Published in final edited form as:

Chembiochem. 2011 July 25; 12(11): 1682–1685. doi:10.1002/cbic.201100278.

Detection of 5-Hydroxymethylcytosine in DNA by Transferring a Keto-Glucose Using T4 phage β -Glucosyltransferase**

Chun-Xiao Song^[a], Yao Sun^[b], Qing Dai^[a], Xing-Yu Lu^[a], Miao Yu^[a], Cai-Guang Yang^{*,[b]}, and Chuan He^{*,[a]}

^[a] Department of Chemistry and Institute for Biophysical Dynamics The University of Chicago 929 East 57th Street, Chicago, IL, 60637 (USA)

^[b] Shanghai Institute of Materia Medica Chinese Academy of Sciences 555 Zuchongzhi Road, Shanghai, 201203 (China)

Keywords

5-Hydroxymethylcytosine; Selective chemical labelling; β -Glucosyltransferase; Single-base resolution detection

5-Methylcytosine (5-mC) is the most important DNA modification found in eukaryotes. This base modification carries epigenetic information that regulates gene expression and disease pathogenesis.^[1] Recently, 5-hydroxymethylcytosine (5-hmC), another form of DNA modification generated by Tet dioxygenases, has been identified in substantial amounts in certain mammalian cell types.^[2] Recent functional studies suggest that Tet1-2 are required for embryonic stem (ES) cell self-renewal and maintenance;^[3] Tet2 is related to normal myelopoiesis;^[4] and Tet3 is involved in zygotic development.^[5] All these results point to 5-hmC as another important epigenetic mark. To further reveal the biology of 5-hmC, sensitive detection and sequencing methods for 5-hmC,^[6] especially high-resolution sequencing methods,^[7] are highly desirable, since traditional bisulfide sequencing cannot differentiate 5-hmC from 5-mC.^[8] We have recently developed a selective chemical labeling method that utilizes T4 bacteriophage β -glucosyltransferase (β -GT) to transfer an azide-substituted glucose onto the hydroxyl group of 5-hmC, followed by a click reaction to install a biotin tag for detection and enrichment of 5-hmC-containing DNA for deep sequencing.^[6a] This strategy also allows versatile tagging systems to be added onto 5-hmC in genomic DNA. Herein we report an alternative bioorthogonal labeling strategy using a keto-substituted glucose that possesses advantages over the azide-substituted glucose and is capable of single-base resolution detection/sequencing of 5-hmC.

As shown in Figure 1, based on our previously results, we expect that wild-type β -GT should be able to transfer a 6-keto glucose, which is similar in size to 6-azide glucose, onto

* Fax: (+1) 773-702-0805 chuanhe@uchicago.edu yangcg@mail.shcnc.ac.cn.

Supporting information for this article is available on the WWW under <http://www.chembiochem.org> or from the author.

Capture with ketone: 5-Hydroxymethylcytosine in DNA can be selectively labelled with a keto-glucose and subsequently linked to biotin via a ketoxime linker for enrichment and detection/sequencing purposes. Keto-glucose can be more efficiently transferred than azideglucose by β -glucosyltransferase, and can provide single-base resolution detection/sequencing of 5-hmC.

5-hmC using a modified cofactor UDP-6-keto-Glc. After selectively labeling 5-hmC in DNA with 6-keto glucose to form keto-5-gmC, a biotinhydroxylamine (*N*-(aminooxyacetyl)-*N'*-(D-biotinoyl) hydrazine) can be added, which can react with the ketone group on keto-5-gmC to form biotin-keto-5-gmC (Figure 1b). The incorporated biotin group can thus serve as a detection and enrichment moiety for sequencing purposes.

We chemically synthesized UDP-6-keto-Glc based on a modified synthetic route,^[6a, 9] and tested the labeling efficiency of this new approach on an 11-mer synthetic DNA.^[10] As expected, the glycosylation reaction of wild-type β -GT with UDP-6-keto-Glc as well as the subsequent reaction with biotin-hydroxylamine to install the biotin group proceeded efficiently (Figure 2). We compared the efficiency of β -GT using UDP-6-keto-Glc versus UDP-6-N₃-Glc.^[6a] As shown in Table 1, UDP-6-keto-Glc is slightly more effective as a cofactor than UDP-6-N₃-Glc on the double-stranded (ds) DNA substrate, although both cofactors are already fairly effective compared to the native cofactor UDP-Glc (Table 1, Entries 1-3). We also tested the glycosylation reaction on the single-stranded (ss) DNA substrate. We found that UDP-6-keto-Glc is more efficient than UDP-6-N₃-Glc in this case, although a large amount of β -GT was required for effective conversion in the case of the ssDNA substrate (Table 1, Entries 4-6). Overall, β -GT transfers the modified glucose from UDP-6-keto-Glc better than from UDP-6-N₃-Glc. UDP-6-keto-Glc could be a better substrate for applications such as labeling 5-hmC in single-stranded RNA and DNA.

This keto-glucose-based labelling of 5-hmC provides a method alternative to the azide-based approach published previously.^[6a] A major advantage of the modified biotinketo-5-gmC is the shorter linker of ketoxime compared to the cyclooctyne-based linker in the original azide-based method.^[6a] We envision that if streptavidin (a protein that recognizes and binds tightly to biotin) is added to bind biotin-keto-5-gmC in duplex DNA, the short ketoxime linker will position the bound streptavidin close to 5-hmC, thus resulting in better blocking of certain enzymatic processes to enable high-resolution detection of 5-hmC. Exonuclease digestion is an example of such enzymatic processes. Exonucleases are enzymes that cleave nucleotides one at a time from the end of a polynucleotide chain. A bulky glucose-biotin-streptavidin complex of 5-hmC on duplex DNA may block the exonuclease from cutting through, allowing single-base resolution detection of 5-hmC independent of DNA sequences. We have tested this idea using the previous azide-glucose/cyclooctyne-linker approach but failed (data not shown). We attribute the reason to the long cyclooctyne-based linker which positions the blocking group far away from the DNA.

We tried the new ketoxime-based biotin labeling strategy in an exonuclease digestion assay using Exonuclease III, a 3' to 5' exonuclease.^[11] The same 11-mer DNA containing biotin-keto-5-gmC complementary to a 15-mer regular DNA was used (Figure 2). After digestion with Exonuclease III in the presence of streptavidin, the remaining DNA was subjected to MALDI analysis. To our delight, addition of streptavidin blocked Exonuclease III from cleaving through. As shown in Figure 3, MALDI result shows that the complementary 15-mer regular DNA is completely digested by Exonuclease III. In contrast, digestion of the 11-mer DNA (5'-CGACT*CAAGGC-3', *C represents biotinketo-5-gmC) generated two DNA fragments with molecular weights corresponding to 5'-CGACT*CA-3' and 5'-CGACT*C-3' (Figure 3). This observation indicates that biotinketo-5-gmC with bound

streptavidin can block digestion by Exonuclease III mainly at one base before the modification, but also right at the modification position. The result has been confirmed by gel analysis of the same digestion assay performed on a 50-mer dsDNA with two modification sites (Figure 4). Substantial blocking was observed in the case of biotin-keto-5-gmC with bound streptavidin but not in the case of 5-hmC (Figure 4, Lanes 4-6 vs Lanes 1-3). The result also shows that Exonuclease III is blocked mostly at the first modification site (Figure 4, Lane 6).

The Exonuclease III digestion assay has a clear blocking pattern that may be developed into be a high-throughput method to provide single-base resolution sequencing of 5-hmC in genomic DNA. Unlike the restriction enzyme digestion assay, this new approach does not depend on the DNA sequence, and thus may be applied to analyse the distribution of 5-hmC in the whole genome. Since regular DNA without modification will be completely digested, this approach has the potential to discriminate fully- versus hemi-hydroxymethylation sites in genomic DNA.

The advances of bioorthogonal chemistry provide key tools for modern biological research.^[12] We present here a chemical labelling method of 5-hmC using keto-glucose transferred by β -GT. This enzymatic transfer chemistry of keto-glucose is more efficient than the previous method that transfers azide-glucose. More importantly, the biotin group can be tagged through a short ketoxime linker from 5-hmC in duplex DNA. With the short linker the modified biotin-keto-5-gmC with bound streptavidin blocks digestion by Exonuclease III near the modification site, thus providing a single-base resolution detection of 5-hmC independent of the DNA sequence. We are in the process of adapting this method into deep sequencing for the genome-wide, high-resolution sequencing of 5-hmC in mammalian genomic DNA samples.

Experimental Section

Labelling of 5-hmC with keto-glucose and subsequent reaction with biotinhydroxylamine

In a typical reaction, 30 μ M synthetic 5-hmC-containing DNA in 50 mM HEPES buffer (pH 7.9), 25 mM $MgCl_2$, 250 μ M UDP-6-keto-Glc, and 2 μ M wild-type β GT was incubated for 1 h at 37 °C. After the reaction, the DNA substrates were purified by Bio-Rad Micro Bio-Spin 6 spin column (exchange buffer to H_2O first). The biotinhydroxylamine reaction was performed with addition of 1 mM of *N*-(aminooxyacetyl)-*N'*-(D-biotinoyl) hydrazine (Invitrogen) in 10 mM MES buffer (pH 6.0), and the reaction mixture was incubated for 2 h at 37 °C. The DNA samples were then purified by Bio-Rad Micro Bio-Spin 6 spin column.

Exonuclease digestion assay

In a typical assay, 200-400 ng labelled synthetic DNA was digested by 100 units of Exonuclease III (New England Biolabs) in the presence of 2 μ M streptavidin (Pierce) for 30 min at 37 °C followed by 20 min heat inactivation at 70 °C. Afterwards, streptavidin was released by denaturing for 10 min at 100 °C in the presence of 1% SDS and 400 μ M free biotin. Digested DNA was purified by Bio-Rad Micro Bio-Spin 6 spin column for MALDI or gel analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institute of Health (GM071440 to C.H.) and The University of Chicago; Hundred Talent Program of the Chinese Academy of Sciences (C.Y.).

References

1. a Goll MG, Bestor TH. *Annu. Rev. Biochem.* 2005; 74:481–514. [PubMed: 15952895] b Klose RJ, Bird AP. *Trends Biochem. Sci.* 2006; 31:89–97. [PubMed: 16403636] c Reik W. *Nature.* 2007; 447:425–432. [PubMed: 17522676] d Weber M, Schubeler D. *Curr. Opin. Cell Biol.* 2007; 19:273–280. [PubMed: 17466503] e Gal-Yam EN, Saito Y, Egger G, Jones PA. *Annu. Rev. Med.* 2008; 59:267–280. [PubMed: 17937590]
2. a Kriaucionis S, Heintz N. *Science.* 2009; 324:929–930. [PubMed: 19372393] b Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. *Science.* 2009; 324:930–935. [PubMed: 19372391]
3. a Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. *Nature.* 2010; 466:1129–1133. [PubMed: 20639862] b Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, Rodig SJ, Daley GQ, Rao A. *Cell Stem Cell.* 2011; 8:200–213. [PubMed: 21295276]
4. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, Liu XS, Aravind L, Agarwal S, Maciejewski JP, Rao A. *Nature.* 2010; 468:839–843. [PubMed: 21057493]
5. a Iqbal K, Jin SG, Pfeifer GP, Szabo PE. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:3642–3647. [PubMed: 21321204] b Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J. *Nat. Commun.* 2011; 2:241. [PubMed: 21407207]
6. a Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C. *Nat. Biotechnol.* 2011; 29:68–72. [PubMed: 21151123] b Munzel M, Globisch D, Bruckl T, Wagner M, Welzmler V, Michalakakis S, Muller M, Biel M, Carell T. *Angew. Chem., Int. Ed.* 2010; 49:5375–5377. c Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. *Nucleic Acids Res.* 2010; 38:e181. [PubMed: 20685817] d Robertson AB, Dahl JA, Vagbo CB, Tripathi P, Krokan HE, Klungland A. *Nucleic Acids Res.* 2011; 39:e55. [PubMed: 21300643] e Davis T, Vaisvila R. *J. Vis. Exp.* 2011;48. doi:10.3791/2661. f Jin SG, Wu X, Li AX, Pfeifer GP. *Nucleic Acids Res.* 2011 doi: 10.1093/nar/gkr120. g Song C-X, Yu M, Dai Q, He C. *Bioorg. Med. Chem. Lett.* 2011 doi:10.1016/j.bmcl.2011.03.118. h Höbartner C. *Angew. Chem., Int. Ed.* 2011; 50:4268–4270. i Song C-X, He C. *Acc. Chem. Res.* 2011 doi: 10.1021/ar2000502.
7. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. *Nat. Methods.* 2010; 7:461–465. [PubMed: 20453866]
8. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. *PLoS One.* 2010; 5:e8888. [PubMed: 20126651]
9. Campbell RE, Tanner ME. *J. Org. Chem.* 1999; 64:9487–9492.
10. a Dai Q, Song C-X, Pan T, He C. *J. Org. Chem.* 2011 doi: 10.1021/jo200566d. b Tardy-Planechaud S, Fujimoto J, Lin SS, Sowers LC. *Nucleic Acids Res.* 1997; 25:553–559. [PubMed: 9016595] c Munzel M, Globisch D, Trindler C, Carell T. *Org. Lett.* 2010; 12:5671–5673. [PubMed: 21082782]
11. a Mol CD, Kuo CF, Thayer MM, Cunningham RP, Tainer JA. *Nature.* 1995; 374:381–386. [PubMed: 7885481] b Rogers SG, Weiss B. *Methods Enzymol.* 1980; 65:201–211. [PubMed: 6246343] c Henikoff S. *Gene.* 1984; 28:351–359. [PubMed: 6235151]
12. a Rostovtsev VV, Green LG, Fokin VV, Sharpless KB. *Angew. Chem., Int. Ed.* 2002; 41:2596–2599. b Speers AE, Cravatt BF. *Chem. Biol.* 2004; 11:535–546. [PubMed: 15123248] c Ning X,

Guo J, Wolfert MA, Boons GJ. *Angew. Chem., Int. Ed.* 2008; 47:2253–2255. d Sletten EM, Bertozzi CR. *Angew. Chem., Int. Ed.* 2009; 48:6974–6998. e Mamidyala SK, Finn MG. *Chem. Soc. Rev.* 2010; 39:1252–1261. [PubMed: 20309485] f Khidekel N, Arndt S, Lamarre-Vincent N, Lippert A, Poulin-Kerstien KG, Ramakrishnan B, Qasba PK, Hsieh-Wilson LC. *J. Am. Chem. Soc.* 2003; 125:16162–16163. [PubMed: 14692737] g Soriano Del Amo D, Wang W, Jiang H, Besanceney C, Yan AC, Levy M, Liu Y, Marlow FL, Wu P. *J. Am. Chem. Soc.* 2010; 132:16893–16899. [PubMed: 21062072] h Zhang MM, Tsou LK, Charron G, Raghavan AS, Hang HC. *Proc. Natl. Acad. Sci. U. S. A.* 2010; 107:8627–8632. [PubMed: 20421494] i Salic A, Mitchison TJ. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:2415–2420. [PubMed: 18272492]

Figure 1.

Selective labelling of 5-hmC in genomic DNA using a keto-glucose strategy. A) The hydroxyl group of 5-hmC in duplex DNA can be glycosylated by β -GT to form β -6-ketoglucosyl-5-hydroxymethylcytosine (keto-5-gmC) using UDP-6-keto-Glc as a co-factor. B) After a ketone group is selectively installed onto 5-hmC, it can be labelled with a biotin moiety using biotin-hydroxylamine for subsequent affinity purification and detection/sequencing.

Figure 2.

Selective chemical labelling of 5-hmC with keto-glucose and the subsequent reaction with biotin-hydroxylamine, showing the molecular structures of each step in the workflow next to the corresponding MALDI spectra. Reactions were performed in duplex DNA with the complementary strand; however, mass spectrometry (MS) monitored the single-stranded DNA containing the modification. Black is theoretical MS; grey is the observed MS.

Figure 3.

Exonuclease III digestion assay of a 11-mer biotin-keto-5-gmC-containing DNA in the presence of streptavidin, showing MALDI spectra after the digestion. Exonuclease III digestion can be blocked mainly at one base before the modification, but also right at the modification position. Black is theoretical MS; grey is the observed MS.

Figure 4.

Gel analysis of Exonuclease digestion assay on a 50-mer dsDNA with two cytosine modification sites in the presence and absence of streptavidin. Samples were analyzed by 16% PAGE/Urea gel and visualized using SYBR Green I staining (Lumiprobe).

Table 1

Glycosylation reaction of p-GT with different cofactors on dsDNA or ssDNA containing 5-hmC.^[a]

Entry	cofactor	DNA	β -GT [%]	Time [min]	Yield [%] ^[b]
1	UDP-Glc	ds	0.4	10	100
2	UDP-6-N3-Glc	ds	0.4	10	75
3	UDP-6-keto-Glc	ds	0.4	10	89
4	UDP-Glc	ss	20	60	100
5	UDP-6-N3-Glc	ss	20	60	20
6	UDP-6-keto-Glc	ss	20	60	61

^[a] Reaction conditions: 30 mM DNA in 50 mM HEPES buffer (pH 7.9), 25 mM MgCl₂, 250 mM cofactor, and 0.12 mM wild-type p-GT (dsDNA) or 6 mM wild-type p-GT (for ssDNA) at 37 °C.

^[b] HPLC yields reported.