ChemPubSoc Europe

DOI: 10.1002/cbic.201300342

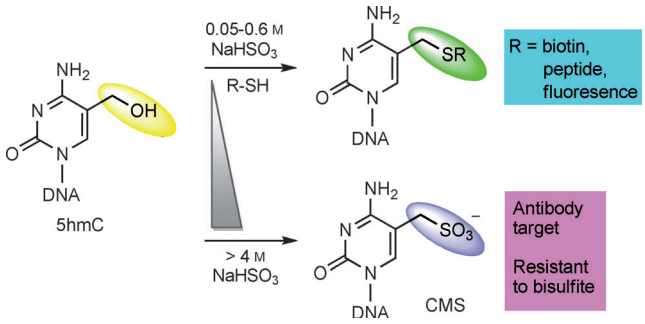
Nonenzymatic Labeling of 5-Hydroxymethylcytosine in Nanopore Sequencing

Xingyu Lu and Chuan He*[a]

The methylation of cytosine is one of the most important epigenetic modifications, and 5-methylcytosine (5mC) is widely accepted as the fifth base in DNA.[1] 5mC encodes crucial epigenetic information and has a significant impact on gene regulation, mammalian development, and human diseases.[1] Since 2009, the rediscovery of a new cytosine variant, 5-hydroxymethylcytosine (5hmC), has attracted a lot of interest from the chemistry and biology communities.[2,3] The Tet family proteins were found to generate 5hmC through the oxidation of 5mC.[3] Later on it was discovered that continued oxidation of 5hmC by Tet enzymes forms 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC).[3–6] A thymine DNA glycosylase (TDG)-mediated base-excision-repair (BER) process was shown to remove 5fC and 5caC in an active DNA demethylation pathway.[6]

The first challenge in understanding the potential function(s) of the new modifications is to detect and map their distributions in genomic DNA. Through a combination of high-throughput sequencing and affinity-based enrichment, genome-wide 5hmC profiling methods have been rapidly developed in order to reveal its potential roles in gene regulation. Covalent labeling methods have been developed to achieve efficient and unbiased modification and enrichment of 5hmC. For example, β-glucosyltransferase (βGT) can catalyze the glucosylation of 5hmC, and this can be followed by a second chemical modification to biotinylate 5hmC.[7–9] Antibodies or binding protein have also been generated as the most straightforward way to specifically bind and enrich 5hmC or its derivatives, such as cytosine 5-methylenesulfonate (CMS) and glucosylated 5hmC.[7,10–17] DNA methyltransferase has also been employed to direct thiol and selenol substitution of 5hmC for potential labeling.[18] Now a new nonenzymatic labeling method has been developed by the group of Bayley.[19] In this method, a bisulfite reagent was used to mediate the selective thiolation of 5hmC with a nucleophilic thiolate (Scheme 1). In a one-step reaction, 5hmC can be substituted to afford a variety of 5-thiomethyl derivatives in high yields. Unlike enzymatic modification, this reaction can label single-stranded DNA.

Highly concentrated bisulfite solution is commonly used in the whole-genome single-base-resolution sequencing of 5mC, which provides accurate and quantitative information on cytosine methylation. In this bisulfite sequencing, cytosine is deaminated to uracil and reads as thymine, whereas 5mC is inert



Scheme 1. Chemical transformation of 5hmC with thiol and bisulfite at low or high concentration.

to the deamination and reads as cytosine. Unfortunately, upon bisulfite treatment, 5hmC is converted to CMS, which behaves similarly to 5mC (Scheme 1).[3] In order to overcome this difficulty, two modified bisulfite sequencing methods, TAB-Seq and OxBS, have been developed to achieve single-base detection of 5hmC.[20,21] In Bayley's work, the bisulfite reagent is used at low concentration and low temperature; therefore, no deamination of cytosine was detected. In the presence of thiol, the thiolation is proposed to occur through either the addition of an *exo*-methylene intermediate or the substitution of sulfite ester to the 5-thiomethyl derivatives.

Besides the regular profiling, this 5hmC-labeling method for single-stranded DNA provides an opportunity to combine the labeling approach with nanopore sequencing. In principle, the readout window of the third-generation sequencing technology can differentiate the base information directly from the unamplified DNA.[22,23] When this technology is combined with selective enzymatic reactions, 5mC and 5hmC signals have been detected by single molecular real-time sequencing (SMRT).[24,25] Another attraction of the technology is that nanopore sequencing has been reported to detect the modified base by differentiating the pore current of each base in single-stranded DNA.[26] In their study, Bayley and co-workers also employed a nanopore system to detect the newly labeled 5hmC in single-stranded DNA. Once the 5hmC site had been modified by glutathione and bisulfite, the pore current was greatly decreased compared to the unlabeled 5hmC due to the bulky size of glutathione (Figure 1). The significant signal shift caused by modified 5hmC allows the detection of the 5hmC site in single-stranded DNA.

The chemical-labeling approach can be simple and cost-effective. The unique chemical properties and reactivity of the

[a] *X. Lu, Prof. C. He*
Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago
929 East 57th Street, Chicago, Illinois, 60637 (USA)
E-mail: chuanhe@uchicago.edu

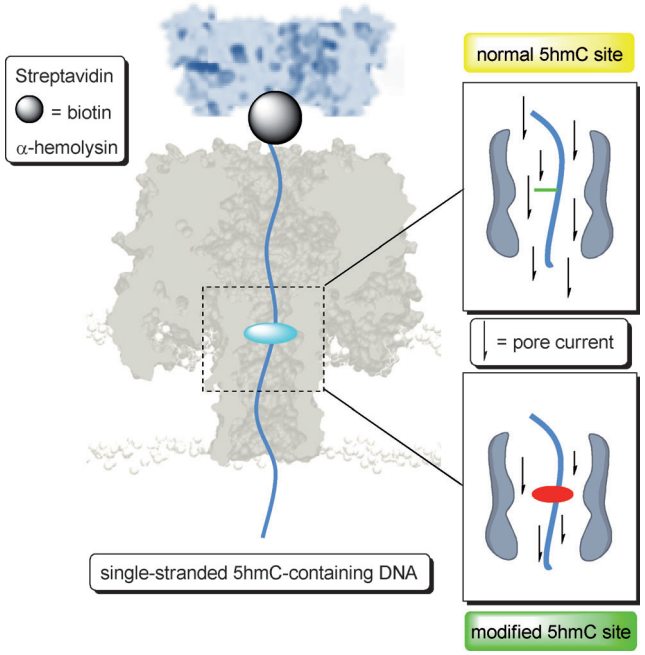


Figure 1. Detection of 5hmC and modified 5hmC in single-stranded DNA by nanopore.

hydroxy, formyl, or carboxyl group in 5hmC, 5fC, or 5caC, respectively, could offer the potential to develop new labeling approaches. Indeed, recent studies have shown nonenzymatic detection of 5fC by capturing the formyl group with aminoxy or hydrazine derivatives.[27,28] Nanopore or related technologies could be used to map these new cytosine derivatives.

Keywords: 5-hydroxylmethylcytosine · DNA methylation · epigenetics · nanopores · nonenzymatic labeling

[1] R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89 – 97.
[2] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929 – 930.
[3] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930 – 935.
[4] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146 – 7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008 – 7012.
[5] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300 – 1303.
[6] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303 – 1307.
[7] W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394 – 397.
[8] C. X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C. H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68 – 72.
[9] C. X. Song, Y. Sun, Q. Dai, X. Y. Lu, M. Yu, C. G. Yang, C. He, *ChemBioChem* **2011**, *12*, 1682 – 1685.
[10] K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343 – 348.
[11] G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398 – 402.
[12] H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *Genes Dev.* **2011**, *25*, 679 – 684.
[13] S. G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, *Nucleic Acids Res.* **2011**, *39*, 5015 – 5024.
[14] Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Mol. Cell* **2011**, *42*, 451 – 464.
[15] H. Stroud, S. Feng, S. Morey Kinney, S. Pradhan, S. E. Jacobsen, *Genome Biol.* **2011**, *12*, R54.
[16] A. A. Serandour, S. Avner, F. Oger, M. Bizot, F. Percevault, C. Lucchetti-Miganeh, G. Palierne, C. Gheeraert, F. Barloy-Hubler, C. L. Peron, T. Madigou, E. Durand, P. Froguel, B. Staels, P. Lefebvre, R. Metivier, J. Eeckhoute, G. Salbert, *Nucleic Acids Res.* **2012**, *40*, 8255 – 8265.
[17] A. B. Robertson, J. A. Dahl, C. B. Vagbo, P. Tripathi, H. E. Krokan, A. Klungland, *Nucleic Acids Res.* **2011**, *39*, e55.
[18] Z. Liutkevičiūtė, E. Kriukienė, I. Grigaitytė, V. Masevičius, S. Klimašauskas, *Angew. Chem.* **2011**, *123*, 2138 – 2141; *Angew. Chem. Int. Ed.* **2011**, *50*, 2090 – 2093.
[19] W. W. Li, L. Gong, H. Bayley, *Angew. Chem.* **2013**, *125*, 4446 – 4451; *Angew. Chem. Int. Ed.* **2013**, *52*, 4350 – 4355.
[20] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368 – 1380.
[21] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934 – 937.
[22] H. Bayley, *Curr. Opin. Chem. Biol.* **2006**, *10*, 628 – 637.
[23] P. K. Gupta, *Trends Biotechnol.* **2008**, *26*, 602 – 611.
[24] C. X. Song, T. A. Clark, X. Y. Lu, A. Kislyuk, Q. Dai, S. W. Turner, C. He, J. Korlach, *Nat. Methods* **2012**, *9*, 75 – 77.
[25] T. A. Clark, X. Lu, K. Luong, Q. Dai, M. Boitano, S. W. Turner, C. He, J. Korlach, *BMC Biol.* **2013**, *11*, 4.
[26] E. V. Wallace, D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, T. J. Donohoe, H. Bayley, *Chem. Commun.* **2010**, *46*, 8195 – 8197.
[27] C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678 – 691.
[28] E. A. Raiber, D. Beraldi, G. Ficz, H. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69.

Received: May 27, 2013
Published online on June 18, 2013