JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Base-Resolution Analysis of 5-Hydroxymethylcytosine by One-Pot Bisulfite-Free Chemical Conversion with Peroxotungstate

Gosuke Hayashi,[†] Kenta Koyama,[†] Hidefumi Shiota,[†] Asuka Kamio,[‡] Takayoshi Umeda,[‡] Genta Nagae,[‡] Hiroyuki Aburatani,[‡] and Akimitsu Okamoto*[,†,‡]

[†]Department of Chemistry and Biotechnology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

[‡]Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

Supporting Information

ABSTRACT: 5-Hydroxymethylcytosine (^{hm}C) is an essential intermediate in the active DNA demethylation pathway. Here we report a new base-resolution method for measuring ^{hm}C by combining peroxotungstate-mediated oxidation and sequencing analysis. We reveal that an oxidized product of ^{hm}C, trihydroxylated thymine (thT), tolerated the incorporation of dATP as a substrate in the process of DNA polymerase elongation. By comparing the results of Sanger sequencing before and after the oxidation, we observed that ^{hm}C sites on single-stranded DNAs could be discriminated from unmethylated cytosines. We found that a thermal cycle condition during peroxotungstate treatment enhanced the oxidation reaction of ^{hm}C in double-stranded DNA. Furthermore, Illumina sequencing analysis of ^{hm}C-containing synthetic genome fragments enabled us to identify simultaneously the positions of ^{hm}C in base resolution. This bisulfite-free simple ^{hm}C detection technique could facilitate the acquisition of epigenomic information.

To regulate the degree of cellular DNA methylation, which defines the regions and levels of gene expression, many multicellular organisms have an active demethylation pathway, where TET family proteins oxidize 5-methylcytosine (^{m}C) to afford an oxidized ^{m}C analogue, 5-hydroxymethylcytosine (^{hm}C), as an intermediate of demethylation.[1] ^{hm}C has been discovered in human and mouse embryonic stem cells[2−6] and brain tissues[7−10] and seems to regulate gene expression through avoiding binding of ^{m}C binding proteins or recruiting ^{hm}C-specific binding proteins.[11,12] As ^{hm}C is involved not only in essential biological processes such as development and cellular reprogramming but also in diseases such as cancer,[13] several research groups have developed useful methods for ^{hm}C analysis.[14]

Base-resolution analysis of ^{hm}C, in which an ^{hm}C-selective chemical reaction is utilized in combination with bisulfite sequencing (BS-Seq), is currently available. Oxidative bisulfite sequencing (oxBS-Seq) uses potassium perruthenate ($KRuO_4$) to oxidize ^{hm}C to 5-formylcytosine, which is deaminated after bisulfite treatment and therefore sequenced as thymine (T).[15] As a result, by comparing the sequence data from standard BS-Seq and oxBS-Seq, ^{hm}C can be discriminated from unmodified cytosine (C) and ^{m}C. In TET-assisted bisulfite sequencing (TAB-Seq), ^{hm}C is first modified with glucose by β-glucosyltransferase to protect ^{hm}C from subsequent TET-mediated oxidation, in which ^{m}C is converted to 5-carboxycytosine (^{ca}C).[16] Because ^{ca}C undergoes bisulfite-mediated deamination, and is sequenced as T, only unoxidized ^{hm}C can be detected as C by comparing sequence data before and after TET/bisulfite treatment. It is notable that both of the ^{hm}C base-resolution analysis methods include multistep reactions and depend on bisulfite-mediated conversion, which often causes depyrimidination-mediated DNA damages such as strand breaks.[17]

We have developed the peroxotungstate-mediated ^{hm}C-selective oxidation reaction. The C5−C6 carbon of the allyl alcohol moiety in ^{hm}C is oxidized to form an epoxide, and a subsequent nucleophilic attack of water generates the trihydroxylated and 4-deaminated product, trihydroxylated thymine (thT) (Figure 1A).[18] As the oxidized DNA strand is easily cleaved at the thT nucleotide by hot piperidine treatment, ^{hm}C-containing single-stranded DNA was identified in PAGE analysis after peroxotungstate and piperidine treatment. However, this method has limitations, such as a need for labeled target DNA and a lack of sequence information. We

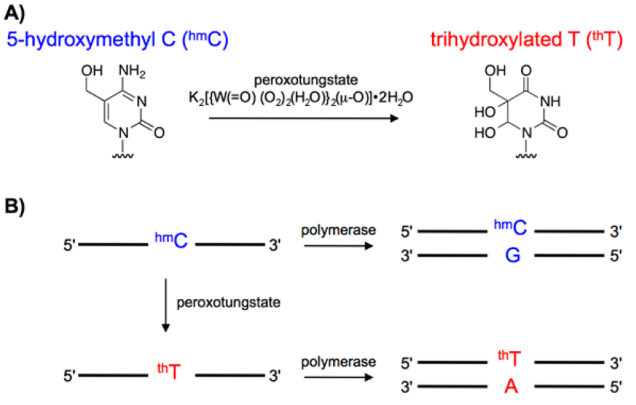

Figure 1. ^{hm}C detection strategy in this study. (A) Peroxotungstate-mediated conversion of 5-hydroxymethyl cytosine (^{hm}C) to trihydroxylated thymine (thT). (B) Base-resolution analysis of ^{hm}C by comparing the sequence before and after the peroxotungstate reaction.

Received: June 22, 2016
Published: October 21, 2016

DOI: 10.1021/jacs.6b06428
J. Am. Chem. Soc. 2016, 138, 14178−14181

hypothesized that thT, which possesses one hydrogen bond donor and two acceptors, as in the case of native T, may undergo base pairing with adenine (A) and therefore incorporate A against it in the polymerase extension reaction (Figure 1B). Here, we present a bisulfite-free ^{hm}C base-resolution analysis using the peroxotungstate-mediated one-pot oxidation and subsequent DNA sequencing.

The oxidation conditions with peroxotungstate against an ^{hm}C-containing oligonucleotide were optimized using HPLC analysis. A ^{hm}C-containing trinucleotide 5′-A^{hm}CA-3′ was treated at 50 °C in the presence of 5 mM peroxotungstate. After 1 h of incubation, the peak of the starting trinucleotide decreased by approximately 60%, and two new peaks appeared (Figure S1A). MALDI-TOF MS revealed that the molecular weights of both new peaks were identical to 5′-AthTA-3′, suggesting that the diastereoisomers of the thT were produced in similar yields. Time-course HPLC analysis showed that more than 99% of ^{hm}C reacted to generate thT in the trinucleotide after 7 h (Figures S1 and S2). On the other hand, trinucleotides containing ^{m}C or C instead of ^{hm}C showed no change in HPLC spectra after 5 h oxidation (Figure S1C,D).

The nucleotide incorporated into a position opposite to thT by DNA polymerase was determined by primer extension assays conducted with ^{hm}C-containing template oligonucleotide (hmODN2) and fluorescein-labeled primer (Table S1). After peroxotungstate oxidation of hmODN2, thT-containing oxidized hmODN2 was isolated and identified by HPLC and MALDI-TOF MS, respectively (Figure S2). The template–primer duplex was incubated with dATP or dGTP in the presence of KOD Dash polymerase at 68 °C for 2 min. Denaturing PAGE analysis revealed that the incorporation of dATP increased in a dose-dependent manner, whereas dGTP showed similar tendency with slower incorporation rate (Figure S3A). On the other hand, dCTP and dTTP did not show any product bands, even in the presence of 300 μM triphosphate on nonpurified oxidized hmODN2 (Figure S3B). To estimate the incorporation efficiency of dATP and dGTP, the steady-state kinetic parameters, V_{max} and K_m, of the primer extension reaction were derived from a Lineweaver–Burk plot created using the intensities of the gel bands (Figure S3C).[19] The efficiency (V_{max}/K_m) of dATP incorporation was more than 2 times higher than that of dGTP, mainly due to the higher V_{max} and lower K_m values (Table 1). The unexpected incorporation

Table 1. Steady-State Kinetic Parameters Derived from Single dNTPs Primer Extension Assay and Lineweaver–Burk Plot

	V_{max} (pmol U^{-1} min^{-1})	K_m (pmol)	rfficiency (V_{max}/K_m)
dATP	35.0	10.6	3.30
dGTP	20.4	14.1	1.44

of dGTP by KOD Dash polymerase might be caused by lactam–lactim tautomerization at 3–4 position of thT, which alters hydrogen-bonding pattern to match guanine base. dATP was not incorporated into the unoxidized template (Figure S3D), suggesting that dATP incorporation depends on formation of thT.

Base-resolution analysis of oxidized hm-ssDNA1 (Table S1) was performed using Sanger sequencing. The sequencing result after 1 h of incubation of 107-mer hm-ssDNA1 with peroxotungstate showed increased incorporation of A at the ^{hm}C position, although A incorporation was not observed without peroxotungstate treatment (Figure 2A). The extended period of the oxidative treatment increased the rate of A

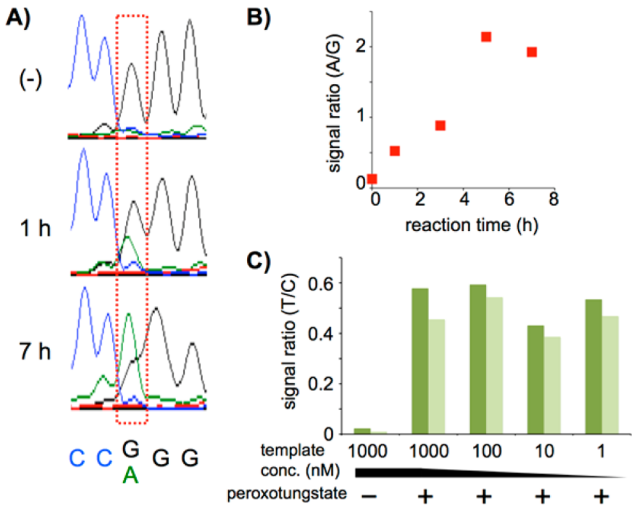

Figure 2. Sanger sequencing analysis of peroxotungstate-treated ssDNAs. (A) Electropherograms of sequencing data before and after 1 or 7 h oxidation of hm-ssDNA1. The base positions originally from the opposite of ^{hm}C are surrounded by red dotted lines. (B) Oxidation-time-dependent change of A/G signal ratio. (C) Template concentration dependence of T/C signal ratio of PCR-amplified samples derived from hm-ssDNA2. The ratio values derived from the 5′- and 3′-side ^{hm}C positions are shown in dark and light green, respectively.

incorporation (Figure 2B). Given that the results of primer extension experiments described above indicated that the incorporation efficiency of dATP was approximately 2 times higher than that of dGTP, the G/A conversion ratio shown in Figure 2B was unexpectedly high. However, it is possible that the specific reaction conditions in Sanger sequencing, which include a different polymerase, salt concentration, and temperature, enabled the incorporation of dATP more efficiently.

To investigate whether our technique is applicable to a low copies of DNA sample such as genomic DNA, the peroxotungstate-treated ssDNA was PCR amplified before Sanger sequencing. After a 7 h oxidation reaction of 75-mer ssDNA (hm-ssDNA2), which contains two ^{hm}C in its sequence, the oxidized DNA was serially diluted to 1000, 100, 10, and 1 nM as a PCR reaction mixture containing KOD Dash polymerase. The sequencing results of each PCR product showed increased T signals at the ^{hm}C positions, irrespective of the diluted concentration. However, no signal gain was observed in the 1000 nM hm-ssDNA2 without oxidation (Figure S4). The signal ratios of T/C of the oxidized samples were clearly higher than those of the unoxidized sample (Figure 2C), indicating that this one-step oxidation reaction offers a new base-resolution analysis of ^{hm}C.

The reaction conditions of ^{hm}C oxidation were optimized to apply this sequencing method to ^{hm}C-containing dsDNA, which is generally more resistant to environmental stimuli than ssDNA. Instead of using isothermal incubation at 50 °C for ssDNA treatment, we used thermal cycle conditions of 90 and 60 °C, which could denature dsDNA molecules and facilitate the oxidation reaction. A dsDNA composed of fluorescein-labeled ^{hm}C-containing oligonucleotide (hmODN3) and its complementary sequence (chmODN3) was treated under the thermal cycle conditions. It was then treated with hot

DOI: 10.1021/jacs.6b06428
J. Am. Chem. Soc. 2016, 138, 14178–14181

piperidine, which cleaves a DNA strand at the thT nucleoside. Denaturing PAGE analysis revealed that the thermal cycle conditions enabled the formation of cleaved DNA (Figure 3),

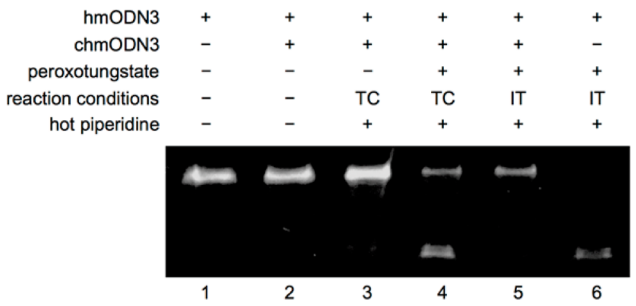

Figure 3. Optimization of oxidation conditions for ^{hm}C-containing dsDNA analyzed by denaturing PAGE analysis. Conditions of peroxotungstate treatment were either TC (thermal cycle conditions: 90 °C/1 min and 60 °C/10 min) or IT (isothermal conditions: 50 °C/5 h).

suggesting that the oxidation of ^{hm}C in dsDNA proceeded. On the other hand, any cleaved bands were not observed under the isothermal conditions, in which ^{hm}C in ssDNA is easily oxidized. We reasoned that the thermal cyclic conditions are required to oxidize ^{hm}C in dsDNA.

We also investigated simultaneously analyzing multiple ^{hm}C-containing dsDNAs using an Illumina sequencing system. We prepared a 150 bp synthetic dsDNA library consisting of four different ^{hm}C-containing dsDNAs whose sequences were extracted from the human reference genome (Table S1). The DNA library was incubated under three different thermal cycle conditions of 90 and 60 °C (10, 20, and 30 cycles) in either the absence or presence of peroxotungstate, and ligated with adapter sequences at both ends for Illumina sequencing. It is notable that more than 90 and 70% of dsDNAs remained even after 10- and 30-cycles of oxidation (Figure S5), suggesting that DNA damage is not problematic in the peroxotungstate-mediated oxidation. The sequencing results showed more than 2000 reads in each DNA (Table S2). At all ^{hm}C positions in DNA1, 2, 3, and 4, the ratios of T reads among total reads for peroxotungstate-treated samples were significantly higher than those for untreated samples (Figure 4). We compared the ratios

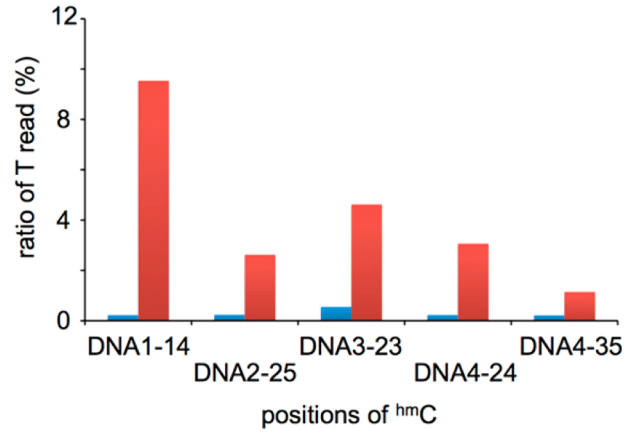

Figure 4. Simultaneous detection of multiple ^{hm}C by Illumina sequencing. The ratio of T reads over total reads derived from the oxidized and unoxidized DNA library is shown in red and blue, respectively.

of T reads at other C positions among terminal 50 nt in DNA1, 2, 3, and 4 (Figure S6). A significant difference in the ratios between peroxotungstate-treated and untreated samples was only observed at ^{hm}C positions, suggesting that multiple ^{hm}Cs included in the dsDNA library can be detected at base-resolution when using the Illumina sequencing method.

Finally, the tandem oxidation-sequencing analysis was applied to human brain genomic DNA, in which ^{hm}C occurs at relatively high levels.[20] Genomic DNA was treated under 10 cycles of thermal cycle oxidation, and then used as a template of strand-specific PCR (SS-PCR), which could increase sensitivity of ^{hm}C detection by ignoring nontarget strand.[21] We sequenced SS-PCR product amplified from ^{hm}C-enriched region on chromosome 2, and observed a distinguishable Sanger sequencing signal (Figure S7A). The ratio of T/C signal at the position was significantly higher than flanking C sites (Figure S7B). It is notable that the presence of ^{m}C or ^{hm}C at the position was confirmed by conventional bisulfite sequencing (Figure S7A), suggesting that this site is potential target of TET oxidation. These results suggest that this bisulfite-free ^{hm}C-to-thT conversion is applicable to genomic DNA analysis.

In this study, we demonstrated that one-step chemical conversion of ^{hm}C to thT leads to a new base-resolution analysis of 5-hydroxymethylcytosine. The polymerase incorporation of dATP at the opposite position of thT was confirmed by primer extension analysis and Sanger sequencing. After optimization of peroxotungstate treatment for dsDNA oxidation, Illumina sequencing analysis clearly distinguished the positions of ^{hm}C from a dsDNA library. Finally, we were able to detect ^{hm}C signal from human brain genomic DNA. This bisulfite-free sequencing method could provide a simple alternative for genome-wide mapping of the epigenetically important DNA modification.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications Web site at DOI 10.1021/jacs.6b06428.

Experimental methods, HPLC traces, primer extension assays, Sanger sequencing results, Illumina sequencing results, and DNA sequences (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*okamoto@chembio.t.u-tokyo.ac.jp

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (A) 15H02190.

■ REFERENCES

(1) Shen, L.; Zhang, Y. *Curr. Opin. Cell Biol.* **2013**, *25*, 289.
(2) Ficz, G.; Branco, M. R.; Seisenberger, S.; Santos, F.; Krueger, F.; Hore, T. A.; Marques, C. J.; Andrews, S.; Reik, W. *Nature* **2011**, *473*, 398.
(3) Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. *Nature* **2010**, *466*, 1129.
(4) Stroud, H.; Feng, S. H.; Kinney, S. M.; Pradhan, S.; Jacobsen, S. E. *Genome Biol.* **2011**, *12*, R54.

DOI: 10.1021/jacs.6b06428
J. Am. Chem. Soc. 2016, 138, 14178−14181

(5) Szulwach, K. E.; Li, X. K.; Li, Y. J.; Song, C. X.; Han, J. W.; Kim, S.; Namburi, S.; Hermetz, K.; Kim, J. J.; Rudd, M. K.; Yoon, Y. S.; Ren, B.; He, C.; Jin, P. *PLoS Genet.* **2011**, *7*, e1002154.
(6) Tahiliani, M.; Koh, K. P.; Shen, Y. H.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
(7) Jin, S. G.; Wu, X. W.; Li, A. X.; Pfeifer, G. P. *Nucleic Acids Res.* **2011**, *39*, 5015.
(8) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
(9) Szulwach, K. E.; Li, X. K.; Li, Y. J.; Song, C. X.; Wu, H.; Dai, Q.; Irier, H.; Upadhyay, A. K.; Gearing, M.; Levey, A. I.; Vasanthakumar, A.; Godley, L. A.; Chang, Q.; Cheng, X. D.; He, C.; Jin, P. *Nat. Neurosci.* **2011**, *14*, 1607.
(10) Wang, T.; Pan, Q.; Lin, L.; Szulwach, K. E.; Song, C. X.; He, C.; Wu, H.; Warren, S. T.; Jin, P.; Duan, R. H.; Li, X. K. *Hum. Mol. Genet.* **2012**, *21*, 5500.
(11) Valinluck, V.; Tsai, H. H.; Rogstad, D. K.; Burdzy, A.; Bird, A.; Sowers, L. C. *Nucleic Acids Res.* **2004**, *32*, 4100.
(12) Yildirim, O.; Li, R. W.; Hung, J. H.; Chen, P. B.; Dong, X. J.; Ee, L. S.; Weng, Z. P.; Rando, O. J.; Fazzio, T. G. *Cell* **2011**, *147*, 1498.
(13) Wu, H.; Zhang, Y. *Cell* **2014**, *156*, 45.
(14) Song, C. X.; Yi, C. Q.; He, C. *Nat. Biotechnol.* **2012**, *30*, 1107.
(15) Booth, M. J.; Branco, M. R.; Ficz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S. *Science* **2012**, *336*, 934.
(16) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C. X.; Zhang, L.; Kim, A.; Li, X. K.; Dai, Q.; Shen, Y.; Park, B.; Min, J. H.; Jin, P.; Ren, B.; He, C. *Cell* **2012**, *149*, 1368.
(17) Tanaka, K.; Okamoto, A. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1912.
(18) Okamoto, A.; Sugizaki, K.; Nakamura, A.; Yanagisawa, H.; Ikeda, S. *Chem. Commun.* **2011**, *47*, 11231.
(19) Okamoto, A.; Tanaka, K.; Saito, I. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 97.
(20) Wen, L.; Li, X.; Yan, L.; Tan, Y.; Li, R.; Zhao, Y.; Wang, Y.; Xie, J.; Zhang, Y.; Song, C.; Yu, M.; Liu, X.; Zhu, P.; Li, X.; Hou, Y.; Guo, H.; Wu, X.; He, C.; Li, R.; Tang, F.; Qiao, J. *Genome Biol.* **2014**, *15*, R49.
(21) An, J.; Yang, T.; Huang, Y.; Liu, F.; Sun, J.; Wang, Y.; Xu, Q.; Wu, D.; Zhou, P. *BMC Biochem.* **2011**, *12*, 2.

DOI: 10.1021/jacs.6b06428
J. Am. Chem. Soc. 2016, 138, 14178−14181