

# Formation and Abundance of 5-Hydroxymethylcytosine in RNA

Sabrina M. Huber,<sup>[a]</sup> Pieter van Delft,<sup>[a]</sup> Lee Mendil,<sup>[b]</sup> Martin Bachman,<sup>[b]</sup> Katherine Smollett,<sup>[c]</sup> Finn Werner,<sup>[c]</sup> Eric A. Miska,<sup>[d]</sup> and Shankar Balasubramanian\*<sup>[a, b]</sup>

RNA methylation is emerging as a regulatory RNA modification that could have important roles in the control and coordination of gene transcription and protein translation. Herein, we describe an *in vivo* isotope-tracing methodology to demonstrate that the ribonucleoside 5-methylcytidine ( $m^5C$ ) is subject to oxidative processing in mammals, forming 5-hydroxymethylcytidine ( $hm^5C$ ) and 5-formylcytidine ( $f^5C$ ). Furthermore, we have identified  $hm^5C$  in total RNA from all three domains of life and in polyA-enriched RNA fractions from mammalian cells. This suggests  $m^5C$  oxidation is a conserved process that could have critical regulatory functions inside cells.

Half a century of research has identified more than 140 distinct post-transcriptional modifications of ribonucleic acids.<sup>[1]</sup> The majority of these chemical modifications are found in the highly abundant transfer and ribosomal RNAs (tRNA, rRNA, 95% of total RNA) where they contribute to RNA folding, provide structural stability and diversify molecular recognition, for example, through “wobble base pairing”.<sup>[2]</sup> Recently RNA modifications have also been identified in the less abundant messenger and long noncoding RNAs (mRNA, lncRNA).<sup>[3,4]</sup> Given the short half-life of many RNA species ( $t_{1/2} < 10$  h), RNA modifications previously have been viewed largely as static and stable marks. Recent studies, however, have shown that they can be dynamic and might have critical regulatory and/or signalling roles inside cells, similar to the reversible epigenetic marks on DNA and histones.<sup>[5]</sup> For instance,  $N^6$ -methyladenosine ( $m^6A$ ), the most abundant internal modification in eukary-

otic mRNA, undergoes oxidative processing to  $N^6$ -hydroxymethyladenosine and  $N^6$ -formyladenosine by the fat mass and obesity-associated protein (FTO) and ALKBH5, respectively.<sup>[6,7]</sup> Although the function of this dynamic adenosine methylation is not yet fully understood, it is proposed to be involved in the regulation of gene expression, mRNA stability and protein translation.<sup>[8,9]</sup> 5-Methylcytidine ( $m^5C$ ) is another known methylated ribonucleoside that is widespread across both coding and noncoding RNAs from all domains of life.<sup>[2]</sup> Although 5-methyl-2'-deoxycytidine ( $m^5dC$ ) is known to be an inheritable epigenetic mark in DNA; the precise cellular and molecular functions of  $m^5C$ -modified nucleobases in RNA are not clearly understood. Recent studies have shown that  $m^5C$  sites in lncRNAs can affect binding to chromatin-modifying complexes *in vitro*,<sup>[4]</sup> this implies that RNA C5-methylation can affect the epigenetic status of a cell. Furthermore,  $m^5C$  in vault RNA has been demonstrated to act as a molecular switch to produce specific small vault RNAs, which can enter the micro RNA pathway and influence the translation of distinct sets of mRNAs.<sup>[10]</sup> These findings provide the first evidence that  $m^5C$  in RNA could be critical for the control and regulation of gene transcription and protein translation, and prompted us to investigate whether  $m^5C$  in RNA is subject to dynamic control inside organisms in a fashion similar to  $m^6A$ .

Subsequent to the recent discovery of 5-hydroxymethyl-2'-deoxycytosine ( $hm^5dC$ ) in mammalian DNA,<sup>[11]</sup> the dynamics and oxidative metabolism of DNA C5-methylation have been the subject of intense investigation. The ten-eleven translocation (TET) family of enzymes can mediate the removal of  $m^5dC$  through an oxidative pathway involving the distinct intermediates  $hm^5dC$ , 5-formyl-2'-deoxycytosine and 5-carboxy-2'-deoxycytosine.<sup>[12]</sup> Furthermore, studies on the lifetime of  $hm^5dC$  have demonstrated that it can be a stable DNA modification, thus suggesting that  $hm^5dC$  might have its own unique biological function in addition to being an intermediate in  $m^5dC$  removal.<sup>[13]</sup>

In RNA,  $hm^5C$  was originally identified in rRNA isolated from wheat seedlings.<sup>[14]</sup> During the preparation of this manuscript, Fu et al. reported the presence of  $hm^5C$  for the first time in mammalian total RNA and demonstrated that it can be formed *in vitro* by TET-mediated oxidation of  $m^5C$ .<sup>[15]</sup>

Herein, we present direct chemical evidence for the *in vivo* oxidative processing pathway of  $m^5C$  in RNA. Using an isotope-tracing strategy in a mouse model followed by liquid chromatography tandem-mass spectrometry (LC-MS/MS) analysis, we were able to follow  $m^5C$  metabolism in RNA. Furthermore, we quantified  $hm^5C$  in total RNA samples from all do-

[a] S. M. Huber, Dr. P. van Delft, Prof. S. Balasubramanian

University of Cambridge, Department of Chemistry  
Lensfield Road, Cambridge, CB2 1EW (UK)  
E-mail: sb10031@cam.ac.uk

[b] L. Mendil, Dr. M. Bachman, Prof. S. Balasubramanian

Cancer Research UK Cambridge Institute, University of Cambridge  
Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE (UK)

[c] Dr. K. Smollett, Prof. F. Werner

University College London, Institute for Structural and Molecular Biology  
Gower Street, London, WC1E 6BT (UK)

[d] Prof. E. A. Miska

University of Cambridge  
Wellcome Trust Cancer Research UK Gurdon Institute  
Tennis Court Road, Cambridge, CB2 1QN (UK)

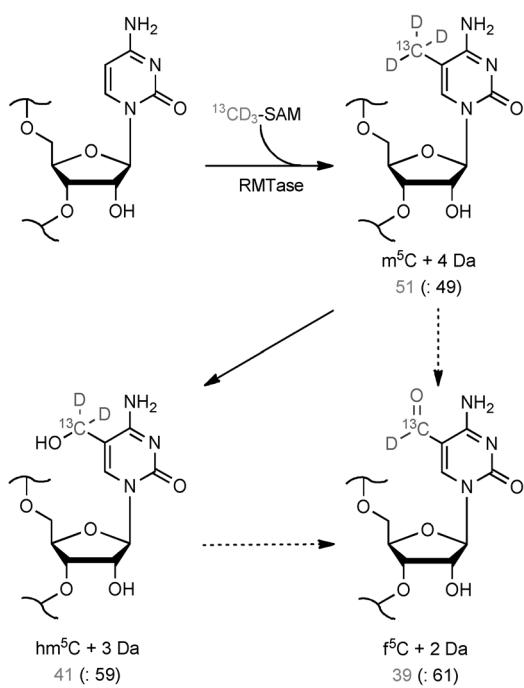
 Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cbic.201500013>.

 © 2015 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA.  
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

means of life and polyA-enriched mammalian RNA to better understand the scope and biological function of this noncanonical base in RNA.

To explore the *in vivo* metabolic pathway of 5-methylcytosine in RNA and determine whether hm<sup>5</sup>C is indeed the result of m<sup>5</sup>C oxidation, we carried out an isotope-tracing experiment by using a mouse model. RNA methyltransferases (RMTases) use S-adenosylmethionine (SAM) as the methyl-group donor in all nucleoside methylation events.<sup>[2]</sup> As SAM is synthesized from L-methionine,<sup>[16]</sup> feeding mice on a diet supplemented with <sup>13</sup>C and D stable-isotope-labelled methionine leads to the incorporation of these isotopes into the C5-methyl group of m<sup>5</sup>C. Consequently, isolation of the murine RNA and the LC-MS/MS analysis thereof, enables the detection of isotope-labelled m<sup>5</sup>C and any derivatives that are formed directly from it.

Mice were fed on a custom diet depleted of L-methionine and supplemented with L-[methyl-<sup>13</sup>CD<sub>3</sub>]methionine for 117 days.<sup>[13]</sup> The mice were then sacrificed, and RNA was extracted from their brain tissue, which was then digested by a mixture of nucleases and phosphatases to form the nucleoside monomers. Incorporation of the <sup>13</sup>CD<sub>3</sub> label was then assessed by LC-MS/MS analysis of the RNA nucleosides, with the mass filters of the triple quadrupole preset for m<sup>5</sup>C and m<sup>5</sup>C + 4 Da. This revealed that the <sup>13</sup>CD<sub>3</sub> label had been incorporated in 51% of total m<sup>5</sup>C (Scheme 1). Next, we detected the masses of hm<sup>5</sup>C and f<sup>5</sup>C and their expected isotopologues hm<sup>5</sup>C + 3 Da and f<sup>5</sup>C + 2 Da, respectively. As depicted in Scheme 1, we observed <sup>13</sup>C and D isotope incorporation into hm<sup>5</sup>C and f<sup>5</sup>C with



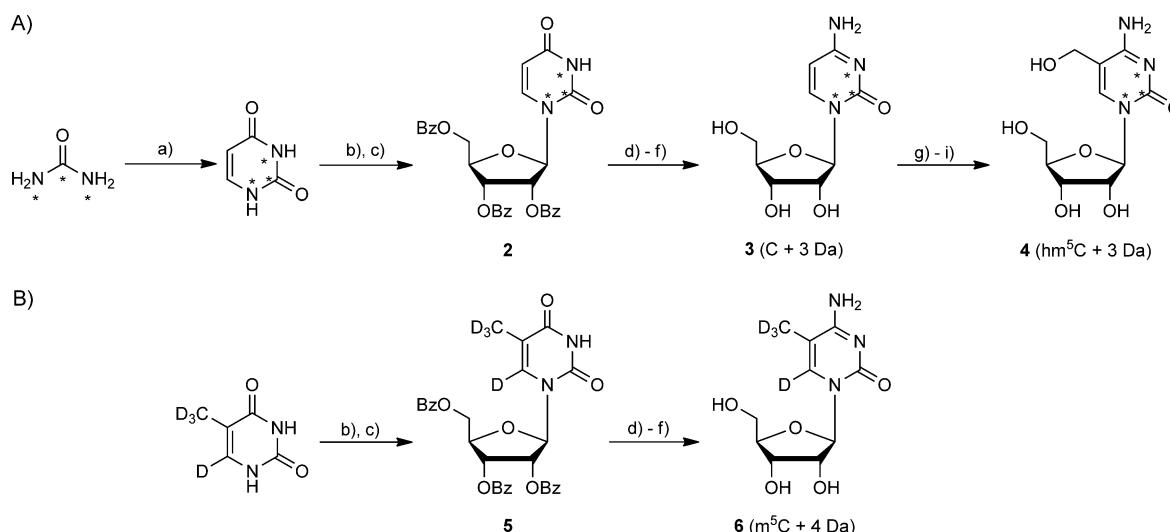
**Scheme 1.** In vivo methylation and subsequent oxidation of cytidine as detected by MS analysis of RNA from the brain of mice fed with L-[methyl-<sup>13</sup>CD<sub>3</sub>]methionine. Numbers correspond to the ratio of peak areas of the depicted labelled nucleosides (in grey) against the respective unlabelled nucleosides (in brackets). See Figure S6.1 and Table S6.3 for extracted ion count chromatograms and absolute values obtained thereof, and Table S6.4 for HR-LC-MS/MS data.

similar efficiencies as for m<sup>5</sup>C (hm<sup>5</sup>C: 41%, f<sup>5</sup>C: 39%). These findings confirm that m<sup>5</sup>C in RNA can be oxidatively metabolized first to hm<sup>5</sup>C and subsequently to f<sup>5</sup>C. In the recently published study on TET-mediated oxidation of m<sup>5</sup>C to hm<sup>5</sup>C in synthetic RNA oligonucleotides, no significant sequential *in vitro* oxidation of hm<sup>5</sup>C to f<sup>5</sup>C was observed.<sup>[15]</sup>

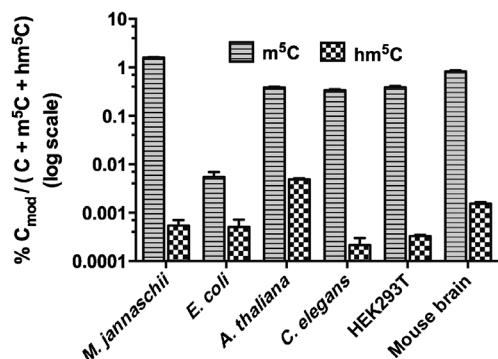
The conservation of m<sup>5</sup>C in RNA across Archaea, bacteria and eukaryotes<sup>[2]</sup> and the apparent metabolic relationship between m<sup>5</sup>C and hm<sup>5</sup>C *in vivo* suggest that RNA C5-hydroxymethylation could be a widespread process. To assess the prevalence of hm<sup>5</sup>C across different species, a quantitative LC-MS/MS method was implemented with stable isotope-labelled (SIL) internal standards, which were spiked into samples and calibration standards.<sup>[17]</sup> Its presence and concentrations were determined by using both the LC and MS/MS selectivity. To apply this method to the detection and quantification of m<sup>5</sup>C and hm<sup>5</sup>C in RNA, SIL standards for cytidine, m<sup>5</sup>C and hm<sup>5</sup>C were synthesized as depicted in Scheme 2.<sup>[18]</sup> Thus, 2-<sup>13</sup>C,<sup>15</sup>N<sub>2</sub>-labelled uracil was synthesized from a commercial source of [2-<sup>13</sup>C,<sup>15</sup>N<sub>2</sub>]urea and used in a Vorbrüggen coupling with an acyl-protected ribose donor in the presence of a Lewis acid to yield 2',3',5'-tri-O-benzoyl-[2-<sup>13</sup>C,1,3-<sup>15</sup>N<sub>2</sub>]uridine (2; Scheme 2A). Treatment of 2 with tri-(1,2,4-triazol-4-yl)-N-phosphoramidate furnished an intermediate N4-triazolyl species. Subsequent aminolysis with concentrated aqueous NH<sub>4</sub>OH and saponification of the benzoyl esters under Zemplén conditions yielded [2-<sup>13</sup>C,1,3-<sup>15</sup>N<sub>2</sub>]cytidine (3). 5-Hydroxymethyl-[2-<sup>13</sup>C,1,3-<sup>15</sup>N<sub>2</sub>]-cytidine (4) was obtained from hydroxymethylation of 3 by adapting published procedures in a microwave-assisted formylation reaction in combination with the use of a cyclic acetal protecting group to aid purification.<sup>[19]</sup> 5-[Methyl-D<sub>3</sub>]-[6D]cytidine (6) was synthesized from [D<sub>4</sub>]thymine by using the same sequence of reactions as was used for the synthesis of 3 (Scheme 2B).

With the stable-isotope-labelled internal standards in hand, we examined the abundance of hm<sup>5</sup>C in total RNA samples from model organisms representing the three domains of life by using a previously described methodology.<sup>[20]</sup> We chose the archaeal species *Methanocaldococcus jannaschii*, the bacterium *Escherichia coli* and two eukaryotic model organisms, namely a plant (*Arabidopsis thaliana*) and a nematode (*Caenorhabditis elegans*). HEK293T cells and mammalian tissues<sup>[21]</sup> were analysed to verify previously reported results.<sup>[15]</sup> To rule out oxidation of m<sup>5</sup>C to hm<sup>5</sup>C as a result of downstream RNA processing, we subjected synthetic RNA 122-mers to identical RNA extraction and enzymatic digestion. No oxidation was detected in the quantitative LC-MS/MS analysis of these control samples.

We detected hm<sup>5</sup>C in all samples studied (Figure 1), thus confirming RNA C5-hydroxymethylation as a common RNA modification. Notably, hm<sup>5</sup>C was found in species that do not exhibit detectable hm<sup>5</sup>dC in their DNA and lack TET homologues in their genomes, such as *C. elegans* and *A. thaliana*.<sup>[22]</sup> This suggests that, in such organisms, the formation of hm<sup>5</sup>C in RNA must occur through a non-TET mechanism. As noted by Fu et al., their observation that TET-null ES cells in which TET1, TET2, and TET3 are genetically deleted still exhibit signifi-



**Scheme 2.** Synthesis of stable-isotope-labelled cytidine, 5-methylcytidine and 5-hydroxymethylcytidine: a) polyphosphoric acid, 17 h, 95 °C, 37%; b) TMSCl, 1,1,1,3,3-hexamethyldisilazane, MeCN, 14 h, 85 °C; c) 1-O-acetyl-2,3,5-tri-O-benzoyl- $\beta$ -D-ribofuranose, TMSOTf, 1,2-dichloroethene, 4 h, 70 °C, 80%; d) tris(3,5-dihydro-4H-1,2,4-triazol-4-yl)phosphine oxide, MeCN, 17 h, 25 °C; e) NH<sub>4</sub>OH (35%), 1,4-dioxane, 2 h, 25 °C, 69%; f) NaOMe, MeOH, 5 h, 25 °C, 97%; g) H<sub>2</sub>SO<sub>4</sub>, acetone, 15 h, 25 °C, 87%; h) paraformaldehyde, KOH (0.5 M), microwave irradiation, 75 min, 60 °C, 16%; i) 90% TFA, 1 h, 0 °C, 98%. Asterisks represent a <sup>15</sup>N or <sup>13</sup>C isotopologue.<sup>[18]</sup>

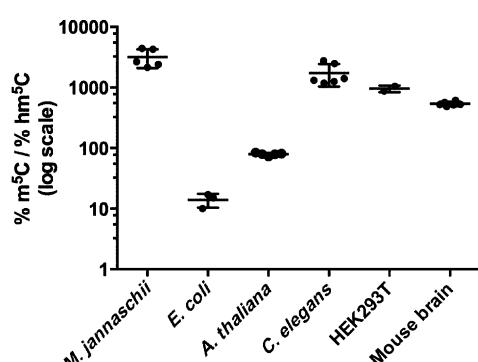


**Figure 1.** Levels of m<sup>5</sup>C and hm<sup>5</sup>C across different model organisms given in amounts relative to the sum of (modified) cytosine residues.<sup>[20]</sup>

cant hm<sup>5</sup>C levels in RNA also supports the idea that hm<sup>5</sup>C in RNA can form from pathways that do not involve TET enzymes.<sup>[15]</sup> Comparison of m<sup>5</sup>C oxidation into hm<sup>5</sup>C across the

different species revealed varying levels (Figure 2). This could also indicate that RNA hydroxymethylation is the result of different enzymatic transformations. It will ultimately be essential to map the position of hm<sup>5</sup>C to specific RNAs in order to understand its function. As a step towards elucidating hm<sup>5</sup>C location, we explored the abundance of hm<sup>5</sup>C in RNA classes other than tRNA and rRNA. Specifically, we isolated mRNA and lncRNAs from HEK293T cells that contain a polyA tail by RNA pulldown using polyT magnetic beads. Subsequent analysis of this fraction by LC-MS/MS allowed us to measure hm<sup>5</sup>C (Figure S6.2 in the Supporting Information). In these samples, the extent of C5-hydroxymethylation in polyA RNA was 40 times higher than in total RNA (Table S6.5). Currently, the function of m<sup>5</sup>C in mRNA and lncRNA is not clearly understood. However, the presence and levels of hm<sup>5</sup>C as an oxidative m<sup>5</sup>C metabolite in polyA RNA suggests that hm<sup>5</sup>C and m<sup>5</sup>C might be part of a dynamic regulatory mechanism. Further studies are needed to assess the functional roles of m<sup>5</sup>C and hm<sup>5</sup>C in RNA.

In summary, we have shown for the first time that 5-hydroxymethylcytosine in RNA is generated by m<sup>5</sup>C oxidation in vivo and is pervasive in all three domains of life across a variety of different species. Our findings suggest the existence of dynamic m<sup>5</sup>C metabolism in RNA that might have important roles in RNA function.



**Figure 2.** The abundance of hm<sup>5</sup>C as a fraction of m<sup>5</sup>C in various organisms.

## Experimental Section

All in vivo experiments were performed under the terms of a UK Home Office licence. C57BL/6J (JAX mice strain) mice were bred and housed according to UK Home Office guidelines. Detailed information on the mouse labelling study, cell culture and subsequent RNA extraction (TRI reagent, Sigma) and purification (RNA clean & concentrator kit, Zymo Research) is given in the Supporting Information. Synthetic procedures and full compound charac-

terization for all compounds and  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of **3**, **4** and **6** can also be found in the Supporting Information.

RNA was digested by using a mixture of Benzonase (Sigma), snake venom phosphodiesterase I (*Crotalus adamanteus* venom, Sigma) and antarctic phosphatase (New England Biolabs) in a Tris-HCl buffer (pH 8) for 12 h. A solution of SIL C, m $^5\text{C}$  and hm $^5\text{C}$  used for LC-MS/MS quantitation was added to samples before the enzymatic digestion.

LC-MS/MS analysis was performed by using an Agilent 1290 Infinity UHPLC system coupled to an AB Sciex 6500 triple-quadrupole mass spectrometer. Chromatographic conditions typically consisted of using a Waters Acquity UPLC HSS T3 column and an  $\text{H}_2\text{O}/\text{MeCN}$  (0.1 % formic acid) linear gradient.

## Acknowledgements

This work was supported by the Cambridge PhD Training Programme in Chemical Biology and Molecular Medicine and the Wellcome Trust (grant number 099232/Z/12/Z). Dr. Donna Bond is thanked for the provision of *A. thaliana* total RNA and Drs. Santiago Uribe-Lewis and Adele Murrell are acknowledged for the labelled mouse tissue supply.

**Keywords:** 5-hydroxymethylcytosine • 5-methylcytosine • isotope tracing • LC-MS/MS • RNA modifications

- [1] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262–D267.
- [2] Y. Motorin, F. Lyko, M. Helm, *Nucleic Acids Res.* **2010**, *38*, 1415–1430.
- [3] J. E. Squires, H. R. Patel, M. Nousch, T. Sibbritt, D. T. Humphreys, B. J. Parker, C. M. Suter, T. Preiss, *Nucleic Acids Res.* **2012**, *40*, 5023–5033.
- [4] T. Amort, M. F. Soulière, A. Wille, X.-Y. Jia, H. Fiegl, H. Wörle, R. Micura, A. Lusser, *RNA Biol.* **2013**, *10*, 1002–1008.
- [5] C. Yi, T. Pan, *Acc. Chem. Res.* **2011**, *44*, 1380–1388.

- [6] G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y.-G. Yang, C. He, *Nat. Chem. Biol.* **2011**, *7*, 885–887.
- [7] G. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C.-M. Huang, C. J. Li, C. B. Vägbo, Y. Shi, W.-L. Wang, S.-H. Song, Z. Lu, R. P. G. Bosmans, Q. Dai, Y.-J. Hao, X. Yang, W.-M. Zhao, W.-M. Tong, X.-J. Wang, F. Bogdan, K. Furu, Y. Fu, G. Jia, X. Zhao, J. Liu, H. E. Krokan, A. Klungland, Y.-G. Yang, C. He, *Mol. Cell* **2013**, *49*, 18–29.
- [8] X. Wang, Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan, C. He, *Nature* **2014**, *505*, 117–120.
- [9] Y. Fu, D. Dominissini, G. Rechavi, C. He, *Nat. Rev. Genet.* **2014**, *15*, 293–306.
- [10] S. Hussain, A. A. Sajini, S. Blanco, S. Dietmann, P. Lombard, Y. Sugimoto, M. Paramor, J. G. Gleeson, D. T. Odom, J. Ule, M. Frye, *Cell. Rep.* **2013**, *4*, 255–261.
- [11] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) 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.
- [12] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
- [13] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 1049–1055.
- [14] I. Rácz, I. Király, D. Lászity, *Planta* **1978**, *142*, 263–267.
- [15] L. Fu, C. R. Guerrero, N. Zhong, N. J. Amato, Y. Liu, S. Liu, Q. Cai, D. Ji, S.-G. Jin, L. J. Niedernhofer, G. P. Pfeifer, G.-L. Xu, Y. Wang, *J. Am. Chem. Soc.* **2014**, *136*, 11582–11585.
- [16] S. C. Lu, *Int. J. Biochem. Cell Biol.* **2000**, *32*, 391–395.
- [17] a) E. Ciccimaro, I. A. Blair, *Bioanalysis* **2010**, *2*, 311–341; b) J. J. Dalluge, T. Hashizume, J. A. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 3242–3245; c) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2009**, *48*, 7932–7934; *Angew. Chem.* **2009**, *121*, 8074–8077.
- [18] See the Supporting Information for detailed synthetic procedures and spectroscopic data.
- [19] a) M. Khurshid, A. Khan, M. Qudrat-E-Khuda, *J. Chem. Soc. Pak.* **1982**, *4*, 167–168; b) A. A.-H. Abdel-Rahman, E. S. H. El Ashry, *Synlett* **2002**, 2043–2044.
- [20] See Tables S6.1 and Table S6.2 for exact values.
- [21] See Figure S6.3 for hm $^5\text{C}$  abundance in mammalian tissues.
- [22] S. Liu, T. L. Dunwell, G. P. Pfeifer, J. M. Dunwell, I. Ullah, Y. Wang, *PLoS One* **2013**, *8*, e84620.

Received: January 9, 2015

Published online on February 12, 2015