Hydroxymethylcytosine

DOI: 10.1002/anie.201002033

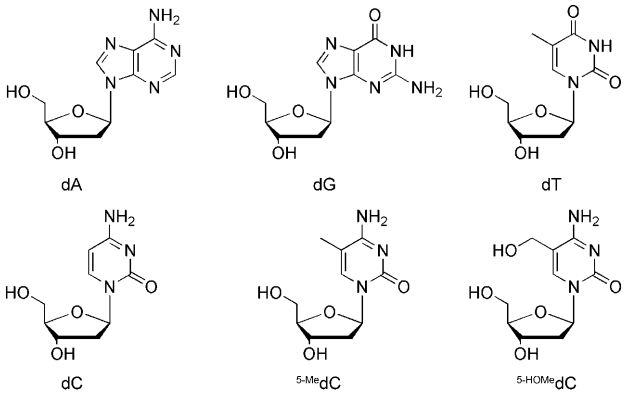
Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain**

*Martin Münzel, Daniel Globisch, Tobias Brückl, Mirko Wagner, Veronika Welzmiller, Stylianos Michalakis, Markus Müller, Martin Biel, and Thomas Carell**

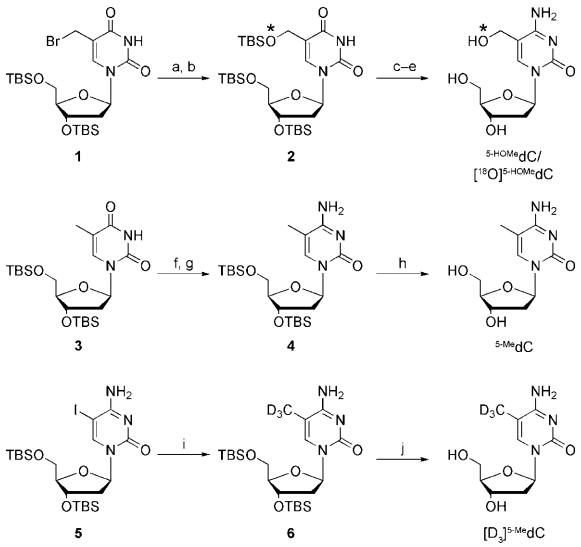
The genetic code is established by the sequence of the four canonical DNA nucleosides dA, dC, dG, and dT.[1] Of these four bases, only the dC base is chemically modified inside cells of higher organisms to control transcriptional activity.[2] Special methyltransferases replace the H atom at position 5 by a methyl group to form methylcytosine ($^{5\text{-Me}}$dC).[3] Methylation occurs only in CpG sequences and is mostly responsible for the silencing of genes.[4] In two recent publications 5-hydroxymethylcytosine ($^{5\text{-HOMe}}$dC) was established as a new post-replicatively formed DNA nucleoside (Scheme 1). Kriaucionis and Heintz detected $^{5\text{-HOMe}}$dC in cerebellar purkinje neurons.[5] Tahiliani et al. reported the presence of traces of $^{5\text{-HOMe}}$dC (ca. 0.032% of all nucleosides) in mouse embryonic stem cells and observed the new base in CpG sequences in human embryonic kidney (HEK) cells when they overexpressed the hydroxylating 2-oxoglutarate and Fe^{II}-dependent enzyme TET in these cells.[6] It was additionally shown that these TET enzymes are able to oxidize the 5-methyl group of $^{5\text{-Me}}$dC to give the hydroxymethyl group in vitro. The new base was detected using thin-layer chromatography after radioactive labeling of the nucleotides. The function of $^{5\text{-HOMe}}$dC is currently unclear, but it is speculated that it may establish another level of transcriptional control or that it is an intermediate of a putative oxidative demethylation mechanism.[7]

We developed a quantitative LC-MS method to investigate the distribution of $^{5\text{-HOMe}}$dC in mammal brains and to determine the relative quantities of $^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC. To this end, we synthesized both nucleosides in natural and isotope-labeled forms[8] (Scheme 2) and quantified their amounts in different mouse brain tissues.

$^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC were labeled as ^{18}O and CD_3 derivatives, respectively. The synthesis of $^{5\text{-HOMe}}$dC started



Scheme 1. Structure of the four canonical nucleosides and of the post-replicatively formed bases $^{5\text{-Me}}$dC and $^{5\text{-HOMe}}$dC.



Scheme 2. Syntheses of the isotope-labeled nucleosides $^{5\text{-HOMe}}$dC, $[^{18}O]^{5\text{-HOMe}}$dC, $^{5\text{-Me}}$dC, and $[D_3]^{5\text{-Me}}$dC. a) H_2O, DIPEA, 72%, b) TBSCl, imid., 46%, c) NaH, TPSCl, d) NH_3/MeOH, 76% over two steps, e) $3\,HF{\cdot}NEt_3$, 52%, f) NaH, TPSCl, g) NH_3/MeOH, 81% over two steps, h) HF·pyr, 88%, i) CD_3MgI, CuCl, $[Pd(PPh_3)_4]$, 90% inseparable mixture of **6** and TBSdC, j) HF·pyr, 84% (based on pure **6**).$[^{18}O]^{5\text{-HOMe}}$dC was synthesized by the same route as $^{5\text{-HOMe}}$dC. The asterisk (*) indicates the ^{18}O label. DIPEA = *N*,*N*-diisopropylethylamine, imid. = imidazole, pyr = pyridine, TBS = *tert*-butyldimethylsilyl, TPS = 2,4,6-triisopropylbenzenesulfonyl.

[*] Dipl.-Chem. M. Münzel,[+] Dipl.-Chem. D. Globisch,[+] Dipl.-Chem. T. Brückl, Dipl.-Chem. M. Wagner, Dipl.-Chem. V. Welzmiller, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPS^M) at the Department of Chemistry, Ludwig-Maximilians-University
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@cup.uni-muenchen.de

Dr. S. Michalakis, Prof. Dr. M. Biel
Center for Integrated Protein Science (CiPS^M) at the Department of Pharmacy, Ludwig-Maximilians-University
Butenandtstrasse 5–13, 81377 Munich (Germany)

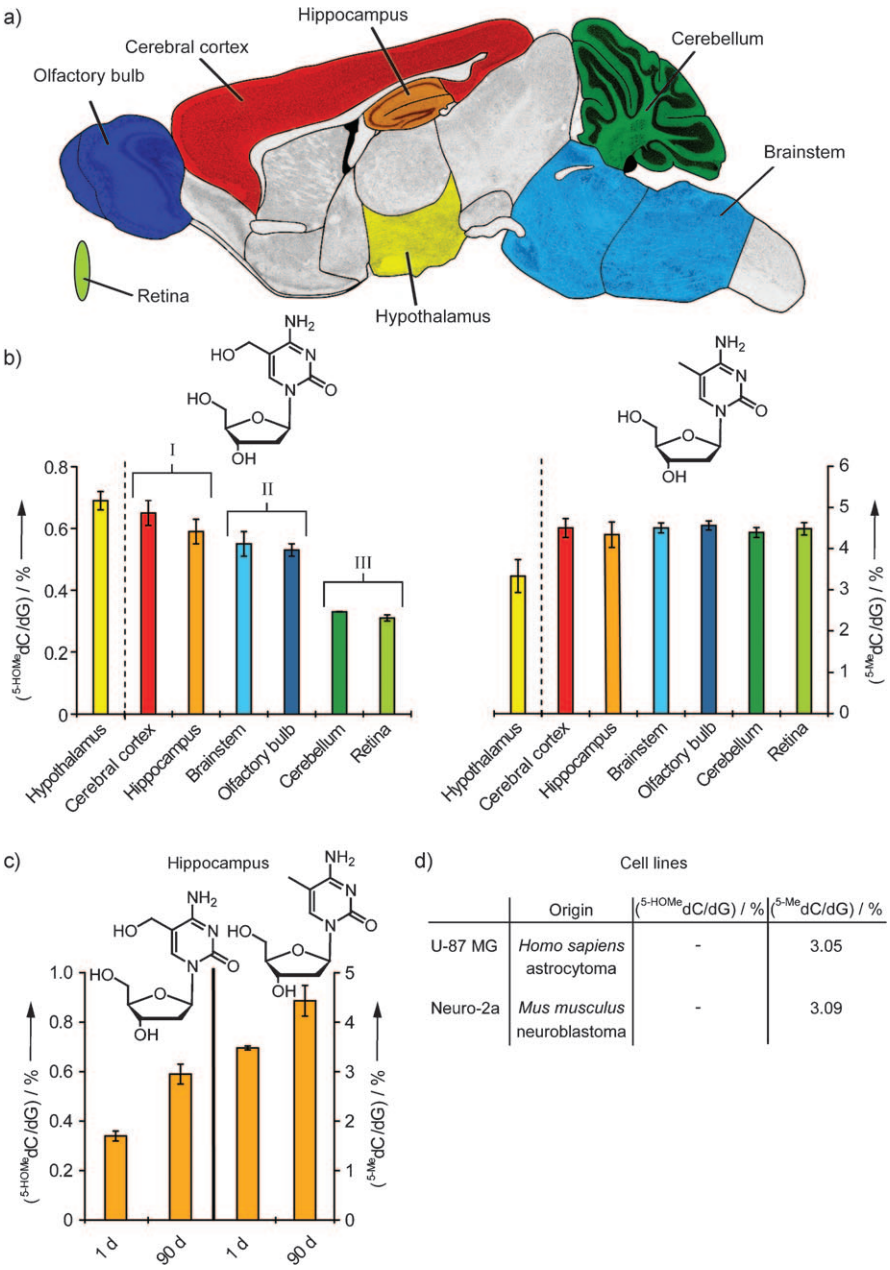
[+] These authors contributed equally to this work.

[**] We thank the excellence cluster CiPS^M, SFB 646, and SFB 749 for generous support. M. Münzel is grateful to the Fonds of the Chemical Industry for a Kekulé fellowship. We thank V. Hammelmann for support during this project.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201002033.

with the TBS-protected bromo-dT derivative **1**.[9] S_N2 substitution with H_2O followed by protection with a TBS group furnished compound **2**. $H_2{}^{18}O$ was used for the labeled compound. Subsequent conversion of **2** into the corresponding dC derivative by activation as a sulfonate[10] and subsequent replacement with NH_3 yielded $^{5\text{-HOMe}}$dC after final deprotection. For $^{5\text{-Me}}$dC we used TBS-protected dT and converted it into TBS-protected $^{5\text{-Me}}$dC by using the strategy described above. The isotope-labeled **6** was prepared from TBS-protected iodocytosine **5**[11] by palladium-catalyzed methylation at C5. Final deprotection furnished $[D_3]^{5\text{-Me}}$dC. These nucleosides were used to generate mass spectrometry calibration curves to enable precise LC-MS quantification (see the Supporting Information).[12] The obtained areas of the ion currents at the exact masses of each nucleoside were plotted against different concentration ratios of natural to labeled nucleosides. In both cases linear plots were obtained with R^2 values of 0.999.

We next used the calibration curve equations to analyze the content of $^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC in various mouse brain tissues (Figure 1 a). To this end, we collected the hypothalamus, cortex, hippocampus, olfactory bulb, brainstem, cerebellum, and retina from four 90 day old mice. The tissues were homogenized and the DNA isolated by phenol/chloroform extraction (see the Supporting Information). From the obtained DNA, 4–10 μg were completely digested to the nucleosides in a two-step procedure by incubating first with nuclease S1 at 37 °C for 3 h. In the second incubation step phosphodiesterase I and antarctic phosphatase were used at 37 °C for 3 h to enable total digestion of the DNA. We then added the isotope-labeled compounds and analyzed the nucleoside mixture by HPLC and using a high-resolution mass spectrometer as detector (Thermo Finnigan LTQ Orbitrap XL). In all experiments $^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC eluted with retention times of 12.3 min and 18.5 min, respectively. One signal for the natural (light) and one for the synthetic (heavy) compound was detected in each experiment. Quantification was performed by comparing the integrals of the ion current of the natural compound (determined amount) with their corresponding heavy atom labeled derivatives (known amounts) by using the calibration curve equations. Repetition of the experiments allowed us to determine an average error for the quantification data of ± 5 % for $^{5\text{-Me}}$dC and $^{5\text{-HOMe}}$dC. All the data for each mouse are listed in the Supporting Information. Figure 1 shows the averaged value.



	Origin	($^{5\text{-HOMe}}$dC/dG) / %	($^{5\text{-Me}}$dC/dG) / %
U-87 MG	*Homo sapiens* astrocytoma	-	3.05
Neuro-2a	*Mus musculus* neuroblastoma	-	3.09

Figure 1. a) Sagittal section of a mouse brain. The brain areas highlighted in color were studied here. b) Ratio of $^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC to dG in the different tissues in percent. dG was chosen as a reference, because it forms base pairs with dC, $^{5\text{-HOMe}}$dC, and $^{5\text{-Me}}$dC in DNA. I–III: see text. c) Ratio of $^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC to dG in the hippocampus of 1 day and 90 day old mice in percent. d) Ratio of $^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC to dG in a neuronal (Neuro-2a) and a glial (U-87 MG) cell line in percent.

The first discovery (Figure 1 b) is that the new base $^{5\text{-HOMe}}$dC is clearly detected in all brain tissues. Between 0.3 % and 0.7 % of all the dC nucleosides are hydroxymethylated at position 5, which clearly establishes $^{5\text{-HOMe}}$dC as a new post-replicatively formed nucleoside in the brain. The second

surprising result is that the amount of $^{5\text{-HOMe}}$dC varies significantly in the investigated tissues. The previous study[5] found $^{5\text{-HOMe}}$dC predominantly in purkinje neurons, which are found in the cerebellum. We now see that the amount of the base is much larger in the cortex and hippocampus, where purkinje cells are not present. Analysis of brain cancer cell lines (Figure 1d) revealed that the new base is absent in neuroblastome and glial cell lines. We found, however, and in agreement with the literature, reduced $^{5\text{-Me}}$dC content (from a typical 4.5% to 3%) in these cancer cell lines.[13]

By using our data we can roughly divide the mouse brain into three different areas (Figure 1b). Most $^{5\text{-HOMe}}$dC is found in the hippocampus and cortex (I), which are the brain areas that have higher cognitive functions. The brainstem and olfactory bulb form a second category, which possess intermediate $^{5\text{-HOMe}}$dC levels (II). The cerebellum and retina contain significantly less $^{5\text{-HOMe}}$dC (III). Surprisingly, initial data also show a relatively high level of $^{5\text{-HOMe}}$dC in the hypothalamus, which is part of the endocrine system that controls hormone-based processes. The distribution of $^{5\text{-HOMe}}$dC was confirmed by immunolocalization (see the Supporting Information).

The relative amounts of $^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC are also interesting. We are unable to correlate the amount of $^{5\text{-HOMe}}$dC and $^{5\text{-Me}}$dC in these tissues, which one would expect if $^{5\text{-HOMe}}$dC is exclusively generated from $^{5\text{-Me}}$dC as a precursor. The $^{5\text{-Me}}$dC values are significantly lower in the hypothalamus and here we indeed detect high $^{5\text{-HOMe}}$dC values. In contrast, and in accord with the literature,[14] we found that the $^{5\text{-Me}}$dC values are stable at a typical value of around 4.5% in all other tissues, while the $^{5\text{-HOMe}}$dC values vary significantly. This suggests that $^{5\text{-HOMe}}$dC has a function that is not correlated with the $^{5\text{-Me}}$dC value.

Finally, we asked the question how the age of the animal influences the $^{5\text{-Me}}$dC and $^{5\text{-HOMe}}$dC values. To this end, we analyzed the hippocampus tissue of one-day old mice (Figure 1c). We indeed detected a significantly lower level of both $^{5\text{-Me}}$dC and $^{5\text{-HOMe}}$dC. The $^{5\text{-Me}}$dC value increases with age from $(3.5 \pm 0.1)\%$ to $(4.3 \pm 0.3)\%$. More significantly, the amount of $^{5\text{-HOMe}}$dC is raised in 90 day old mice by approximately 75% from $(0.34 \pm 0.02)\%$ to $(0.59 \pm 0.04)\%$. To exclude that $^{5\text{-HOMe}}$dC accumulates because of oxidative stress in older animals we also analyzed the damaged base 8-oxodG, which is a typical oxidative stress marker. Here, however, a difference could not be detected, which shows that $^{5\text{-HOMe}}$dC in the brain does not result from an accumulation of oxidative DNA damage.[15]

In summary, we have quantified the amount of $^{5\text{-Me}}$dC and $^{5\text{-HOMe}}$dC in brain tissues with excellent accuracy and we confirmed $^{5\text{-HOMe}}$dC as a new post-replicatively formed nucleoside. We were able to show that the base is widely distributed in the brain and that the nucleoside is particularly prominent in those brain tissues which are involved in higher cognitive functions. Finally, we obtained initial data that show that the $^{5\text{-HOMe}}$dC nucleoside is present in new born mice at reduced levels.

Received: April 6, 2010
Revised: May 3, 2010
Published online: June 25, 2010

Keywords: demethylation · DNA methylation · DNA modification · epigenetics · hydroxymethylcytosine

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
[2] a) A. Bird, *Nature* **2007**, *447*, 396–398; b) P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068–1070.
[3] M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **2005**, *74*, 481–514.
[4] a) A. P. Feinberg, *Nature* **2007**, *447*, 433–440; b) W. Reik, *Nature* **2007**, *447*, 425–432.
[5] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
[6] 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.
[7] a) C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580–583; b) Z. Liutkevičiūtė, G. Lukinavičius, V. Masevičius, D. Daujotytė, S. Klimašauskas, *Nat. Chem. Biol.* **2009**, *5*, 400–402; c) J.-K. Zhu, *Annu. Rev. Genet.* **2009**, *43*, 143–166.
[8] For an alternative synthesis, see C. J. LaFrancois, J. Fujimoto, L. C. Sowers, *Chem. Res. Toxicol.* **1998**, *11*, 75–83.
[9] a) R. K. Grover, S. J. K. Pond, Q. Cui, P. Subramaniam, D. A. Case, D. P. Millar, P. Wentworth, *Angew. Chem.* **2007**, *119*, 2897–2901; *Angew. Chem. Int. Ed.* **2007**, *46*, 2839–2843.
[10] N. Bischofberger, M. D. Matteucci, *J. Am. Chem. Soc.* **1989**, *111*, 3041–3046.
[11] a) C. H. Hwang, J. S. Park, J. H. Won, J. N. Kim, E. K. Ryu, *Arch. Pharmacal Res.* **1992**, *15*, 69–72; b) M. Kuwahara, J.-I. Nagashima, M. Hasegawa, T. Tamura, R. Kitagata, K. Hanawa, S.-I. Hososhima, T. Kasamatsu, H. Ozaki, H. Sawai, *Nucleic Acids Res.* **2006**, *34*, 5383–5394.
[12] a) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem.* **2009**, *121*, 8074–8077; *Angew. Chem. Int. Ed.* **2009**, *48*, 7932–7934; b) E. P. Quinlivan, J. F. Gregory, *Nucleic Acids Res.* **2008**, *36*, e119; c) Y. Yang, D. Nikolic, S. M. Swanson, R. B. van Breemen, *Anal. Chem.* **2002**, *74*, 5376–5382.
[13] M. Esteller, *N. Engl. J. Med.* **2008**, *358*, 1148–1159.
[14] a) M. A. Gama-Sosa, R. M. Midgett, V. A. Slagel, S. Githens, K. C. Kuo, C. W. Gehrke, M. Ehrlich, *Biochim. Biophys. Acta Gene Struct. Expression* **1983**, *740*, 212–219; b) B. F. Vanyushin, N. G. Lopatina, C. K. Wise, F. R. Fullerton, L. A. Poirier, *Eur. J. Biochem.* **1998**, *256*, 518–527.
[15] C. Bienvenu, J. R. Wagner, J. Cadet, *J. Am. Chem. Soc.* **1996**, *118*, 11406–11411.

