DNA Modification

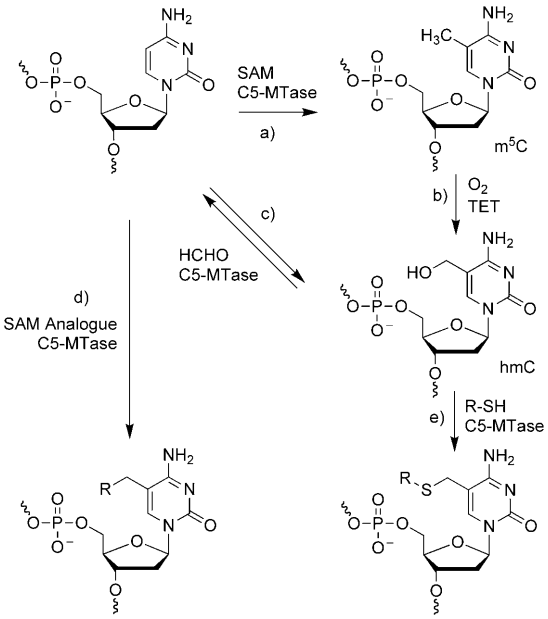
DOI: 10.1002/anie.201100350

Enzymatic Labeling of 5-Hydroxymethylcytosine in DNA**

*Claudia Höbartner**

DNA methylation · epigenetics · labeling · nucleobases · transferase

DNA methylation is an important epigenetic modification that plays crucial roles for regulation of gene expression and chromatin structure. The most abundant natural DNA modification in vertebrates is the methylation of cytosine at position C5 (m^5C). The methyl group is enzymatically installed by cytosine-5 methyltransferases (C5-MTases) that use *S*-adenosyl-L-methionine (SAM or AdoMet) as methyl-group donor (Scheme 1a). Misregulated DNA methylation has been connected to human diseases, including neuro-developmental disorders and cancer.[1] The desire to understand the mechanisms and consequences of DNA modification has been further stimulated by recent reports on 5-hydroxymethylcytosine (hmC) as an additional cytosine modification in mammalian DNA. The presence of hmC in animal DNA was suggested in the early 1970s[2] but has since received little scientific attention. In 2009, Kriaucionis and Heintz found hmC in cerebellar purkinje neurons,[3] and Tahiliani et al. reported the presence of hmC in mouse embryonic stem cells and human embryonic kidney cells and showed that genomic m^5C can be converted to hmC by oxygenase enzymes of the TET family (ten–eleven translocation gene; Scheme 1b).[4] These 2-oxoglutarate and Fe^{II}-dependent enzymes are also able to oxidize the methyl group of m^5C in vitro. Interestingly, prokaryotic C5-MTases have been shown to generate hmC in vitro by reversible addition of formaldehyde to cytosines (Scheme 1c).[5] The functional relevance of hmC as another level of epigenetic control in mammalian genomes and as an intermediate in oxidative demethylation is the subject of current intense investigation.[6]



Scheme 1. a, b) Natural and c–e) artificial pathways for enzymatic DNA modification of cytosine at position C5.

Understanding the role of hmC in biological pathways requires reliable quantification and localization of hmC in genomic DNA. This task has been technically challenging, since m^5C and hmC are not easily distinguishable by standard biochemical methods. The most prominent technique for analysis of DNA methylation patterns, bisulfite sequencing,[7] leads to identical readouts for m^5C and hmC.[8] Moreover, methylation-specific restriction enzymes cannot reliably distinguish m^5C and hmC. This situation illustrates the need for specific detection methods for hmC.

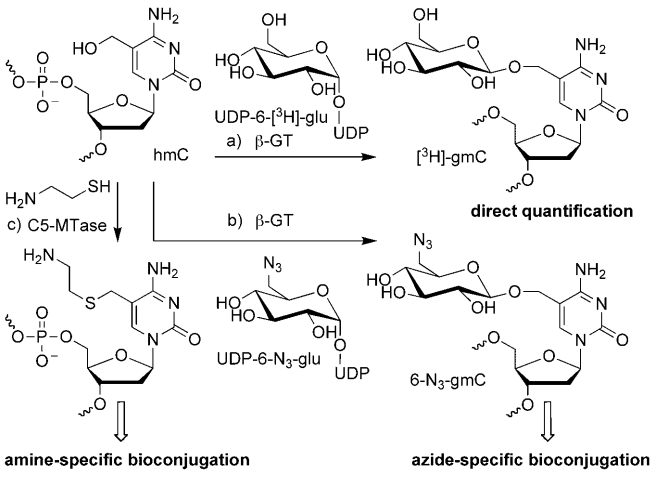
In the 2009 reports of hmC in mammalian DNA,[3,4] the modified nucleoside was detected by TLC and HPLC analysis of digested DNA as an additional compound that was characterized by ESI-MS/MS. A detailed analysis of hmC modification levels in different sections of the mouse brain was reported by Carell and co-workers, who developed a quantitative MS method using chemically synthesized stable isotope-labeled reference compounds.[9] The high sensitivity of mass spectrometry allowed the precise quantification of modified cytosine levels in various mouse tissues. The cellular level of hmC ranges from 0.03 to 0.7% relative to guanine, whereas m^5C is uniformly distributed in all cells at a level of 4–5%.[6] To visualize modified cytosines in different cellular locations, specific antibodies for m^5C, hmC, and the bisulfite adduct of hmC have been reported. These antibodies were also used in quantitative dot-blot assays for the determination of modification levels in genomic DNA from human bone marrow samples.[10] Mass spectrometry and immunostaining

[*] Dr. C. Höbartner
Research Group Nucleic Acid Chemistry
Max Planck Institute for Biophysical Chemistry
Am Fassberg 11, 37077 Göttingen (Germany)
Fax: (+49) 551-201-1680
E-mail: claudia.hoebartner@mpibpc.mpg.de

[**] The Max Planck Society is gratefully acknowledged for generous support of our work.

methods, however, cannot provide single-base-resolution information on the precise location of hmC in the DNA sequence. In this respect, recent proof-of-principle studies on single-molecule sequencing show promising developments. Single-molecule real-time (SMRT) sequencing based on sequencing-by-synthesis has been shown to discriminate cytosine, m^5C, and hmC.[11] Efforts towards application of nanopore amperometry using protein or solid-state nanopores for modification-specific sequencing have also been reported.[12] To date, these methods have only been demonstrated on synthetic DNA samples with known modifications. Further developments are required for direct sequencing of genomic DNA. An important challenge is the selective enrichment of hmC-containing DNA. In this context, bacteriophage and bacterial enzymes have recently been used for selective chemoenzymatic labeling of hmC residues.

In T-even bacteriophages, hmC nucleotides are collectively glycosylated by glycosyltransferases using UDP-glucose as glycosyl donor. Two research groups independently reported the application of T4 β-glycosyltransferase (β-GT) for the development of new hmC quantification methods. In the first implementation, Leonhardt and co-workers applied UDP-[^{3}H]-glucose to radioactively label hmC nucleotides (Scheme 2a).[13] Using appropriate calibration curves, scintillation counting revealed the abundance of hmC in different mouse tissues, although at higher levels than measured by quantitative mass spectrometry.[6] In a second report, He, Jin, and co-workers used β-GT to install a chemically modified, azide-containing glucose derivative at the hydroxy group of hmC to yield 6-N_3-gmC (Scheme 2b).[14] The azido group was further derivatized by a copper-free [3+2] cycloaddition reaction with dibenzocyclooctyne-linked biotin. Both glycosylation and click labeling were efficient reactions that allowed for selective enrichment of hmC-containing genomic DNA from different cell lines and mouse tissues. The streptavidin-bound biotinylated DNA product was also shown to block polymerase extension, thus providing the potential for determination of hmC locations with nucleotide resolution.



Scheme 2. Enzymatic labeling of hmC in DNA. T4 β-GT transfers radioactive [^{3}H]-glucose (a) or azide-modified glucose (b) to the hydroxymethyl group. The glycoslyated gmC-modified DNA can be either directly detected[13] or further derivatized using azide-specific conjugation reactions.[14] Bacterial C5-MTases mediate derivatization of hmC with sulfur nucleophiles, for example, cysteamine (c), which introduces a reactive amino group for further modification.[15]

A recent addition to the previously described methods is based on bacterial methyltransferase enzymes, as reported by Klimašauskas and co-workers.[15] MTases had been known to accept synthetic SAM analogues to install artificial functional groups at cytosines (Scheme 1d).[16] In the course of mechanistic studies on DNA-methylating enzymes, it was found that C5-MTases can use non-cofactor-like small molecules as substrates. This finding led to the discovery that hmC can be generated in vitro directly from unmodified DNA upon the action of methyltransferases using formaldehyde instead of SAM.[5]

As another unexpected activity involving non-cofactor-like substrates, bacterial MTases (M.HhaI and M.SssI) have now been shown to use sulfur and selenium nucleophiles to displace the hydroxy group of hmC and form a new thio- or selenoether functionality (Scheme 1e).[15] This activity requires a conserved cysteine in the enzyme's active site and is thought to be mechanistically related to MTase-catalyzed methylation involving a covalently bound enzyme–DNA intermediate. This finding is interesting from the mechanistic point of view and is potentially applicable as a tool for selective derivatization of hmC residues in natural DNA. As proof of principle, hmC-containing plasmid DNA was derivatized with cysteamine, thereby installing a primary amino functionality (Scheme 2c) that was further conjugated with biotin NHS ester (NHS = *N*-hydroxysuccinimide). The product was affinity-captured using streptavidin beads, thus demonstrating selective enrichment of hmC-containing DNA. It remains to be demonstrated that the sequence-specificity of bacterial MTases and the potential reversibility of the hmC modification in the absence of the natural cofactor do not limit the accuracy of modification analyses in natural DNA samples. From a different perspective, however, this novel activity of C5-MTases to covalently modify hydroxymethyl groups in DNA offers exciting new opportunities for sequence-specific derivatization and labeling of DNA, which is also of current interest in the context of constructing functionalized DNA architectures.

In summary, enzymatic labeling of hmC in DNA adds new aspects to site-specific targeting of DNA and to the investigation of epigenetic DNA modifications. Although quantitative mass spectrometry is currently one of the most widely used and highly reliable methods for accurate quantification of natural DNA modification levels, enzymatic labeling methods are expected to contribute valuable information in the future. An enzyme-based epigenetic analysis kit that combines T4 β-GT and modification-specific restriction enzymes with quantitative PCR analysis has recently become commercially available.[17] Innovative methods will be highly useful for investigating the abundance and genomic distribution of hmC, which has also been implicated in embryonic stem cell maintenance[18] and reprogramming of the paternal genome in the early life cycle of mammals.[19] Insights into the dynamic regulation of DNA methylation will contribute to a deeper understanding of changing epigenetic profiles during

development. With the recent findings of mutant TET proteins involved in myeloid cancers and impaired hydroxylation of m^5C,[10] new hmC detection methods may also add valuable tools for disease diagnosis.

Received: January 14, 2011
Revised: February 22, 2011
Published online: April 11, 2011

[1] a) E. N. Gal-Yam, Y. Saito, G. Egger, P. A. Jones, *Annu. Rev. Med.* **2008**, *59*, 267; b) S. Kriaucionis, A. Bird, *Hum. Mol. Genet.* **2003**, *12 Suppl 2*, 221R.
[2] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781.
[3] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929.
[4] 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.
[5] Z. Liutkevičiūtė, G. Lukinavicius, V. Masevičius, D. Daujotytė, S. Klimašauskas, *Nat. Chem. Biol.* **2009**, *5*, 400.
[6] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Bruckl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367.
[7] Bisulfite sequencing is based on selective chemical deamination of unmethylated cytosine to uracil by sodium bisulfite. The modified cytosine analogue 5-methylcytosine remains unchanged and is read as cytosine during sequencing. hmC reacts with bisulfite to yield cytosine 5-methylenesulfonate, which does not promote deamination and therefore also codes as cytosine.
[8] a) S. G. Jin, S. Kadam, G. P. Pfeifer, *Nucleic Acids Res.* **2010**, *38*, e125; b) Y. Huang, W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLoS One* **2010**, *5*, e8888.
[9] M. Münzel, D. Globisch, T. Bruckl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem.* **2010**, *122*, 5503; *Angew. Chem. Int. Ed.* **2010**, *49*, 5375.
[10] M. Ko, Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, A. Rao, *Nature* **2010**, *468*, 839.
[11] B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, *Nat. Methods* **2010**, *7*, 461.
[12] a) E. V. Wallace, D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, T. J. Donohoe, H. Bayley, *Chem. Commun.* **2010**, *46*, 8195; b) M. Wanunu, D. Cohen-Karni, R. R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M. L. Klein, M. Drndic, *J. Am. Chem. Soc.* **2011**, *133*, 486.
[13] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res.* **2010**, *38*, e181.
[14] 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.* **2010**, *29*, 68.
[15] Z. Liutkevičiūtė, E. Kriukienė, I. Grigaitytė, V. Masevičius, S. Klimašauskas, *Angew. Chem.* **2011**, *123*, 2138; *Angew. Chem. Int. Ed.* **2011**, *50*, 2090.
[16] a) C. Dalhoff, G. Lukinavicius, S. Klimašauskas, E. Weinhold, *Nat. Chem. Biol.* **2006**, *2*, 31; b) G. Lukinavicius, V. Lapiene, Z. Stasevskij, C. Dalhoff, E. Weinhold, S. Klimašauskas, *J. Am. Chem. Soc.* **2007**, *129*, 2758.
[17] EpiMark hmC and m^5C analysis kit from New England Biolabs (NEB E3317).
[18] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129.
[19] K. Iqbal, S. G. Jin, G. P. Pfeifer, P. E. Szabó, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3642.

