

Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA

Xingyu Lu,[†] Chun-Xiao Song,[†] Keith Szulwach,[‡] Zhipeng Wang,[†] Payton Weidenbacher,[†] Peng Jin,[‡] and Chuan He^{*,†}

[†]Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, United States

[‡]Department of Human Genetics, Emory University, School of Medicine, 615 Michael Street, Atlanta, Georgia 30322, United States

S Supporting Information

ABSTRACT: 5-Methylcytosine (5mC) in DNA can be oxidized stepwise to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) by the TET family proteins. Thymine DNA glycosylase can further remove 5fC and 5caC, connecting 5mC oxidation with active DNA demethylation. Here, we present a chemical modification-assisted bisulfite sequencing (CAB-Seq) that can detect 5caC with single-base resolution in DNA. We optimized 1-ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride (EDC)-catalyzed amide bond formation between the carboxyl group of 5caC and a primary amine group. We found that the modified 5caC can survive the bisulfite treatment without deamination. Therefore, this chemical labeling coupled with bisulfite treatment provides a base-resolution detection and sequencing method for 5caC.

5-Methylcytosine (5mC), the fifth base in the mammalian genome encoding critical epigenetic information, significantly impacts cellular biological functions and human diseases.¹ While the mechanism of cytosine methylation has been well studied, active demethylation pathways have remained elusive.² Recent studies have reported the accumulation of 5-hydroxymethylcytosine (5hmC) in certain mammalian cells and tissues,^{3,4} and the ten-eleven translocation (TET) family of enzymes that can oxidize 5mC stepwise to 5-hydroxymethylcytosine (5hmC),⁴ 5-formylcytosine (5fC), and finally 5-carboxylcytosine (5caC).^{4–7} The oxidation products 5fC and 5caC can be recognized and excised by mammalian DNA glycosylase (TDG) which generates an abasic site that can be further transformed to normal cytosine through base excision repair (BER), completing an active demethylation process.^{7–9}

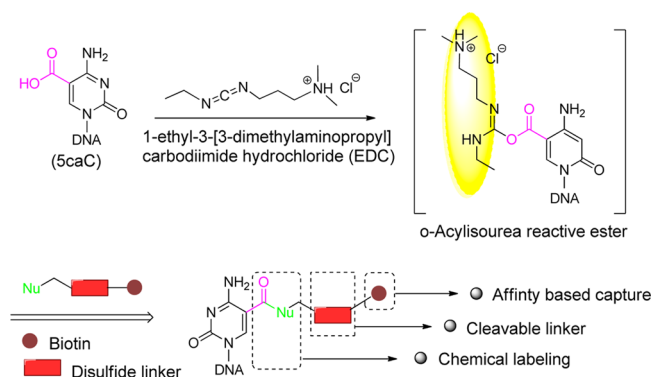
Quantification of these cytosine modifications in mouse embryonic stem (ES) cells showed decreased abundances from 5hmC to 5fC and further to 5caC, which are roughly 1300, 20, and 5 ppm per cytosine, respectively.^{5,6,10,11} The importance and the limited abundance of 5hmC and its further oxidized derivatives present challenges for effective methods to selectively label and sequence these modifications in genomic DNA.^{11–14} The developments of labeling and profiling technologies for 5hmC have significantly contributed to the study of 5hmC.^{11,13,15–17} Patterns of 5hmC distribution unique from those of 5mC were revealed.^{15,18} To further understand

5mC and the subsequent active demethylation, methods to profile and/or sequence 5fC and 5caC are required. Recently, we have developed selective chemical labeling and profiling methods for 5fC, indicating its function in TDG-mediated demethylation at gene regulatory elements.¹⁹ Similar conclusions were also obtained using an antibody-based approach.²⁰ Under typical bisulfite conditions, both 5fC and 5caC are deaminated and read as T, making them indistinguishable from unmodified cytosine. We showed that the labeling of 5fC by hydroxylamine protects it from bisulfite-mediated deamination, thereby allowing detection and sequencing of 5fC with base resolution.¹⁹ We present here a chemical modification-assisted bisulfite sequencing method that allows base-resolution detection of 5caC in DNA.

We chose the 1-ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride (EDC)-catalyzed amide bond formation between carboxyl group and primary amine group as our chemical selective labeling method. The EDC-based coupling reaction has been well developed and widely used.^{21–23} The reaction can be performed in aqueous solution under relatively mild conditions around neutral pH which prevents DNA degradation. The carboxylic acid group can react with EDC to form an *o*-acylisourea ester intermediate that can then be attacked by a primary amine (Scheme 1). Since the 3' and 5' terminal phosphate groups can be easily removed through treatment with phosphatase, the carboxyl group of 5caC will be the most reactive functional group with EDC. After the coupling reaction, a biotin-modified amine may be installed onto the 5caC-containing DNA (Scheme 1). Then, the streptavidin coated beads can be employed to capture the 5caC-containing fragments for subsequent enrichment. The designed disulfide bond between amine group and biotin can be smoothly cleaved by DTT treatment (see Figure S1). As a result, the 5caC-containing DNA may be pulled down, enriched, and subjected to high-throughput sequencing. The key to this strategy is to optimize the labeling efficiency and selectivity.

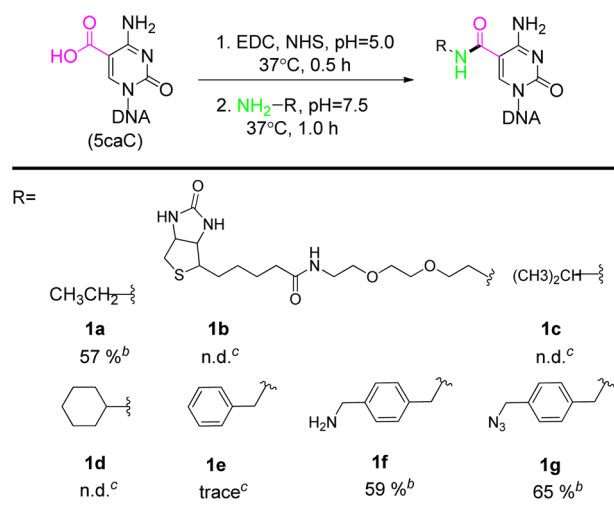
We started by screening reactions on the short double-stranded 5caC-containing oligonucleotides and monitoring the reaction by mass spectrometry, quantifying the yield by HPLC (Table 1). To our delight, ethylamine can react with 5caC

Received: May 5, 2013

Scheme 1. EDC-Catalyzed Chemical Labeling of ScaC^a

^aDesigned labeling strategy for ScaC. The carboxyl group of ScaC in duplex DNA can be labeled by EDC-catalyzed coupling reaction to form biotin-S-S-ScaC using biotin-S-S-amine.

Table 1. EDC-mediated ScaC coupling with different amines.^a



^aGeneral reaction conditions: double stranded ScaC containing DNA (10 μ mol), NHS (20 mM), EDC (2 mM), Mes (pH = 5.0, 75 mM) were incubated in aqueous solution at 37 $^{\circ}$ C for 0.5 h. Buffer exchange to sodium phosphate (pH = 7.5, 100 mM), NaCl (150 mM), amine (10 mM), incubated at 37 $^{\circ}$ C for 1.0 h. ^bReaction yield based on HPLC analysis (see Figure S5). ^cUndetectable product by mass spectrometry. (see Figure S6).

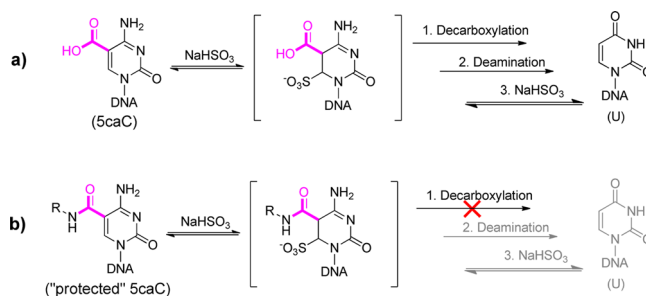
DNA with moderate labeling efficiency (1a in Table 1). This approach also showed high selectivity to ScaC among other cytosine modifications such as 5hmC and 5fC (see Figure S2). The side reaction between the DNA phosphate backbone and ethylamine is undetectable on mass spectrometry. A two-step EDC coupling strategy, which generates the *N*-hydroxysuccinimide (NHS) ester intermediate at pH = 5.0 and then formation of the amide bond at pH = 7.5, showed higher reactivity than the direct coupling at pH = 6.0. Unfortunately, when the biotin-PEG2-amine was employed under the optimized reaction conditions, the coupling reaction failed as a result of the increase in steric hindrance (1b in Table 1). Subsequently, a series of primary amines were tested in order to optimize labeling efficiency. As expected, the bulky amines greatly inhibited the labeling reaction (1c–1e in Table 1).

Further screening revealed that the xylene group significantly facilitates this transformation on DNA. The xylenediamine

showed high reaction activity comparable to ethylamine (1f in Table 1). Notably, the benzylamine only gave trace amounts of coupling product (1e in Table 1). Subsequently, we synthesized (4-aminomethyl)benzylazide which possesses the xylene structure but also contains an azide group for the well-developed bio-orthogonal coupling of the biotin tag.^{24,25} As expected, the (4-aminomethyl)benzylazide showed as good reactivity as xylenediamine (1g in Table 1). We show that the azide group can be coupled with a strained cyclooctyne bearing a disulfide linker to a biotin.^{26,27} After a simple DTT treatment, the disulfide bond can be selectively cleaved with high efficiency. The high reaction activity and selectivity during each step were monitored by mass spectrometry (see Figure S3). The biotin-labeled ScaC nucleoside was further confirmed with high resolution mass spectrometry after DNA digestion (see Figure S4). This chemical labeling step has no noticeable sequence bias.

With this efficient labeling method in hand, we tested if the protection of the carboxyl group could prevent deamination of ScaC under typical bisulfite treatment. It has been shown that under typical bisulfite conditions ScaC will be deaminated and read as T after subsequent PCR amplification and sequencing.⁷ This mechanism likely goes through a process of decarboxylation followed by deamination (Scheme 2a).²⁸ Protection of

Scheme 2. Proposed Deamination Mechanism of ScaC under Bisulfite Treatment



the carboxyl group as an amide bond could block the decarboxylation step and significantly slow down the bisulfite-mediated deamination (Scheme 2b). To probe this mechanism, we performed preliminary studies on ScaC single nucleoside monitored by HPLC. We found that after bisulfite treatment, ScaC forms the same bisulfite adduct intermediate as normal cytosine, which will convert to the uridine bisulfite adduct, supporting decarboxylation as a potentially critical step (see Figure S7). To further confirm this observation, we labeled a synthetic 76mer ScaC-containing DNA with (4-aminomethyl)benzylazide and applied the labeled DNA to bisulfite treatment. The labeled DNA did not block DNA polymerase and can be successfully amplified with PCR under common conditions. No PCR amplification bias was detected after the EDC-mediated labeling of ScaC (Figure S8). After Sanger sequencing, the ScaC sites were mainly read as cytosine (C) instead of thymine (T) while the unlabeled ScaCs were still read as T (Figure 1). Thus, the ScaC can be “protected” from bisulfite-based deamination after the EDC-mediated labeling reaction. The different behaviors of ScaC before and after EDC labeling in bisulfite sequencing provide the single-base resolution detection of ScaC in DNA. In real biological applications, combined with high-throughput sequencing technology or colony picking, the modified cytosine can be detected at very low abundance

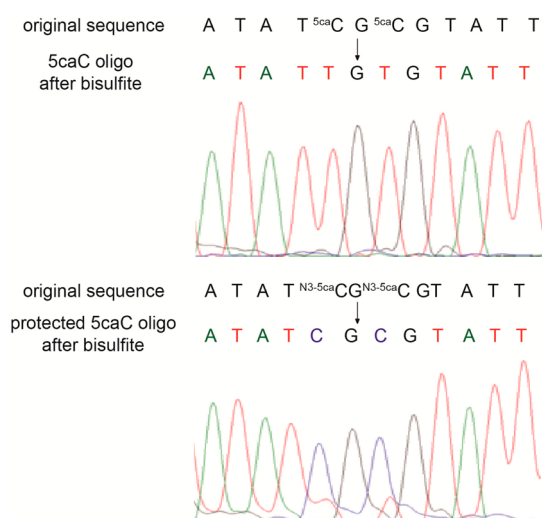


Figure 1. Sanger sequencing results of unlabeled 76 mer 5caC (upper panel) and EDC-labeled 5caC (bottom panel) after bisulfite treatment. The original sequence and the sequence after bisulfite were shown above the corresponding peak. In the bottom panel, the labeled DNA was not further purified before bisulfite treatment. The small background peak of T comes from unlabeled cytosine, which reflects the labeling efficiency.

through modified bisulfite sequencing as we have demonstrated recently.^{15,19} Together with enrichment-based approach,²⁰ this chemical method has potential to gain base-resolution information of 5caC genome-wide.

In summary, we show here an efficient chemical selective labeling of 5caC in DNA catalyzed by EDC using xylene-based primary amine. The xylene structure is critical for highly efficient labeling. Importantly, we showed that the chemically labeled 5caC could survive from bisulfite-mediated deamination and be differentiated at single-base resolution using this new version of CAB-Seq specific for 5caC. This effective method, communicated here, can help the broad community interested in 5caC biology to detect this modified base in various biological samples.

■ ASSOCIATED CONTENT

● Supporting Information

Complete experimental procedures, mass spectrum data, qPCR data and HPLC data for reactions, spectral data for new compounds. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

chuanhe@uchicago.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by National Institutes of Health HG006827 (C.H.), NS079625, and HD073162 (P.J.). We thank S. F. Reichard, MA for editing the manuscript.

■ REFERENCES

- (1) Klose, R. J.; Bird, A. P. *Trends Biochem. Sci.* **2006**, *31*, 89.
- (2) Bhutani, N.; Burns, D. M.; Blau, H. M. *Cell* **2011**, *146*, 866.
- (3) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
- (4) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
- (5) Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008.
- (6) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.
- (7) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. *Science* **2011**, *333*, 1303.
- (8) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334.
- (9) Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G. L.; Luo, C.; Jiang, H.; He, C. *Nat. Chem. Biol.* **2012**, *8*, 328.
- (10) Globisch, D.; Munzel, M.; Muller, M.; Michalak, S.; Wagner, M.; Koch, S.; Bruckl, T.; Biel, M.; Carell, T. *PLoS One* **2010**, *5*, e15367.
- (11) Song, C. X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C. H.; Zhang, W.; Jian, X.; Wang, J.; Zhang, L.; Looney, T. J.; Zhang, B.; Godley, L. A.; Hicks, L. M.; Lahn, B. T.; Jin, P.; He, C. *Nat. Biotechnol.* **2011**, *29*, 68.
- (12) Song, C. X.; Sun, Y.; Dai, Q.; Lu, X. Y.; Yu, M.; Yang, C. G.; He, C. *ChemBioChem* **2011**, *12*, 1682.
- (13) Pastor, W. A.; Pape, U. J.; Huang, Y.; Henderson, H. R.; Lister, R.; Ko, M.; McLoughlin, E. M.; Brudno, Y.; Mahapatra, S.; Kapranov, P.; Tahiliani, M.; Daley, G. Q.; Liu, X. S.; Ecker, J. R.; Milos, P. M.; Agarwal, S.; Rao, A. *Nature* **2011**, *473*, 394.
- (14) Raiber, E. A.; Beraldi, D.; Ficiz, G.; Burgess, H.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, *13*, R69.
- (15) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C. X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J. H.; Jin, P.; Ren, B.; He, C. *Cell* **2012**, *149*, 1368.
- (16) Robertson, A. B.; Dahl, J. A.; Vagbo, C. B.; Tripathi, P.; Krokan, H. E.; Klungland, A. *Nucleic Acids Res.* **2011**, *39*, e55.
- (17) Booth, M. J.; Branco, M. R.; Ficiz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S. *Science* **2012**, *336*, 934.
- (18) Stadler, M. B.; Murr, R.; Burger, L.; Ivanek, R.; Lienert, F.; Scholer, A.; van Nimwegen, E.; Wirbelauer, C.; Oakeley, E. J.; Gaidatzis, D.; Tiwari, V. K.; Schubeler, D. *Nature* **2011**, *480*, 490.
- (19) Song, C. X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S. Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G. L.; Jin, P.; He, C. *Cell* **2013**, *153*, 678.
- (20) Shen, L.; Wu, H.; Diep, D.; Yamaguchi, S.; D'Alessio, A. C.; Fung, H. L.; Zhang, K.; Zhang, Y. *Cell* **2013**, *153*, 692.
- (21) Panchaud, A.; Hansson, J.; Affolter, M.; Bel Rhlid, R.; Piu, S.; Moreillon, P.; Kussmann, M. *Mol. Cell. Proteomics* **2008**, *7*, 800.
- (22) Kim, J. H.; Kushi, K.; Graham, N. A.; Asthagiri, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11149.
- (23) Naue, N.; Fedorov, R.; Pich, A.; Manstein, D. J.; Curth, U. *Nucleic Acids Res.* **2011**, *39*, 1398.
- (24) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004.
- (25) Tornoe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057.
- (26) Ning, X.; Guo, J.; Wolfert, M. A.; Boons, G. J. *Angew. Chem., Int. Ed.* **2008**, *47*, 2253.
- (27) Sletten, E. M.; Bertozzi, C. R. *Angew. Chem., Int. Ed.* **2009**, *48*, 6974.
- (28) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Muller, M.; Hagemeier, C.; Truss, M.; Carell, T. *Angew. Chem., Int. Ed.* **2012**, *51*, 6516.