

Published in final edited form as:

*Bioorg Med Chem Lett.* 2011 September 1; 21(17): 5075–5077. doi:10.1016/j.bmcl.2011.03.118.

## Detection of 5-Hydroxymethylcytosine in a Combined Glycosylation Restriction Analysis (CGRA) Using Restriction Enzyme Taq<sup>q</sup>I

Chun-Xiao Song<sup>1</sup>, Miao Yu<sup>1</sup>, Qing Dai, and Chuan He\*

Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois, USA

### Abstract

5-Hydroxymethylcytosine (5-hmC) is a newly discovered DNA base in mammalian cells that is believed to be another important epigenetic modification. Here we report the use of a methylation-insensitive restriction enzyme Taq<sup>q</sup>I coupled with selective chemical labeling of 5-hmC in a Combined Glycosylation Restriction Analysis (CGRA) to detect 5-hmC in TCGA sequences. This method, differentiates fully versus hemi-hydroxymethylated cytosine in the CpG dinucleotide, adds a new tool to facilitate biological studies of 5-hmC.

### Keywords

5-Hydroxymethylcytosine; Selective chemical labeling; Restriction enzyme assay

In addition to the four standard DNA bases, base modifications can add extra epigenetic information to the DNA that is not encoded in the sequence. For example, in mammalian genome, 5-methylcytosine (5-mC), which constitutes about 1% of total nucleotide and mainly occurs in the CpG dinucleotide, has a profound effect on gene regulation and dysregulation.<sup>1–3</sup> In 2009, two groups independently reported the existence of 5-hydroxymethylcytosine (5-hmC) in mammalian genomic DNA.<sup>4, 5</sup> Meanwhile the TET family enzymes, which are methylcytosine dioxygenases, have been identified to oxidize 5-mC to 5-hmC, and play roles in embryonic stem (ES) cell regulation and myelopoiesis.<sup>5–8</sup> These studies alter our perception of how DNA methylation impacts mammalian cells by suggesting that 5-hmC might also play an important role in epigenetic regulation.

To further reveal the function of 5-hmC, sensitive detection and sequencing methods are required, as the widely used sequencing methods for 5-mC, such as bisulfite sequencing and methylation-sensitive restriction digestion, cannot discriminate between the two cytosine modifications.<sup>9–12</sup> Commercially available anti-5-hmC antibodies are unlikely to provide high-resolution sequencing of 5-hmC. Recently, we have developed a chemical labeling method to selectively label 5-hmC with an azide-modified glucose by  $\beta$ -glucosyltransferase ( $\beta$ GT). A subsequent coupling of the azide group with an alkyne-biotin probe allows

© 2011 Elsevier Ltd. All rights reserved.

\*To whom correspondence should be addressed. chuanhe@uchicago.edu.

<sup>1</sup>These authors contributed equally to this work.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

detection of 5-hmC in genomic DNA (Fig. 1).<sup>13</sup> With this method, 5-hmC-containing genomic DNA fragments can be enriched and sequenced to provide the genomic distribution of this modification.<sup>13</sup> However, a high-throughput method for single-base resolution detection of 5-hmC has yet to be achieved.

The use of methylation-sensitive restriction enzymes is a classic approach to the study of DNA methylation at specific loci.<sup>14</sup> However, due to the similar size of the methyl and hydroxymethyl groups, 5-mC and 5-hmC are indistinguishable to most restriction enzymes.<sup>10, 11, 15, 16</sup> We anticipated that after adding the bulky glucose group or a subsequent biotin attachment to 5-hmC, the glucosylated base would have different properties from 5-mC.<sup>17</sup> For example, the resulting 5-N<sub>3</sub>-gmC or biotin-5-N<sub>3</sub>-gmC in a duplex DNA may be able to block digestion from the methylation-insensitive restriction enzyme, which can digest both 5-hmC- and 5-mC-containing DNA. While we are developing this method, two companies, Zymo Research and New England Biolabs, have launched products based on this combined glycosylation restriction analysis (CGRA). They utilize  $\beta$ GT to transfer a regular glucose to 5-hmC and show that it can block the methylation-insensitive restriction enzyme MspI, which has a recognized sequence of C<sup>^</sup>CGG.<sup>18</sup> Although the use of MspI in CGRA can detect the presence of 5-hmC on CCGG site, it has several limitations: i) MspI is also blocked if the outer C is 5-mC or 5-hmC, regardless of the cytosine modification status of the inner C, which limits the use of this approach on many CCGG sites where the outer C methylated;<sup>15</sup> ii) it cannot tell whether 5-hmC occurs on only one strand or both strands of the CpG dinucleotide (see below). Herein, we show that Taq<sup>q</sup>I, another methylation-insensitive restriction enzyme that recognizes and cuts T<sup>^</sup>CGA, can also be used in CGRA when coupled with our chemical labeling method. This new approach can differentiate fully versus hemi-hydroxymethylated states in the CpG dinucleotide.

We first synthesized a 32-mer double strand DNA bearing T<sup>^</sup>CGA (\*C = 5-hmC) on both strands (Fig. 2A). Instead of glucose, we employed  $\beta$ GT to transfer our chemically modified 6-N<sub>3</sub>-glucose, onto 5-hmC (Fig. 1). Then we subjected the modified DNA to Taq<sup>q</sup>I-mediated digestion. As expected, Taq<sup>q</sup>I can completely cut unmodified 5-hmC, but only partially for N<sub>3</sub>-5-gmC. (Fig. 2A, lane 1-4).<sup>17</sup> The relatively high tolerance of Taq<sup>q</sup>I to the cytosine modification requires more bulky modifications in order to achieve a satisfactory difference in CGRA. The presence of azide group on the glucose allows us to add further modification using click chemistry.<sup>19-21</sup> We therefore coupled N<sub>3</sub>-5-gmC with dibenzocyclooctyne-modified biotin using copper-free click chemistry to introduce a sterically bulky dibenzocyclooctyne moiety (Fig. 1).<sup>22, 23</sup> To our delight, when we applied the biotin-N<sub>3</sub>-5-gmC modification into Taq<sup>q</sup>I digestion, it showed an almost complete blocking effect (Fig. 2A, lane 5-6).

Due to the semi-conservative DNA replication, besides a full methylation state, hemi-methylation state also exists in mammalian genome. The conversion of 5-mC to 5-hmC suggests that fully and hemi-hydroxymethylation states may also exist in the mammalian genome. If this is the case, developing a method to distinguish between these two states will be very important to understanding the formation of 5-hmC and conversion process between 5-mC and 5-hmC. Since the blocking efficiency of Taq<sup>q</sup>I is largely dependent on the size of the modification group on the hydroxyl group, we wondered if Taq<sup>q</sup>I can behave differently over fully and hemi-hydroxymethylation states. We prepared the same 32-mer double strand DNA with hemi-hydroxymethylation, performed the same labeling procedure, and subjected to Taq<sup>q</sup>I digestion (Fig. 2B). While the hemi-5-hmC can be cut, hemi-N<sub>3</sub>-5-gmC cannot block digestion as well as the fully-modified one (Fig. 2B, lane 1-4). Even with the bulkier group, biotin-N<sub>3</sub>-5-gmC, present, the majority of DNA was still digested (Fig. 2B, lane 5-6). Thus, Taq<sup>q</sup>I digests the hemi-modified sequence but is blocked by the fully-modified one

with biotin-N<sub>3</sub>-5-gmC. This noticeable difference of Taq<sup>q</sup>I in response to fully- and hemi-hydroxymethylation states after modification provides a method to distinguish these two states on TCGA sites.

To further investigate if the difference in digestion between fully and hemi-hydroxymethylation states is universal in restriction enzymes, we replaced the T\*CGA site with C\*CGG (\*C = 5-hmC) in the previous 32-mer duplex DNA and performed the same assays on MspI (Fig. 3). Both fully and hemi-5-hmC can be cut by MspI completely before modification (Fig. 3A, lane 1-2 and Fig. 3B, lane 1-2). For a fully 5-hmC site, both N<sub>3</sub>-gmC and biotin-N<sub>3</sub>-5-gmC can block MspI digestion completely as expected (Fig. 3A, lane 3-6), suggesting that MspI is more sensitive towards the steric hindrance of cytosine modification of the CpG dinucleotide; the presence of a glucose group on the inner C is enough for the protection from digestion. For the hemi-5-hmC site, it gave the same results as the fully-modified one: both the N<sub>3</sub>-5-gmC and biotin-N<sub>3</sub>-5-gmC can be protected from digestion (Fig. 3B, lane 3-6). This result is in accordance with our assumption that MspI is much easier to block and that it cannot distinguish the fully-modified 5-hmC site from hemi-5-hmC site as Taq<sup>q</sup>I.

In conclusion, we discovered the formerly known methyl-insensitive restriction enzyme Taq<sup>q</sup>I can be blocked by a biotin-N<sub>3</sub>-5-gmC modification in order to detect the existence of 5-hmC on TCGA sites in CGRA. Furthermore, Taq<sup>q</sup>I is unique in that it is almost completely inhibited by the fully-modified biotin-N<sub>3</sub>-gmC site while cutting most of the hemi-modified site, therefore may be used to distinguish fully versus hemi-hydroxymethylation sites on specific loci. Given the possible importance of 5-hmC in gene regulation, the development of this detection method will greatly facilitate future epigenetic studies. Work to employ this assay to discover 5-hmC in real biological samples is currently underway in our laboratory.

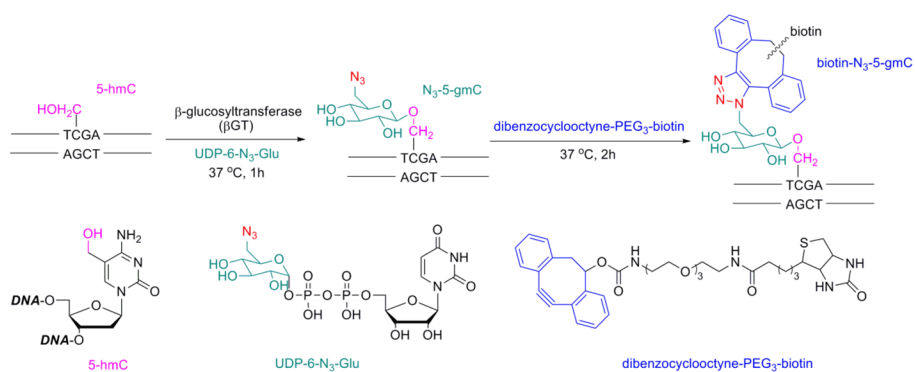
## Acknowledgments

This work was partially supported by the National Institutes of Health (GM 071440 to C.H.). We thank Sarah Frank Reichard, M.A. for help with editing the manuscript.

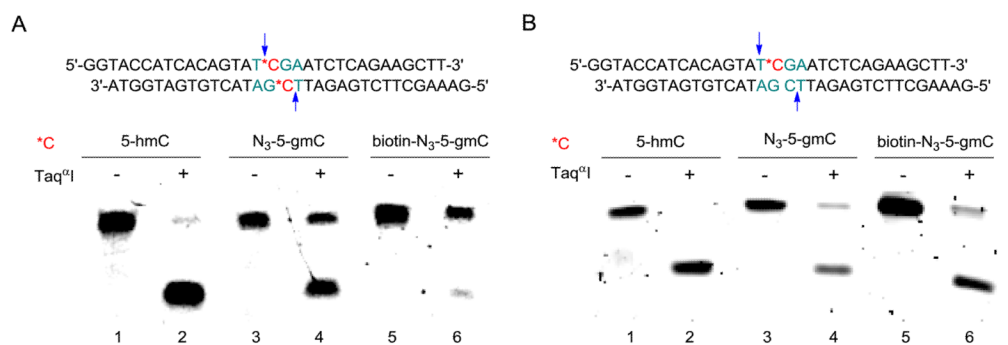
## References

1. Goll MG, Bestor TH. *Annu Rev Biochem.* 2005; 74:481. [PubMed: 15952895]
2. Reik W. *Nature.* 2007; 447:425. [PubMed: 17522676]
3. Gal-Yam EN, Saito Y, Egger G, Jones PA. *Annu Rev Med.* 2008; 59:267. [PubMed: 17937590]
4. Kriaucionis S, Heintz N. *Science.* 2009; 324:929. [PubMed: 19372393]
5. 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. [PubMed: 19372391]
6. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. *Nature.* 2010; 466:1129. [PubMed: 20639862]
7. 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. [PubMed: 21057493]
8. 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. [PubMed: 21295276]
9. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. *PLoS One.* 2010; 5:e8888. [PubMed: 20126651]
10. Jin SG, Kadam S, Pfeifer GP. *Nucleic Acids Res.* 2010; 38:e125. [PubMed: 20371518]
11. Nestor C, Ruzov A, Meehan RR, Dunican DS. *BioTechniques.* 2010; 48:317. [PubMed: 20569209]

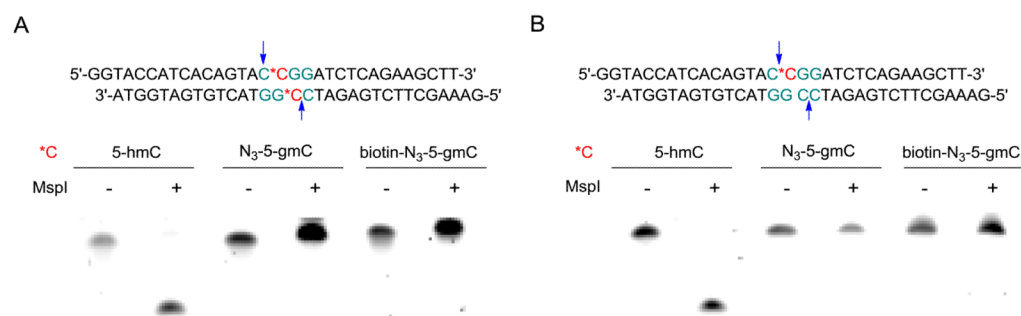
12. Zheng Y, Cohen-Karni D, Xu D, Chin HG, Wilson G, Pradhan S, Roberts RJ. *Nucleic Acids Res.* 2010; 38:5527. [PubMed: 20444879]
13. Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen C-H, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C. *Nat Biotechnol.* 2011; 29:68. [PubMed: 21151123]
14. Singer-Sam J, Grant M, LeBon JM, Okuyama K, Chapman V, Monk M, Riggs AD. *Mol Cell Biol.* 1990; 10:4987. [PubMed: 1697035]
15. Tardy-Planechaud S, Fujimoto J, Lin SS, Sowers LC. *Nucleic Acids Res.* 1997; 25:553. [PubMed: 9016595]
16. Szwagierczak A, Brachmann A, Schmidt CS, Bultmann S, Leonhardt H, Spada F. *Nucleic Acids Res.* 2011; 10.1093/nar/gkr118
17. Huang LH, Farnet CM, Ehrlich KC, Ehrlich M. *Nucleic Acids Res.* 1982; 10:1579. [PubMed: 6280151]
18. Davis T, Vaisvila R. *J Vis Exp.* 2011; 4810.3791/2661
19. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB. *Angew Chem, Int Ed.* 2002; 41:2596.
20. Sletten EM, Bertozzi CR. *Angew Chem, Int Ed.* 2009; 48:6974.
21. Speers AE, Cravatt BF. *Chem Biol.* 2004; 11:535. [PubMed: 15123248]
22. Ning X, Guo J, Wolfert MA, Boons GJ. *Angew Chem, Int Ed.* 2008; 47:2253.
23. Jewett JC, Bertozzi CR. *Chem Soc Rev.* 2010; 39:1272. [PubMed: 20349533]



**Figure 1.** The βGT-catalyzed formation of N<sub>3</sub>-5-gmC and the subsequent click chemistry to yield biotin-N<sub>3</sub>-5-gmC on the TCGA site in duplex DNA. Modification on only one strand is shown.

**Figure 2.**

Taq<sup>q</sup>I-mediated digestion of 5-hmC-, N<sub>3</sub>-5-gmC-, and biotin-N<sub>3</sub>-5-gmC-containing DNA with the sequences showing on top. \*C indicates the modified position; arrows indicate Taq<sup>q</sup>I cutting sites. (A), Digestion of fully-modified DNA. (B), Digestion of hemi-modified DNA. The 32-mer dsDNA (1 pmol) was digested with 100 U of Taq<sup>q</sup>I (New England BioLabs) for 1 hr at 65 °C. Samples were analyzed by 16% PAGE/Urea gel and visualized using SYBR Green I staining (Lumiprobe).

**Figure 3.**

MspI digestion of 5-hmC-, N<sub>3</sub>-5-gmC-, and biotin-N<sub>3</sub>-5-gmC-containing DNA with the sequences showing on top. \*C indicates the modified position; arrows indicate MspI cutting sites. (A), Digestion of fully-modified DNA. (B), Digestion of hemi-modified DNA. The 32-mer dsDNA (1 pmol) was digested with 100 U of MspI (New England BioLabs) for 1 hr at 37 °C. Samples were analyzed by 16% PAGE/Urea gel and visualized using SYBR Green I staining (Lumiprobe).