Epigenetics

DOI: 10.1002/anie.201101547

5-Hydroxymethylcytosine, the Sixth Base of the Genome

*Martin Münzel, Daniel Globisch, and Thomas Carell**

5-hydroxymethylcytosine · DNA modification · epigenetics · nucleobases · TET enzymes

***5-H**ydroxymethylcytosine (hmC) was recently discovered as a new constituent of mammalian DNA. Besides 5-methylcytosine (mC), it is the only other modified base in higher organisms. The discovery is of enormous importance because it shows that the methylation of cytosines to imprint epigenetic information is not a final chemical step that leads to gene silencing but that further chemistry occurs at the methyl group that might have regulatory function. Recent progress in hmC detection—most notably LC-MS and glucosyltransferase assays—helped to decipher the precise distribution of hmC in the body. This led to the surprising finding that, in contrast to constant mC levels, the hmC levels are strongly tissue-specific. The highest values of hmC are found in the central nervous system. It was furthermore discovered that hmC is involved in regulating the pluripotency of stem cells and that it is connected to the processes of cellular development and carcinogenesis. Evidence is currently accumulating that hmC may not exclusively be an intermediate of an active demethylation process, but that it functions instead as an important epigenetic marker.*

1. History of 5-Hydroxymethylcytosine

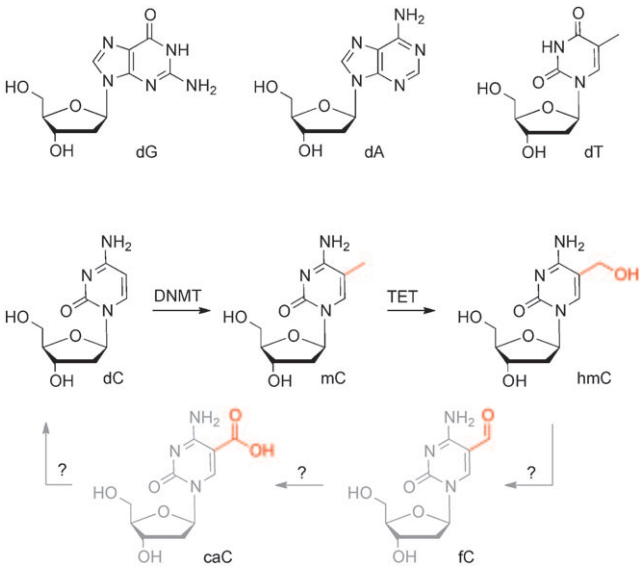
The genetic material is constructed from the four DNA nucleosides 2′-deoxyadenosine (dA or dAdo), 2′-deoxyguanosine (dG or dGua), thymidine (Thd or dT), and 2′-deoxycytidine (dC or dCyt; Scheme 1). These four building blocks are assembled inside a DNA double helix whose structure was deciphered by Watson and Crick in 1953.[1] In multicellular organisms, all cells possess the same genetic content; however, cells might perform vastly different functions. A neuronal cell, for example, is designed to conduct electrical signals along axons while a muscle cell is involved in performing contractive motions. The same genetic material manifests in the different functions of different cells, which is only possible because cells differ in the kind and number of active genes. During cellular development, cells switch unneeded genes off on their way to specialization. Chemically, long-term gene silencing is achieved by the methylation of the base cytosine (dC) at position C5 in special CpG islands.[2]

The methylation pattern is a crucial part of the epigenetic information and a critical marker that distinguishes cells. The underlying nucleoside 5-methyldeoxycytidine (mC or $^{5\text{-Me}}$dC or M or 5 mC) is often considered to be the fifth base of the genome. After fertilization and at certain points of the embryonic development, a large number of the methylation marks are erased.[3] As a consequence, this allows embryonic stem cells to differentiate into any possible specialized cell. In certain cases this genome-wide demethylation occurs without cell division and thus without the synthesis of new DNA. The methyl group therefore has to be actively removed by enzymes. The mechanism underlying this active demethylation is of utmost interest, because it is speculated that some cells might have the possibility to actively demethylate their genetic material in order to re-differentiate.[4] In other words: One might be able to move differentiated cells into a less differentiated state by a process called active demethylation, which would have great potential for research and therapeutic applications.

Recently, the search for this active demethylation process moved in a new direction. In 2009 two independent reports published back-to-back in *Science* showed that mammalian genomic DNA contains not only 5-methylcytosine but also 5-hydroxymethylcytosine (hmC or $^{5\text{-HOMe}}$dC or 5 hmC), which is now considered to be the sixth base of the genome of higher

[*] Dipl.-Chem. M. Münzel,[+] Dipl.-Chem. D. Globisch,[+] Prof. Dr. T. Carell
Center for Integrated Protein Science
Department of Chemistry
Ludwig-Maximilians-University Munich
Butenandtstr. 5–13, 81377 Munich (Germany)
Fax: (+49) 89-2180-77756
E-mail: thomas.carell@cup.uni-muenchen.de
Homepage: http://www.carellgroup.de

[+] These authors contributed equally.



Scheme 1. The canonical DNA nucleosides dG, dA, dT, and dC. Cytosine can be modified to mC and hmC in mammalian tissues. hmC could be further oxidized to the putative demethylation intermediates fC and caC, but have never been found in vivo.

organisms.[5] Interestingly, hmC was already found in mammalian DNA 40 years ago.[6] These findings, however, remained in doubt because later studies were unable to reproduce the data and even the new publications report much lower levels of hmC.[7]

1.1. 5-Hydroxymethylcytosine as a Nucleobase in Bacteriophage DNA

The discovery of hmC in mammals was a milestone for epigenetic research. The base itself, however, has long been known to be a component of the DNA of special bacteriophages. This was discovered in 1952, even before the 3D structure of DNA was deciphered.[8] Certain phages incorporate hmC, with help from the corresponding triphosphate, into their genetic material. A subsequently acting glucosyltransferase finally loads a glucose moiety onto the base, thereby creating highly α- or β-glucosylated phage DNA.[9] This glycosylation is part of a protecting strategy used by DNA bacteriophages to escape the action of bacterial restriction enzymes. These otherwise degrade invading double-stranded DNA in a bacterial defense process. In addition, the phage reprograms the cell to produce nucleases which degrade non-hmC-containing DNA in an attempt to clear the infected cell from its own genetic material.[10]

1.2. 5-Hydroxymethylcytosine as an Oxidatively Damaged Nucleobase

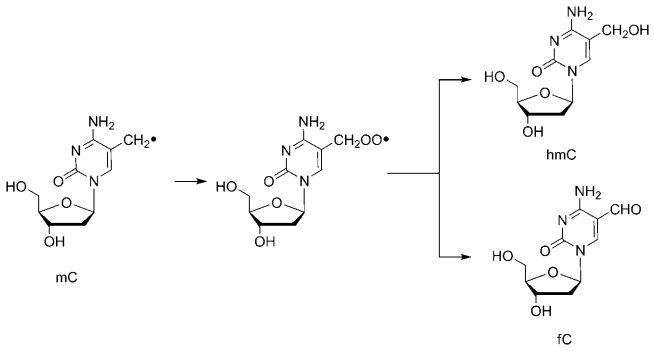
5-Hydroxymethylcytosine was also suspected to be formed at mC sites in response to oxidative stress in normal DNA. It was shown in in vitro studies that one-electron oxidation of mC causes formation of hmC in an aerated aqueous solution under UV-A irradiation.[11] This is possible in the presence of the type I photosensitizer menadione, which yields the unstable peroxyl radical or hydroxyperoxide intermediates. These decompose to hmC.[11,12] Another study showed that treatment of mC under Fenton conditions (Fe^{2+}/Cu^{2+} and H_2O_2) or with benzoylperoxide in $CBrCl_3$ also yielded hmC.[13] Since then it is speculated that hmC could be one of the major oxidative degradation products of mC (Scheme 2).

On the basis of data obtained with dT, it was proposed that 20 molecules of hmC per cell per day are formed as an oxidatively generated lesion.[14] Interestingly, hmC can also be photochemically converted into dC after elimination of formaldehyde.[12b] However, its oxidative formation under in vivo conditions could not yet be confirmed. An important observation was that under oxidative conditions not only hmC but also 5-formyldeoxycytydine (fC or $^{5\text{-CHO}}$dC or 5fC) and 5-carboxydeoxycytidine (caC or $^{5\text{-COOH}}$dC or 5caC) are formed.[11,12] While fC can be deformylated via the hydrate, caC can be decarboxylated, which leads to the regeneration of cytosine. It is this reactivity which spurs the current idea that

Daniel Globisch studied chemistry at the Technical University Kaiserslautern. He completed his diploma at the Nucleic Acids Research Center of the Southern University of Denmark in Odense in the group of Prof. Erik B. Pedersen, where he worked on HIV-1 inhibitors and triplex-forming oligonucleotides. In 2007 he started his PhD in the group of Prof. Thomas Carell at the Ludwig-Maximilians-University in Munich. His research is focused on the synthesis and quantification of natural modified tRNA and DNA nucleosides.

Martin Münzel studied chemistry in Marburg, Germany and Stockholm, Sweden. His diploma thesis in the group of Prof. Gerhard Klebe in Marburg focused on the synthesis of thrombin inhibitors. He then joined the group of Prof. Dirk Trauner at the University of California in Berkeley to work on the total synthesis of complex natural products. In 2008 he started his PhD in the group of Prof. Thomas Carell at the LMU Munich. His research interests include the synthesis and detection of modified DNA bases and novel photolesions.

Thomas Carell studied chemistry at the Universities of Münster and Heidelberg and obtained his PhD with Prof. H. A. Staab at the Max-Planck Institute of Medical Research. After postdoctoral research with Prof. J. Rebek at MIT he joined the group of Prof. F. Diederich at ETH Zürich to start independent research. In 2000 he became full professor at the Philipps-University in Marburg and in 2004 he moved to his current position at the LMU Munich. He received the Gottfried Wilhelm Leibniz Award in 2003. He is a member of the German Academy of Science, Leopoldina, and of the Berlin-Brandenburg Academy of Science and Humanities.



Scheme 2. One-electron oxidation of mC yields the two main products hmC and fC.[12a]

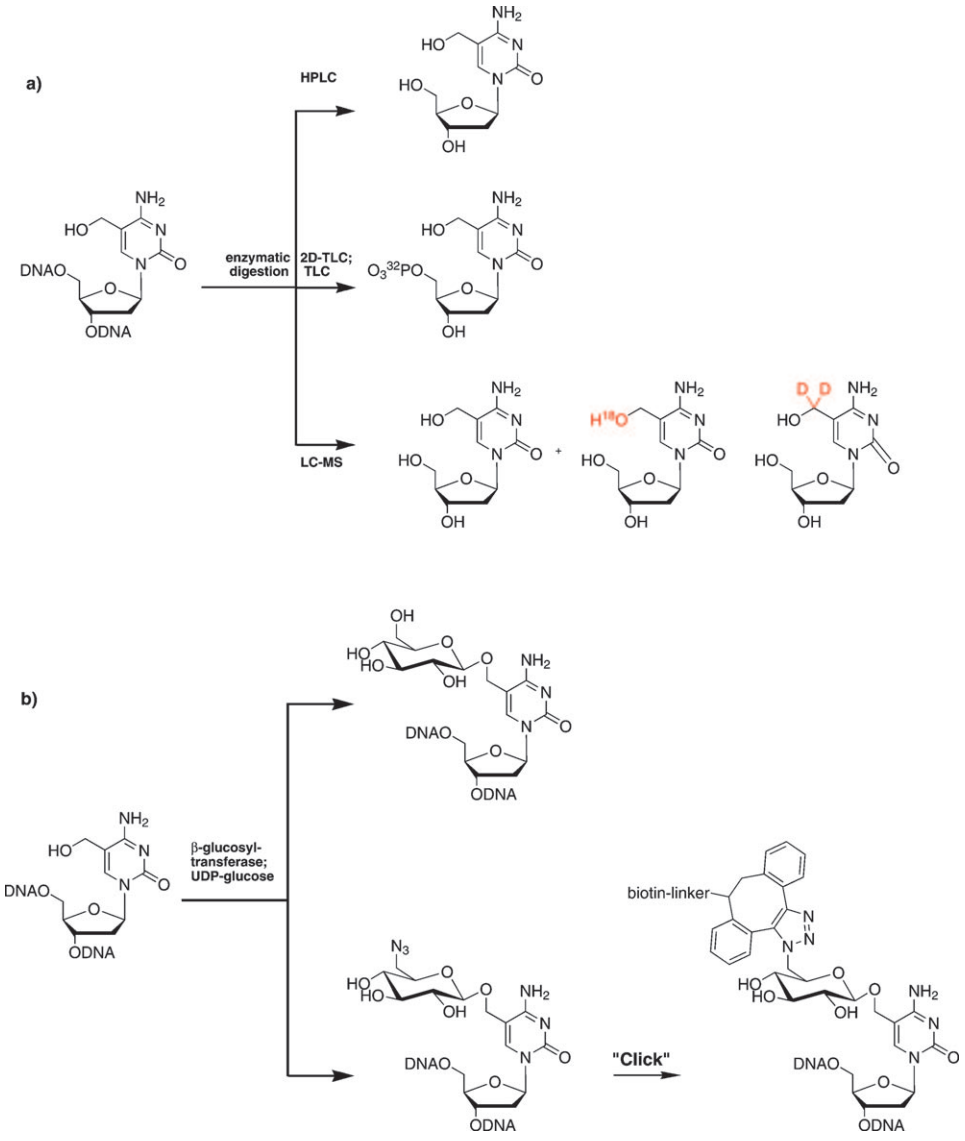
hmC might be an intermediate in the long searched for pathway that enables active demethylation.

2. Detection Methods

2.1. Global Quantification of 5-Hydroxymethylcytosine

In the two reports that appeared in *Science* 2009, Heintz and Rao and their co-workers isolated DNA from purkinje neurons and embryonic stem cells, respectively.[5] The DNA was digested enzymatically to the nucleoside level. Finally, the nucleoside mixture was treated with a kinase and ^{32}P-ATP to transfer a radioactive phosphate group onto the primary 5′-hydroxy group. Subsequent 2D chromatography, HPLC, and MS analysis of the nucleotides enabled both groups to detect hmC as a new component in genomic DNA (Scheme 3a). Kriaucionis and Heintz serendipitously discovered that 0.6% of all the cytosine residues in purkinje neurons and about 0.2% of all the cytosine residues in granule cells are hydroxymethylated. They further found that hmC is present in many tissue types, although the detection limit (ca. 0.08%) of their technique did not allow this to be studied in detail.[5a] The amount of hmC in embryonic stem cells studied by Rao and co-workers was even lower, with only 0.1% of the cytosine moieties carrying a 5-hydroxymethyl group.[5b]

With the help of isotope-labeled hmC as an internal standard for an accurate HPLC-MS method, Carell and co-workers investigated the distribution of hmC in mouse tissues (Scheme 3a).[15] It was found that the amount of 5-hydroxymethylcytosine varies substantially between different tissue types. Whereas the amount of 5-methylcytosine is constant at around 4.5% (values are always given in % of all cytosine derivatives) in all tissues, it was discovered that organs such as liver and testes have very low, heart and kidney medium, and tissues of the central nervous system high amounts of hmC (Figure 1a).[15a] A more detailed survey of various brain tissues again showed that the 5-hydroxymethylcytosine levels vary from 0.3% in retina and cerebellum to 0.7% in cerebral cortex and hippocampus.[15b] The latter are the brain areas involved in higher cognitive functions. These results are



Scheme 3. Methods for the detection and quantification of hmC. a) Quantification of DNA nucleosides or nucleotides after enzymatic digestion. b) Enzymatic glucosylation of hmC and subsequent detection by radioactivity or click reaction with biotin.

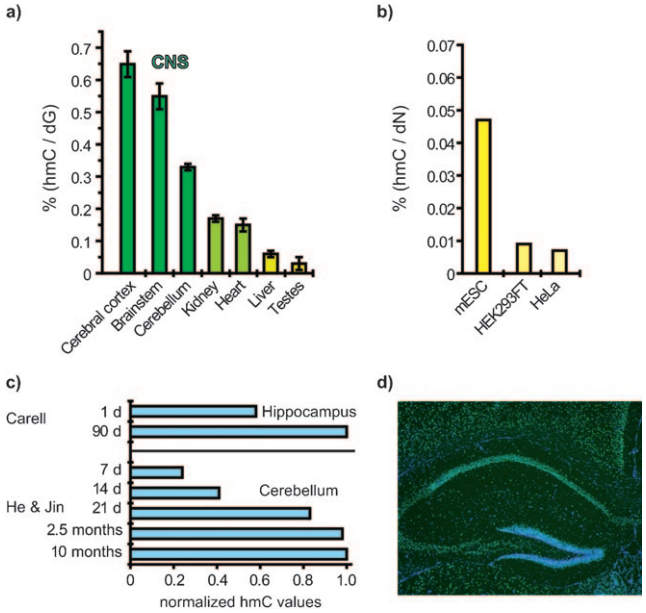


Figure 1. Distribution of hmC in the mammalian body. a,b) Quantification of hmC in different mammalian tissues and different cell lines. c) hmC values rise during the development of the mouse brain (normalized values). d) Immunolocalization of hmC in mouse hippocampus. Anti-hmC (green) and Hoechst 33342 nuclear dye (blue), mESC: mouse embryonic stem cells.

exciting, especially in light of recent findings that locus-specific active demethylation occurs in fully differentiated neurons as a response to learning and memory formation.[4,16] Another interesting result is the discovery that the hmC values increase during the development of the brain. The hmC value in the hippocampus of 90 day old adult mice is almost double that in 1 day old mice (Figure 1c).[15b] A similar correlation was later shown for the cerebellum (Figure 1c).[17] Together, this hints at a role of hmC in brain maturation and neuronal development.

In an alternative approach for the detection of hmC, Leonhardt and co-workers used the phage β-glucosyltransferase to load radioactive labeled UDP-[^{3}H]glucose onto the hydroxy group of the hmC base within DNA (Scheme 3b).[18] They determined the amounts of hmC with a liquid scintillation analyzer. By using this method the researchers were able to measure the tissue distribution of hmC and to establish that the presence of mC is required for the formation of hmC. This was achieved by measuring hmC values in DNMT triple knockout cells which had lost the ability to methylate cytosine residues.

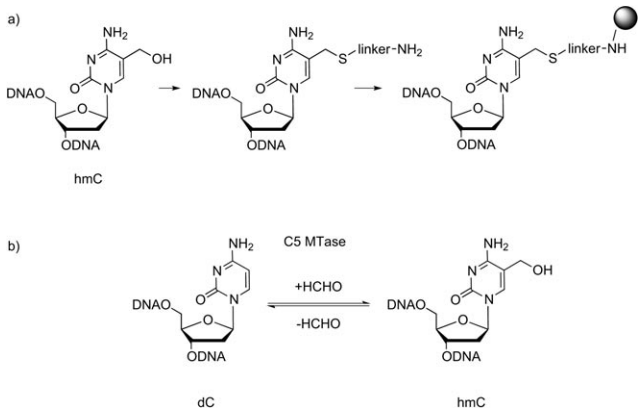
Jin, He, and co-workers beautifully extended the method developed by Leonhardt and co-workers by employing UDP-6-azidoglucose and the same phage β-glucosyltransferase to load hmC-containing DNA with azide moieties.[17] By using click chemistry they were able to attach biotin units to each hmC residue, which allowed them to isolate hmC-containing DNA strands for subsequent sequencing analysis. They were furthermore able to quantify the amount of hmC in DNA by using biotin-specific antibodies. With this method Jin, He, and co-workers not only confirmed the age effect but showed, furthermore, that hydroxymethylation is accumulated in intragenic regions. In contrast, mC is predominantely found in CpG islands, which occur in promoter regions. Interestingly, hmC was found to be enriched in genes that are linked to angiogenesis and hypoxia. HIF-prolyl hydroxylases, a well-known class of oxygen-sensing enzymes, are related to the TET enzymes, which form hmC (see below).[19] This led Jin, He, and co-workers to speculate that hmC might be involved in an oxygen-sensing and regulation process.

The authors further confirmed that hmC is present only at very low levels in cultured cells (Figure 1b).[17] Recent staining experiments with a polyclonal antibody against hmC showed that hmC is present in nuclei of all cell types and confirmed the uneven distribution found in the quantification experiments.[15a] High amounts were detected in the differentiated neurons of the dentate gyrus (Figure 1d). Areas that are rich in neuronal stem cells, in contrast, did not stain well, thus indicating that the hmC is present in substantially lower amounts in the corresponding nuclei.

Further detection methods that enable isolation of hmC-containing DNA strand fragments include the use of J-binding protein 1 (JBP1) bound to magnetic beads.[20] These proteins bind to glucosylated 5-hydroxymethyluracil, which is called the J base and is present in trypanosomes and related kinetoplastids.[21] Robertson et al. could show that JBP1 binds tightly to glucosylated hmC, thereby allowing isolation of hmC-containing DNA by simple incubation of the DNA with magnetic beads.[20] By using this method it was discovered that hmC levels are higher in the promoter regions of the *NANOG* and *OCT4* genes which are associated with pluripotency.

Further LC-MS/MS-based quantification of hmC was carried out by several research groups.[22] The hmC as well as the mC values were lower in *Dnmt1* and *Dnmt3a* double conditional knockout (DKO) mice than in the control experiments, while no difference in the hmC levels was observed for each single conditional knockout (SKO) experiment.[22a] This additionally supports the formation of hmC from mC. With the development of specific antibodies against hmC, it additionally became possible to quantify the amount of the base by classical dot-blot techniques.[23]

An interesting discovery by Klimasauskas and co-workers was that methyltransferases are able to bind to hmC-containing DNA and that they can replace the OH group by a sulfur- or selenium-containing moiety, which allows hmC-containing DNA strands to be chemically modified (Scheme 4a).[24] In an earlier study, the same authors showed that bacterial methyltransferases are, furthermore, capable of adding formaldehyde to cytosine residues to yield hmC.[25] Interestingly, the reverse reaction to remove the hydroxymethyl group was also observed and could be a conceivable mechanism of active demethylation (Scheme 4b). The formation of formaldehyde in vivo would not be problematic for cells, because formaldehyde is a known by-product of the metabolism and is, for example, also formed by histone demethylases or certain DNA repair enzymes.[26]



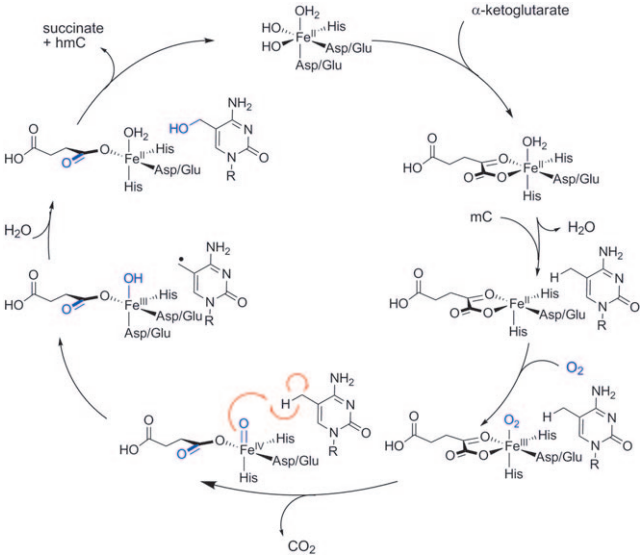
Scheme 4. a) Methyltransferase-dependent derivatization of hmC with either cysteamine or selenocysteamine and further functionalization with any fluorescent group. b) Reversible conversion of dC into hmC with formaldehyde and methyltransferases.

2.2. Site-Specific Detection of 5-Hydroxymethylcytosine

The method of choice for sequencing epigenetic information, bisulfite, unfortunately fails to detect hmC, which makes it currently tremendously difficult to sequence genes directly for the position of hmC bases.[27] Bisulfite deaminates cytosine moieties to uracil, while leaving mC intact.[28] In addition, bisulfite reacts with the hydroxymethyl group by converting it into the 5-methylenesulfonate.[29] As this reaction product does not lead to deamination and possibly blocks polymerases, discrimination of hmC and mC is currently impossible. The 5-methylenesulfonate generated from hmC, however, was used to raise an antibody that allows detection and quantification of hmC via the sulfonate.[30] Alternative bisulfite-independent sequencing principles also cannot distinguish between mC and hmC or have not yet been tested.[31] In addition, standard restriction enzyme based sequencing does not distinguish mC and hmC. However, recent progress has been made in this direction with the use of glucosyltransferases in combination with the restriction enzymes MspI, HpaII, and GlaI, which are, however, very restricted in their sequence specificity.[32] Furthermore, the endonuclease PvuRts1I has been identified to specifically act on hmC.[33] In the future, nanopore sequencing might be a solution to the problem. First studies showed that this sequencing method is indeed able to distinguish between mC and hmC in both single- and double-stranded DNA.[34] Finally, modern sequencing-by-synthesis methods could yield sequence-specific information about hmC on the basis of kinetics of dNTP binding and incorporation.[35]

3. Enzymology of 5-Hydroxymethylcytosine Formation

Parallel to the development of new quantification and site-specific detection methods for hmC, the enzymology behind the formation of this new base was investigated. In their initial report Rao and co-workers discovered by using bioinformatic methods that hmC is formed post-replicatively by oxidation of mC with the help of the TET1 (ten-eleven translocation) enzyme.[5b,36] Currently, three TET enzymes (TET1–3) are known. All of them are Fe^{II}- and 2-ketoglutarate-dependent oxidases, which efficiently convert mC into hmC both in vitro and in vivo.[5b,37] Besides the catalytic subunit with the iron-binding site, they contain a binuclear zinc-chelating CXXC domain that is found in many chromatin-interacting proteins.[5b,38] The consensus reaction mechanism of α-ketoglutarate-dependent dioxygenases is shown in Scheme 5.[39]



Scheme 5. Reaction mechanism of α-ketoglutarate-dependent dioxygenases. The reaction proceeds by a one-electron reduction of a key iron(IV) intermediate. During the reaction the cofactor α-ketoglutarate is decarboxylated to succinate.

3.1. The Role of 5-Hydroxymethylcytosine and the TET Enzymes during Cellular Development

Levels of 5-hydroxymethylcytosine are dynamically regulated in stem cells and during differentiation.[40] Only TET3 is highly expressed in oozytes and zygotes, and a recent study showed that after fertilization mC is converted into hmC only in the male but not the female pronucleus.[23,41] The data suggest that the event that was always considered to be a genome-wide demethylation of the parental genome is actually a genome-wide hydroxymethylation which is mediated by TET3. After the reprogramming of the paternal genome, TET3 levels decrease rapidly in the two-cell stage. In cultured embryonic stem cells, in contrast, it was discovered that TET1 and TET2 are highly expressed while the TET3 levels were found to be low.[42] The TET3 levels increase again when the stem cell differentiates. Differentiation is furthermore accompanied by a reduction in the TET1 and TET2 levels. If adult cells are de-differentiated to induced pluripotent stem cells, TET3 levels were found to decrease while TET1 and TET2 levels again increased, thus showing the reversibility of the process. In conclusion, these data show TET1 and TET2 are both linked to the pluripotency network and the differentiation potential of stem cells.

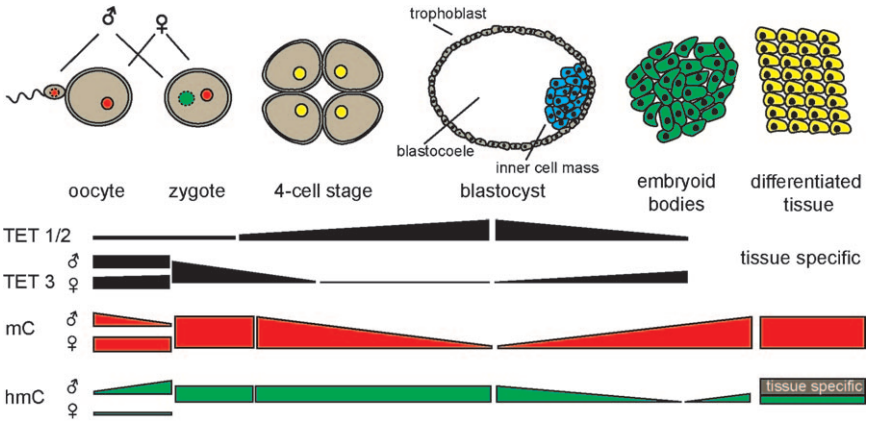


Figure 2. Levels of mC, hmC, and the TET enzymes during cellular development.

siRNA-mediated knockdown of TET1 in stem cells showed that the absence of this enzyme causes morphological changes to the embryonic stem cells.[37] These results could, however, not be reproduced by another research group.[42a] In addition, the embryonic stem cells moved towards differentiation, again showing that high TET1 and thus hmC levels are important for the maintenance of embryonic stem cells.[37] In a further study, Koh et al. showed that TET1 depletion leads to a mesoderm/endoderm bias, whereas knockout of TET2 alters neuroectoderm differentiation.[42a] At the genetic level, the TET1 defect caused increased methylation of the *NANOG* promoter, which leads to the loss of pluripotency.[37] The relationship between methylation of the *NANOG* promoter, TET1, and TET2 is, however, very controversial.[37,42a,43] It was also discussed that the *NANOG* and *OCT4* gene products might be involved in regulation of the TET1 and TET2 enzymes.[42a] Chromatin immunoprecipitation revealed that OCT4 binds to noncoding sequences of the TET1 and TET2 genes, which supports the theory that the DNA hydroxylases are regulated by the pluripotency factor *OCT4*.[42a] An overview of the investigated TET, hmC, and mC values in each developmental stage is depicted in Figure 2.

In combination, the data suggest that TET1 and TET2 are important to maintain pluripotency, while TET3 is associated with differentiation processes. hmC is consequently involved in the formation of pluripotent cells, maintenance of pluripotency, and differentiation into the germ cell lines.

3.2. The role of 5-Hydroxymethylcytosine and the TET Enzymes in Cancer

It was found that many myeloid cancers (leukemia) feature a mutation in TET2, which in turn reduces the ability of cells to hydroxymethylate cytosine bases. It was found that TET2 has an important role in normal myelopoiesis and that impaired formation of hmC is associated with myeloid tumorgenesis.[30] Surprisingly, it was discovered that patients with an impaired TET2 function also have lower mC levels. Since hmC is known to be formed from the precursor mC, this decrease is counterintuitive, again showing the complexity of epigenetic programming mechanisms. It was hypothesized that loss of TET2 function may reactivate a stem cell like state that is characterized by hypomethylation and general genomic instability.

Another enzyme that is impaired in many cancer types is isocitratedehydrogenase (IDH).[44] A mutation of IDH leads to the accumulation of 2-hydroxyglutarate in the cells.[45] This small molecule is a known competitive inhibitor of the 2-ketoglutarate-dependent enzymes, which is now considered to contribute to cancer development through inhibition of the TET enzymes.[46] Further support for impaired hydroxylation in cancers comes from the observation that all the cancer cell lines studied so far only show extremely small hmC values.[17] The measurement of hmC levels by one of the above described methods may consequently be a valuable tool for the early diagnosis of cancer.

To gain evidence that reduced hmC values are the cause and not just the consequence of cancer, Koh et al. injected normal and TET-depleted stem cells into mice. While wild-type cells formed highly differentiated benign teratomas, TET-deficient stem cells developed very aggressive tumors that were in a highly proliferative state.[42a] This elegant experiment showed that hmC is of utmost importance for the correct function of cells and that disturbing the hydroxymethylation of DNA can lead to serious defects.

4. Regulatory Function

Despite the tremendous research effort in the last two years, the exact biochemical functions of hmC remain enigmatic. Although it is established that the sixth base is involved in gene regulation during development and carcinogenesis, and evidence is increasing that hydroxymethylation is associated with actively transcribed genes, the exact biochemical mechanisms behind these findings are still unclear.[37,40,47] How the information on the DNA is translated into cellular signals needs to be clarified in particular. For mC, special methyl-CpG binding domains (MBD) enable the cross-talk between epigenome and proteome, however, specific hydroxymethyl-binding domains remain to be found (Figure 3). Importantly, mC-binding proteins such as MeCP2 or the DNMTs are known not to bind hydroxymethylated DNA, thus suggesting complementary roles of the two epigenetic DNA modifications.[48]

Another possible function is that hmC is an intermediate in an oxidative demethylation process. However, Carell and co-workers were not able to find either 5-formyl- or 5-carboxycytosine in somatic cells, even when large amounts of DNA were analyzed.[15a] This shows that if hmC is involved in an oxidative active demethylation mechanism, the intermediates do not accumulate to any detectable level. Furthermore, no evidence for an alternative active demethylation mecha-

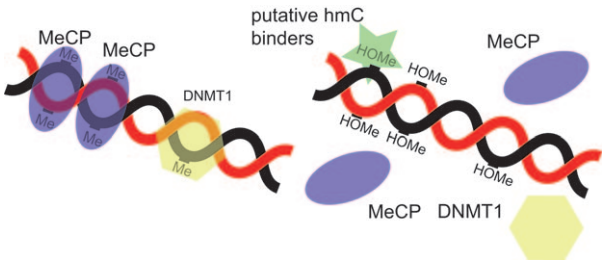


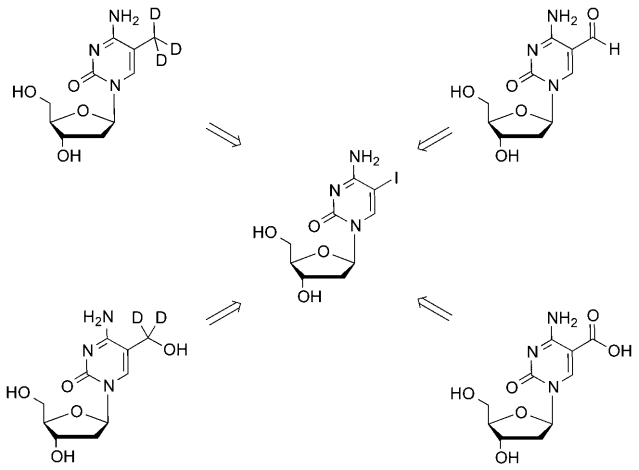
Figure 3. Effect of hmC on 5-methylcytosine-binding proteins.

nism which would proceed via deamination to hmU and subsequent base excision repair could be found. These results were supported by a study by Iqbal et al., whose data suggest the absence of these demethylation mechanisms in the zygote.[23] Recently, Guo et al. were able to find an active demethylation process when they overexpressed TET and APOBEC proteins in neurons.[16c]

5. Chemical Synthesis of 5-Hydroxymethylcytosine-Containing DNA and Biophysical Studies

Two main areas arose in the chemical research of hmC. Firstly, it was important to develop phosphoramidite building blocks that allow efficient incorporation of hmC into DNA. Secondly, isotope-labeled derivatives of the cytosine modifications were synthesized, which can be used as internal standards in MS-based quantification or for metabolic labeling.[13b,15,22b,49]

Synthetically, two strategies evolved: the synthesis of hmC via hmU[14,15b,50] or by metal-catalyzed functionalization of 5-iododeoxycytidine (Scheme 6).[15a,51] The first strategy was used by Sowers and co-workers to develop the first phosphoramidite building block, in which the 5-hydroxymethyl group is protected with a cyanoethyl moiety.[14] Recently, the synthesis was elegantly streamlined by Hansen et al.[50b] The building block is routinely used for the generation of hmC-containing DNA but requires rather harsh deprotection conditions (60 h, 60 °C, conc. aq NH_3), thus limiting the ability to synthesize DNA strands which contain further functionalities such as biotin, which are needed for many biochemical experiments. A further phosphoramidite was developed by de Kort et al., used an acetate to protect the primary hydroxy group. It could be deprotected under mild conditions in a two-step sequence (NaOH and NH_4OH).[50a] Carell and co-workers designed a phosphoramidite building block which utilizes a cyclic carbamate group to protect the N(4)-amino and 5-hydroxymethyl group simultaneously.[51] It can be deprotected under mild conditions of dilute NaOH at room temperature and now allows the generation of highly functionalized hmC-containing DNA strands. In addition, two related phosphoramidites were developed by Dai et al., who protected the 5-hydroxymethyl group as a *tert*-butylsilyl (TBS) ether. This moiety can be cleaved either by base treatment (NH_4OH, 65 °C, 15 h), or in a two-step procedure that includes tetrabutylammonium fluoride (TBAF).[52]



Scheme 6. Retrosynthesis of $[D_3]$mC, $[D_2]$hmC, fC, and caC from one precursor: 5-iododeoxycytidine.[15,51]

Drndic and co-workers incorporated a hmC triphosphate by PCR.[34a] In this way, they synthesized DNA strands containing only C, mC, or hmC, respectively. The strands were used to measure fluorescence annealing curves. The lowest annealing temperature was measured for DNA-containing hmC, medium values were obtained for C-containing DNA, and the highest ones for mC-DNA. The decrease in stability of a hmC-G base pair compared to mC-G and C-G base pairs was confirmed by molecular dynamics simulations. The authors concluded that the organization of water molecules around the base influences the duplex stability. The least polar nucleoside mC leads to increased rigidity and thereby stabilizes the duplex. These results were confirmed in sophisticated isothermal titration calorimetry studies, which revealed that hydroxylation of the methyl group exactly reverses the duplex-stabilization effect which is observed for mC.[47]

6. Conclusion and Outlook

After only two years of research, the new sixth base hmC has attracted tremendous attention in different research fields. It is already established that it is a major component in the mammalian genetic system and that it is generated from 5-methylcytosine with the help of TET enzymes. These enzymes oxidize the methyl group of 5-methylcyctosine and hence modulate the epigenetic marks set by the process of epigenetic imprinting. There is evidence that hmC is involved in celluar development processes and that it is connected to pluripotency. Furthermore, evidence is increasing that incorrect hydroxymethylation can lead to carcinogenesis. The result that hmC is found at high levels in the central nervous system and particularly in those brain areas that are associated with higher cognitive functions is very interesting.

Surprisingly, the hmC levels increase with age until they reach a final stable plateau. Young animals before birth or a short period thereafter possess low amounts of hmC even though the mC values are only slightly decreased. The hmC levels then increase up to a factor of five in mice shortly after

birth. It is currently not clear why, but one could speculate that hmC is involved in arresting cells at a final development stage. Future research will have to clarify the exact molecular consequences of hydroxymethylation. hmC-binding proteins have to be uncovered and a powerful methodology for the sequence-specific detection of hmC needs to be developed. This will lead to a more detailed understanding of hydroxymethylation and will help to complete the knowledge about epigenetic regulation mechanisms.

Note added in proof: While this manuscript was under review, four articles appeared that show the genome-wide distribution of hmC in stem cells.[40,53]

M.M. thanks the Fonds der Chemischen Industrie for a Kekulé fellowship. We thank the Deutsche Forschungsgemeinschaft for generous funding through the SFB 646 and 749 and the Normalverfahren, and we thank COST action "Free Radicals in Biology" for support. We are particularly grateful to the Excellence Cluster CiPS^{M} for supporting our research. We thank Dr. Markus Müller for careful reading of the manuscript.

Received: March 2, 2011
Published online: June 17, 2011

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737 – 738.
[2] a) J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204 – 220; b) R. Bonasio, S. J. Tu, D. Reinberg, *Science* **2010**, *330*, 612 – 616; c) P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068 – 1070.
[3] a) A. Meissner, *Nat. Biotechnol.* **2010**, *28*, 1079 – 1088; b) S. H. Feng, S. E. Jacobsen, W. Reik, *Science* **2010**, *330*, 622 – 627.
[4] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607 – 620.
[5] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929 – 930; b) M. Tahiliani et al., *Science* **2009**, *324*, 930 – 935.
[6] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781 – 790.
[7] R. M. Kothari, V. Shankar, *J. Mol. Evol.* **1976**, *7*, 325 – 329.
[8] a) G. R. Wyatt, S. S. Cohen, *Nature* **1952**, *170*, 1072 – 1073; b) G. R. Wyatt, S. S. Cohen, *Biochem. J.* **1953**, *55*, 774 – 782.
[9] a) R. L. Sinsheimer, *Science* **1954**, *120*, 551 – 553; b) E. Volkin, *J. Am. Chem. Soc.* **1954**, *76*, 5892 – 5893.
[10] a) R. A. J. Warren, *Annu. Rev. Microbiol.* **1980**, *34*, 137 – 158; b) J. H. Gommersampt, P. Borst, *FASEB J.* **1995**, *9*, 1034 – 1042.
[11] C. Bienvenu, J. R. Wagner, J. Cadet, *J. Am. Chem. Soc.* **1996**, *118*, 11406 – 11411.
[12] a) J. R. Wagner, J. Cadet, *Acc. Chem. Res.* **2010**, *43*, 564 – 571; b) E. Privat, L. C. Sowers, *Chem. Res. Toxicol.* **1996**, *9*, 745 – 750.
[13] a) G. D. Castro, C. J. Stamato, J. A. Castro, *Free Radical Biol. Med.* **1994**, *17*, 419 – 428; b) H. Cao, Y. Wang, *Nucleic Acids Res.* **2007**, *35*, 4833 – 4844.
[14] S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, *Nucleic Acids Res.* **1997**, *25*, 553 – 558.
[15] a) D. Globisch et al., *PLoS One* **2010**, *5*, e15367; b) M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem.* **2010**, *122*, 5503 – 5505; *Angew. Chem. Int. Ed.* **2010**, *49*, 5375 – 5377.
[16] a) J. J. Day, J. D. Sweatt, *Nat. Neurosci.* **2010**, *13*, 1319 – 1323; b) C. A. Miller, J. D. Sweatt, *Neuron* **2007**, *53*, 857 – 869; c) J. U. Guo, Y. Su, C. Zhong, G.-l. Ming, H. Song, *Cell* **2011**, *145*, 423 – 434. The Carell group discovered 5-formyl-2-deoxycytidine as a new base in DNA from embryonic stem cells, making the existence of an oxidative demethylation pathway according to Lit. [15a] plausible (T. Carell, oral communication, 9.5.2011, ISIS, Strasbourg).
[17] C.-X. Song, K. E. Szulwach Y. Fu, Q. Dai, C. Yi, X. 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.* **2011**, *29*, 68 – 72.
[18] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res.* **2010**, *38*, e181.
[19] C. J. Schofield, P. J. Ratcliffe, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 343 – 354.
[20] A. B. Robertson, J. A. Dahl, C. B. Vågbø, P. Tripathi, H. E. Krokan, A. Klungland, *Nucleic Acids Res.* **2011**, *39*, e55.
[21] M. Cross, R. Kieft, R. Sabatini, M. Wilm, M. de Kort, G. A. van der Marel, J. H. van Boom, F. van Leeuwen, P. Borst, *EMBO J.* **1999**, *18*, 6573 – 6581.
[22] a) J. Feng, Y. Zhou, S. L. Campbell, T. Le, E. Li, J. D. Sweatt, A. J. Silva, G. Fan, *Nat. Neurosci.* **2010**, *13*, 423 – 430; b) H. Cao, Y. Wang, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1335 – 1341; c) A. Burdzy, K. T. Noyes, V. Valinluck, L. C. Sowers, *Nucleic Acids Res.* **2002**, *30*, 4068 – 4074; d) T. Le, K.-P. Kim, G. Fan, K. F. Faull, *Anal. Biochem.* **2011**, *412*, 203 – 209.
[23] K. Iqbal, S.-G. Jin, G. P. Pfeifer, P. E. Szabó, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3642 – 3647.
[24] a) Z. Liutkevičiūtė, E. Kriukienė, I. Grigaitytė, V. Masevičius, S. Klimašauskas, *Angew. Chem.* **2011**, *123*, 2138 – 2141; *Angew. Chem. Int. Ed.* **2011**, *50*, 2090 – 2093.
[25] Z. Liutkevičiūtė, G. Lukinavicius, V. Masevičius, D. Daujotyte, S. Klimašauskas, *Nat. Chem. Biol.* **2009**, *5*, 400 – 402.
[26] a) P. O. Falnes, R. F. Johansen, E. Seeberg, *Nature* **2002**, *419*, 178 – 182; b) R. J. Hopkinson, R. B. Hamed, N. R. Rose, T. D. W. Claridge, C. J. Schofield, *ChemBioChem* **2010**, *11*, 506 – 510; c) Y. Shi, *Nat. Rev. Genet.* **2007**, *8*, 829 – 833; d) S. C. Trewick, T. F. Henshaw, R. P. Hausinger, T. Lindahl, B. Sedgwick, *Nature* **2002**, *419*, 174 – 178; e) C.-G. Yang, C. Yi, E. M. Duguid, C. T. Sullivan, X. Jian, P. A. Rice, C. He, *Nature* **2008**, *452*, 961 – 965.
[27] a) Y. Huang, W. A. Pastor, Y. H. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLoS One* **2010**, *5*, e8888; b) S.-G. Jin, S. Kadam, G. P. Pfeifer, *Nucleic Acids Res.* **2010**, *38*, e125; c) C. Nestor, A. Ruzov, R. Meehan, D. Dunican, *BioTechniques* **2010**, *48*, 317 – 319.
[28] a) H. Hayatsu, Y. Wataya, K. Kai, S. Iida, *Biochemistry* **1970**, *9*, 2858 – 2865; b) R. Shapiro, R. E. Servis, M. Welcher, *J. Am. Chem. Soc.* **1970**, *92*, 422 – 424; c) M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827 – 1831.
[29] M. Shiragami, H. Hayatsu, *Nucleic Acids Res.* **1976**, S31 – S32.
[30] 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 – 843.
[31] a) S. Bareyt, T. Carell, *Angew. Chem.* **2008**, *120*, 187 – 190; *Angew. Chem. Int. Ed.* **2008**, *47*, 181 – 184; b) M. Münzel, L. Lercher, M. Müller, T. Carell, *Nucleic Acids Res.* **2010**, *38*, e192; c) A. Okamoto, *Org. Biomol. Chem.* **2009**, *7*, 21 – 26.
[32] a) T. Davis, R. Vaisvila, *J. Vis. Exp.* **2011**, e2661; b) L. S. Shock, P. V. Thakkar, E. J. Peterson, R. G. Moran, S. M. Taylor, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3630 – 3635.
[33] A. Szwagierczak, A. Brachmann, C. S. Schmidt, S. Bultmann, H. Leonhardt, F. Spada, *Nucleic Acids Res.* **2011**, DOI: 10.1093/nar/gkr118.
[34] a) M. Wanunu et al., *J. Am. Chem. Soc.* **2011**, *133*, 486 – 492; b) E. V. B. Wallace, D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, T. J. Donohoe, H. Bayley, *Chem. Commun.* **2010**, *46*, 8195 – 8197.
[35] a) 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 – 465; b) D. Summerer, *ChemBioChem* **2010**, *11*, 2499 – 2501.

[36] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, *Cell Cycle* **2009**, *8*, 1698 – 1710.
[37] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129 – 1133.
[38] C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580 – 583.
[39] N. R. Rose, M. A. McDonough, O. N. F. King, A. Kawamura, C. J. Schofield, *Chem. Soc. Rev.* **2011**, DOI: 10.1039/C0CS00203H.
[40] G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398 – 402.
[41] M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, *Nat. Commun.* **2011**, *2*, 241.
[42] a) K. P. Koh, A. Yabuuchi, S. Rao, K. Cunniff, A. Laiho, M. Tahiliani, Y. Huang, E. Thompson, J. Nardone, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley, A. Rao, *Cell Stem Cell* **2011**, *8*, 200 – 213; b) J. Walter, *Cell Stem Cell* **2011**, *8*, 121 – 122.
[43] H. Wu, A. C. D'Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, Nature **2011**, *473*, 389 – 393.
[44] L. Dang, S. Jin, S. M. Su, *Trends Mol. Med.* **2010**, *16*, 387 – 397.
[45] L. Dang, D. W. White, S. Gross, B. D. Bennett, M. A. Bittinger, E. M. Driggers, V. R. Fantin, H. G. Jang, S. Jin, M. C. Keenan, K. M. Marks, R. M. Prins, P. S. Ward, K. E. Yen, L. M. Liau, J. D. Rabinowitz, L. C. Cantley, C. B. Thompson, M. G. Vander Heiden, S. M. Su, *Nature* **2009**, *462*, 739 – 744.
[46] a) W. Xu, H. Yang, Y. Liu, Y. Yang, P. Wang, S.-H. Kim, S. Ito, C. Yang, P. Wang, M.-T. Xiao, L. Liu, W. Jiang, J. Liu, J. Zhang, B. Wang, S. Frye, Y. Zhang, Y. Xu, Q. Lei, K.-L. Guan, S. Zhao, Y. Xiong, *Cancer Cell* **2011**, *19*, 17 – 30; b) R. Chowdhury, K. K. Yeoh, Y.-M. Tian, L. Hillringhaus, E. A. Bagg, N. R. Rose, I. K. H. Leung, X. S. Li, E. C. Y. Woon, M. Yang, M. A. McDonough, O. N. King, I. J. Clifton, R. J. Klose, T. D. W. Claridge, P. J. Ratcliffe, C. J. Schofield, A. Kawamura, *EMBO Rep.* **2011**, *12*, 463 – 469.
[47] A. Thalhammer, A. S. Hansen, A. H. El-Sagheer, T. Brown, C. J. Schofield, *Chem. Commun.* **2011**, *47*, 5325 – 5327.
[48] a) V. Valinluck, L. C. Sowers, *Cancer Res.* **2007**, *67*, 946 – 950; b) V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird, L. C. Sowers, *Nucleic Acids Res.* **2004**, *32*, 4100 – 4108.
[49] C. J. LaFrancois, J. Fujimoto, L. C. Sowers, *Chem. Res. Toxicol.* **1998**, *11*, 75 – 83.
[50] a) M. de Kort, P. C. de Visser, J. Kurzeck, N. J. Meeuwenoord, G. van der Marel, W. Rüger, J. H. van Boom, *Eur. J. Org. Chem.* **2001**, 2075 – 2082; b) A. S. Hansen, A. Thalhammer, A. H. El-Sagheer, T. Brown, C. J. Schofield, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1181 – 1184.
[51] M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671 – 5673.
[52] Q. Dai, C.-X. Song, T. Pan, C. He, *J. Org. Chem.* **2011**, *76*, 4182 – 4188.
[53] a) W. A. Pastor et al., *Nature* **2011**, *473*, 394 – 397; b) K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343 – 348; c) H. Wu, A. C. D'Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. Eve Sun, Y. Zhang, *Nature* **2011**, *473*, 389 – 393.

