showed minor amounts of modification products upon treatment with benzoyloxyacetaldehyde (**7**) and betainic aldehyde (**8**) (Supplementary Fig. 5); however, corresponding products were not detectable with HPLC-MS. 2-Chloroacetaldehyde (CAA, **6**) also afforded modification of the target cytosine in the presence of M.HhaI, which is consistent with a recently reported DNA strand cleavage at CAA-modified sites upon piperidine treatment¹¹. HPLC-MS analysis of the modification product (Supplementary Fig. 6b) indicated that it corresponds to 5-(2-chloro-1-hydroxyethyl)-cytosine (**12**) (see Supplementary Table 2 for coupling yields and HR-MS spectra; see Supplementary Fig. 1 for UV spectra). Altogether, our results demonstrate that the addition reaction is general for short aliphatic aldehydes as a class. We found the other classes of electrophilic compounds to be inactive.

To assess the enzymatic generality of this reaction, a series of commercially available bacterial C5-MTases (M.HpaII, target site CCGG; M.SssI, CG; M.AluI, AGCT) were examined in a similar manner using appropriate DNA substrates¹² (see Supplementary Table 1). All of the interrogated MTases showed substantial catalytic activity with the aldehydes (Supplementary Fig. 5). The mouse Dnmt1 MTase also followed the trend with minor but detectable formation of the modified product in the presence of formaldehyde (Supplementary Fig. 5c). Despite variations in the coupling

¹Laboratory of Biological DNA Modification, Institute of Biotechnology, Vilnius, Lithuania. ²Department of Chemistry, Vilnius University, Vilnius, Lithuania. ³Present addresses: Ecole Polytechnique Fédérale de Lausanne, Institute of Chemical Sciences and Engineering, Lausanne, Switzerland (G.L.); MRC Laboratory of Molecular Biology, Cambridge, UK (D.D.). Correspondence should be addressed to S.K. (klimasau@ibt.lt).

Received 16 December 2008; accepted 3 April 2009; published online 10 May 2009; doi:10.1038/nchembio.172



Scheme 1 Transformations of a target cytosine catalyzed by DNA C5-MTases. **(a)** Biological methylations by C5-MTases occur via an S_N2 reaction between the activated cytosine intermediate (ACI) and cofactor SAM (**1**), yielding covalent methylated intermediate (CMI) and then 5-methylcytosine (mC, **2**) (route 1). The ACI can undergo nucleophilic addition reaction with an exogenous aldehyde (**3–8**) (route 2) to give a covalent coupled intermediate (CCI) with subsequent release of corresponding 5-(α -hydroxyalkyl)cytosine (**9–12**). This modification is reversed back to unmodified DNA by the enzyme in the absence of the exogenous aldehyde. Modifying reagents are shown in red, C5-MTase and its catalytic moieties are shown in blue, and boxed areas denote species and reactions within the catalytic center of the enzyme. **(b)** Proposed mechanism of decay of CCI (center). Depending on whether the exocyclic hydroxyl group or the ring 5 position is deprotonated, the release of enzyme yields either the unmodified cytosine and free formaldehyde **3** (left) or the coupling product **9** (right), respectively. Color coding as in **a**.

efficiency observed with individual enzymes, the identity of a modified nucleotide was fully dependent on the aldehyde used. On the other hand, the MTase-assisted coupling reactions occurred with high sequence specificity on both short DNA duplexes and large natural substrates (Fig. 1a,c).

Altogether, our findings reveal a previously unknown ability of C5-MTases to catalyze the addition of exogenous aliphatic aldehydes to the C5 position of the target cytosine, thereby yielding nucleobase derivatives with α -hydroxylated carbon chains (Scheme 1a). To our knowledge, this is the first demonstration of wild-type cofactor-dependent enzymes catalyzing an atypical chemical reaction using non-cofactor-like exogenous substrates¹³. The reactive aldehydes are not bona fide cofactors of SAM-dependent MTases because they lack an anchor moiety (such as adenosyl) that would assist in the formation of a discrete, specific complex with the enzyme. The chemical reaction (nucleophilic addition) itself is different from the S_N2 transfer (nucleophilic substitution) naturally catalyzed by MTases³. Normally, aldehydes attack exocyclic amino groups of nucleobases in DNA, which leads to *N*-hydroxymethyl derivatives and further disubstituted products^{14,15}. These reactions are largely responsible for the cytotoxicity of formaldehyde *in vivo*¹⁴, and are exploited for cross-linking of interacting proteins¹⁰ and mapping of unpaired nucleotides^{11,15} in DNA. The MTase-directed aldehyde coupling occurs under mild conditions and with high sequence and base specificity. The triple role of an MTase is thus to (i) recognize a specific DNA sequence, (ii) present a target nucleobase in the catalytic center and (iii) covalently activate the C5 position in the base. In crystal structures of the M.HhaI reaction complexes, the Watson-Crick edge of the ACI/CMI residues is obscured¹⁶, whereas the C5 position would be exposed to solvent in the absence of bound cofactor (Supplementary Fig. 7 online). These steric factors explain the switch in aldehyde regioselectivity in the presence of C5-MTases (from N4 to C5) and suggest that the stereochemistry of the methylation reaction will be followed (Scheme 1a). Consistent with this model, a steric enlargement of the cofactor binding pocket in M.HhaI (ref. 17) enhances the efficiency of the coupling reaction (Supplementary Fig. 8 online).

In genomic DNA, 5-hydroxymethylated pyrimidines (hmC and hmU) are known to arise via two major pathways. They are incorporated during DNA synthesis to replace the major bases in certain bacteriophages¹⁸. In higher eukaryotes, chemical¹⁹ and enzymatic^{18,20,21} oxidation of mC and thymine is thought to be the sole source of these bases; however, their roles in epigenetic regulation are currently unclear. The unveiled promiscuity of C5-MTases presents a third possible pathway for the formation of genomic 5-hydroxymethylpyrimidines, as formaldehyde occurs in millimolar concentrations in certain tissues of rats and humans²².

Besides the biological implications, a stable C-C bond generated in the MTase-directed reactions opens new ways for sequence-specific derivatization of DNA. The coupling reactions involving formaldehyde are simple, fast and robust, and are thus suitable for routine laboratory applications; modifications with longer aldehydes can be improved by steric engineering of the enzymes (see above). Although the α -hydroxyalkyl groups themselves are not sound reporters, they add a unique functionality to DNA (analogous to benzylic hydroxyl) that can be exploited for chemical or enzymatic derivatization. For example, a mild oxidation to formyl or keto groups would enable a further conjugation with compounds carrying hydrazine or hydroxylamine functions²³. Alternatively, hmC residues can be enzymatically glucosylated¹⁸, thereby permitting selective DNA labeling through application of glycan modification/recognition techniques²⁴. On the other hand, a wide diversity of known DNA C5-MTases¹² and RNA pyrimidine-5-MTases⁶ offers an immense selection of potential labeling targets. Importantly, because the MTase-directed aldehyde coupling is blocked by prior 5-methylation of the same residue (Fig. 1a), the above-mentioned schemes can be used to query the methylation status of CpG sites in mammalian genomic DNA.

We also examined whether C5-MTases can promote the reverse reaction—the removal of formaldehyde from hmC. For this, a DNA duplex that contained enzymatically produced hmC residues at the target position was again treated with the same MTase in the absence of the aldehyde. The amount of hmC was substantially reduced after such treatment, whereas the amount of cytosine was increased (Fig. 1d,e), and the cleavage of a previously resistant DNA fragment by a cognate

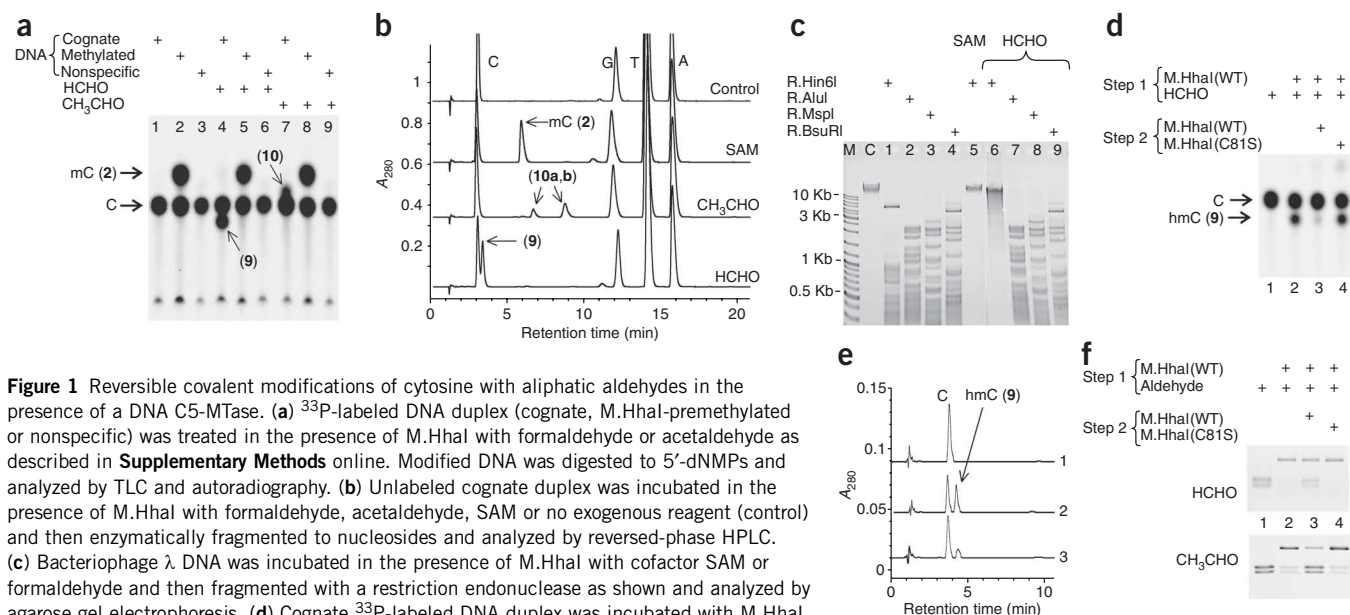


Figure 1 Reversible covalent modifications of cytosine with aliphatic aldehydes in the presence of a DNA C5-MTase. **(a)** ³³P-labeled DNA duplex (cognate, M.HhaI-premethylated or nonspecific) was treated in the presence of M.HhaI with formaldehyde or acetaldehyde as described in **Supplementary Methods** online. Modified DNA was digested to 5'-dNMPs and analyzed by TLC and autoradiography. **(b)** Unlabeled cognate duplex was incubated in the presence of M.HhaI with formaldehyde, acetaldehyde, SAM or no exogenous reagent (control) and then enzymatically fragmented to nucleosides and analyzed by reversed-phase HPLC. **(c)** Bacteriophage λ DNA was incubated in the presence of M.HhaI with cofactor SAM or formaldehyde and then fragmented with a restriction endonuclease as shown and analyzed by agarose gel electrophoresis. **(d)** Cognate ³³P-labeled DNA duplex was incubated with M.HhaI and formaldehyde (step 1) and then treated with M.HhaI (WT or C81S mutant) alone (step 2) and analyzed as in **a**. **(e)** Cognate DNA duplex was incubated with M.HhaI and formaldehyde and then treated with M.HhaI alone and analyzed as in **b** (trace assignments as in **d**). **(f)** pUC-618 DNA (a PCR fragment containing a single GCGC site) was incubated with M.HhaI and with formaldehyde or acetaldehyde (step 1). Modified DNA was then separately incubated with M.HhaI (step 2), fragmented with R.Hin6I and analyzed as in **c**.

restriction endonuclease was largely restored (**Fig. 1f**). Similar results were obtained with acetaldehyde-modified DNA (**Fig. 1f**). These observations demonstrate that C5-MTases catalyze the removal of the coupled aldehyde, thus restoring the original cytosine residue in DNA.

The MTase-assisted removal of the C5-bound hydroxymethyl group is noteworthy in light of the fact that the enzymatic methyl transfer is irreversible^{4,5}. The reverse reaction also requires covalent catalysis (**Fig. 1d,f**), which suggests that it proceeds via a covalently coupled intermediate (CCI, **Scheme 1a**). A mechanism for this reaction (**Scheme 1b**) can be derived from analogy with the light- and alkali-induced conversions of hmC to cytosine, which occur via corresponding 5,6-dihydro-6-hydroxy derivatives¹⁹. The biological importance of dehydroxymethylation reactions is bolstered by their relevance to the mechanisms of DNA demethylation and repair. Certain DNA repair enzymes can reverse alkylation damage via enzymatic oxidation of N-alkylated nucleobases to corresponding α -hydroxyalkyl derivatives, which spontaneously release an aldehyde from the ring nitrogen to generate the unmodified base²⁵. The enzymology of genomic mC demethylation remains elusive and highly debated²; among others, a similar two-step route has been proposed based on the above examples^{2,20,21} and preliminary observations²⁶. Our findings (**Fig. 1d–f**) offer a direct demonstration of an enzymatic hmC-to-cytosine conversion, thus providing a plausible chemical precedent for an oxidative mechanism of DNA demethylation.

Note: Supplementary information and chemical compound information is available on the Nature Chemical Biology website.

ACKNOWLEDGMENTS

We thank S. Morera (Laboratoire d'Enzymologie et Biochimie Structurales) and L.C. Sowers (Loma Linda University) for the gift of an hmC-containing oligonucleotide, G. Vilkaitis (Institute of Biotechnology) for a sample of mouse Dnmt1, and M. Petrušytė (Fermentas) for M.HpaII. Thanks are due to L. Peil for HR-MS analysis of modified nucleosides, M. Krenevičienė for NMR spectra and E. Weinhold for valuable discussions. This work was supported by grants

from the Lithuanian State Science and Study Foundation and the Ministry of Education and Science of Lithuania.

Published online at <http://www.nature.com/naturechemicalbiology/>
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Goll, M.G. & Bestor, T.H. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
- Morgan, H.D., Santos, F., Green, K., Dean, W. & Reik, W. *Hum. Mol. Genet.* **14**, R47–R58 (2005).
- Klimasauskas, S. & Weinhold, E. *Trends Biotechnol.* **25**, 99–104 (2007).
- Klimasauskas, S. & Lukinavicius, G. in *Wiley Encyclopedia of Chemical Biology* (ed. Begley, T.P.) Ch. 335 (John Wiley & Sons, New York, 2008).
- Wu, J.C. & Santi, D.V. *J. Biol. Chem.* **262**, 4778–4786 (1987).
- Bujnicki, J.M., Feder, M., Ayres, C.L. & Redman, K.L. *Nucleic Acids Res.* **32**, 2453–2463 (2004).
- Graves, K.L., Butler, M.M. & Hardy, L.W. *Biochemistry* **31**, 10315–10321 (1992).
- Cheng, X. & Roberts, R.J. *Nucleic Acids Res.* **29**, 3784–3795 (2001).
- Zingg, J.-M., Shen, J.-C., Yang, A.S., Rapoport, H. & Jones, P.A. *Nucleic Acids Res.* **24**, 3267–3275 (1996).
- Solomon, M.J. & Varshavsky, A. *Proc. Natl. Acad. Sci. USA* **82**, 6470–6474 (1985).
- Daujotyte, D., Liutkeviciute, Z., Tamulaitis, G. & Klimasauskas, S. *Nucleic Acids Res.* **36**, e57 (2008).
- Roberts, R.J., Vincze, T., Posfai, J. & Macelis, D. *Nucleic Acids Res.* **35**, D269–D270 (2007).
- Bornscheuer, U.T. & Kazlauskas, R.J. *Angew. Chem. Int. Ed.* **43**, 6032–6040 (2004).
- Wang, M., Cheng, G., Villalta, P.W. & Hecht, S.S. *Chem. Res. Toxicol.* **20**, 1141–1148 (2007).
- Kusmieriek, J.T. & Singer, B. *Biochemistry* **21**, 5717–5722 (1982).
- Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. *Cell* **76**, 357–369 (1994).
- Lukinavicius, G. *et al. J. Am. Chem. Soc.* **129**, 2758–2759 (2007).
- Gommers-Ampt, J.H. & Borst, P. *FASEB J.* **9**, 1034–1042 (1995).
- Privat, E. & Sowers, L.C. *Chem. Res. Toxicol.* **9**, 745–750 (1996).
- Tahiliani, M. *et al. Science* published online, doi:10.1126/science.1170116 (16 April 2009).
- Kriaucionis, S. & Heintz, N. *Science* published online, doi:10.1126/science.1169786 (16 April 2009).
- Heck, H. & Casanova, M. *Regul. Toxicol. Pharmacol.* **40**, 92–106 (2004).
- Prescher, J.A. & Bertozzi, C.R. *Nat. Chem. Biol.* **1**, 13–21 (2005).
- Chiattolina, S., Xie, F. & Wang, Q. *Tetrahedron Lett.* **46**, 2331–2336 (2005).
- Sedgwick, B., Bates, P.A., Paik, J., Jacobs, S.C. & Lindahl, T. *DNA Repair (Amst.)* **6**, 429–442 (2007).
- Hamm, S. *et al. Biorg. Med. Chem. Lett.* **18**, 1046–1049 (2008).