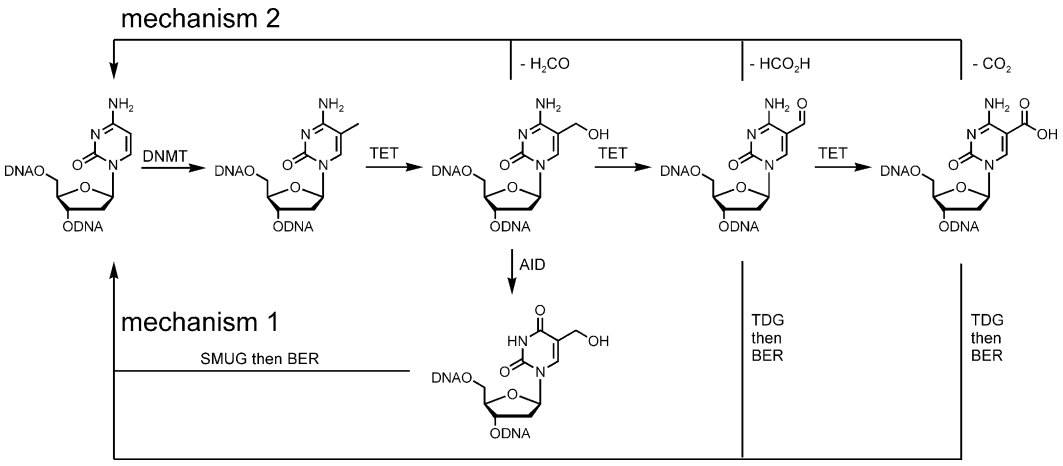
Epigenetics

DOI: 10.1002/anie.201202583

Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing**

Stefan Schiesser, Benjamin Hackner, Toni Pfaffeneder, Markus Müller, Christian Hagemeier, Matthias Truss, and Thomas Carell**

5-Methylcytosine (mC) is an important, well-known nucleobase modification that is involved in many biological processes, including gene expression, genomic imprinting, X-chromosome inactivation, and suppression of transposable elements.[1–3] Recently it was discovered that mC can be further processed in neurons and in embryonic stem cells by TET enzymes, which utilize 2-ketoglutarate, to oxidize mC first to 5-hydroxymethylcytosine (hmC)[4,5] and further to 5-formylcytosine (fC)[6] and 5-carboxycytosine (caC)[7,8] (Scheme 1).[9,10] It is currently believed that this oxidation chemistry is involved in a process of active demethylation, which allows cells to exchange mC with unmodified 2′-deoxycytidine (dC) to revert the biological effects caused by cytosine methylation.[11] Two possible active demethylation mechanisms are currently being intensively investigated (Scheme 1). Mechanism 1 is based on special glycosylases,[12–15] such as the thymine–DNA glycosylase (TDG), which was shown to also cleave the glycosidic bonds of fC and caC, as well as of 5-hydroxymethyluridine (hmU).[15,16] The latter is formed after deamination of hmC.[15] This base cleavage reaction first generates an abasic site, which is repaired by the base excision repair system (BER) by insertion of dC (Scheme 1, lower part). This whole process leads in summary to an exchange of mC by dC. The second mechanism is thought to involve retro-Aldol-type chemistry



Scheme 1. Active demethylation pathways involving elimination (mechanism 2) of formaldehyde from hmC, formic acid from fC, or decarboxylation of caC. In contrast, mechanism 1 involves the base excision repair (BER)-based replacement of fC and caC by dC, or of hmC by dC after prior deamination of hmC to hmU.

[*] S. Schiesser,[+] B. Hackner,[+] T. Pfaffeneder, M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians Universität München
Butenandtstrasse 5–13, 81377 München (Germany)
E-mail: matthias.truss@charite.de
thomas.carell@lmu.de

Dr. C. Hagemeier, Dr. M. Truss
Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie
Ziegelstrasse 5–9, 10098 Berlin (Germany)

[+] These authors contributed equally to this work.

[**] We thank the Excellence cluster CiPS[M] and the collaborative research center SFB749 as well as the DFG, grant CA275/8-4, the Volkswagen Foundation and NGFNplus (01GS0870) for financial support. T.P. and S.S. thank the Fonds der Chemischen Industrie for pre-doctoral fellowships.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201202583.

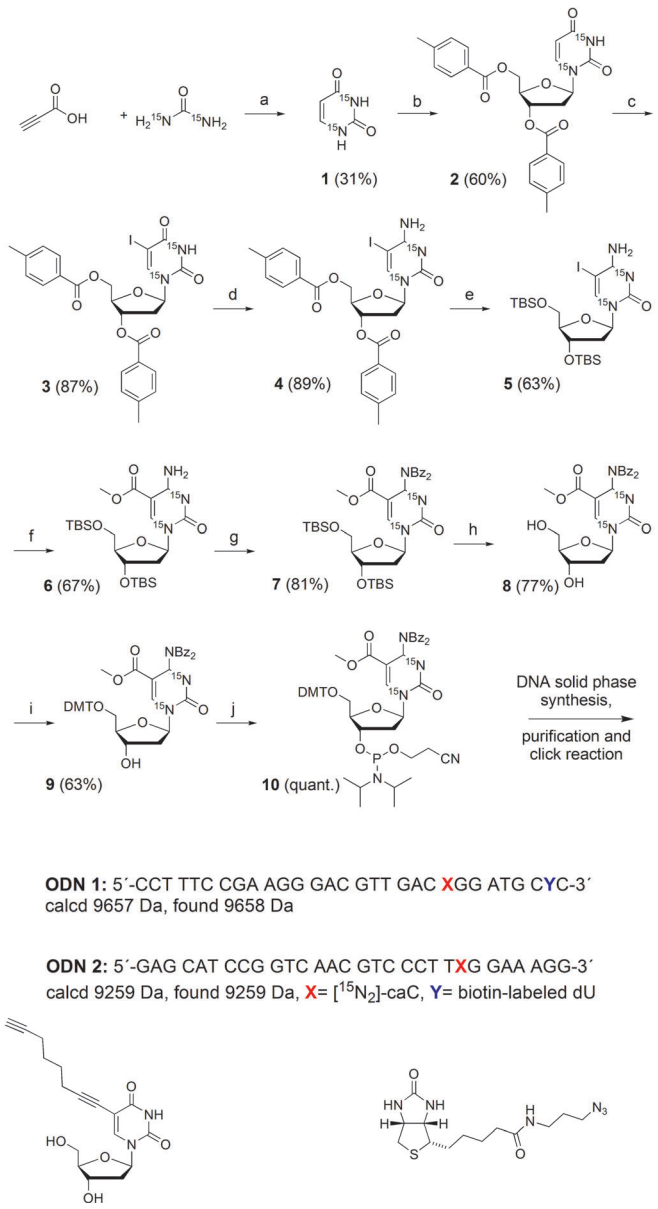
starting with hmC[17,18] or fC. Alternatively, decarboxylation of caC (Scheme 1) is a highly attractive alternative, because this mechanism allows exchange of mC by dC without formation of intermediate strand breaks, which are DNA lesions that are known to cause genome instability, and without formation of potentially harmful side products, such as formaldehyde. Although decarboxylation of caC has never been observed, similar transformations are known for orotate and isoorotate, which are decarboxylated to uracil.[19]

To gain deeper insight into the putative decarboxylation of caC in stem cells, we developed a sensitive isotope tracing experiment. A double ^{15}N-labeled caC phosphoramidite reagent [$^{15}N_2$]-caC was developed, which we incorporated into the 30 mer oligonucleotide **D1** with a sequence from the Ecat1 promoter[23] that is known to be the subject of active demethylation. The natural occurrence of the double ^{15}N isotopologue of dC ([$^{15}N_2$]-dC) is extremely low, which allows us to monitor the chemistry that operates on the caC

compound by high-resolution mass spectrometry. The synthesis of the [$^{15}N_2$]-caC phosphoramidite is depicted in Scheme 2 (see also the Supporting Information). The starting point is [$^{15}N_2$]-uracil **1**, which was prepared from [$^{15}N_2$]-urea and propiolic acid.[20] Vorbrüggen nucleosidation with the bis(toluoyl)-protected chlororibofuranoside[21] furnished the β-configured nucleoside **2** in 60% yield. Electrophilic iodination to **3**,[22] conversion of the C4-keto function into the amine as needed for the cytosine base (**4**), and protecting-group exchange gave the key intermediate **5**. Pd-catalyzed CO insertion and quenching of the Pd–CO intermediate with methanol provided the [$^{15}N_2$]-caC precursor **6** in about 65% yield. Double benzoyl protection of the C4-amino group (**7**) and cleavage of the TBS groups (**8**) was followed by protection of the 5′OH group with dimethoxytritylchloride (DMT-Cl; **9**). We subsequently converted the 3′OH group into the final phosphoramidite **10** using standard procedures. The phosphoramidite was used to synthesize the 30 mer oligonucleotides **ODN1** and **ODN2**. The two strands were finally hybridized to give the DNA duplex **D1** with the sequence of the Ecat1 promoter containing two [$^{15}N_2$]-labeled caC derivatives.

In one of the DNA strands (**ODN1**), we also inserted the base 5-octadienyl-dU (Scheme 2, bottom) and subsequently used mild click chemistry with the depicted biotin azide (Scheme 2, bottom; Supporting Information) to attach a biotin label to 5-octadienyl-dU, giving the biotinylated uridine Y.[23] All of the DNA strands needed for the experiment were carefully purified by HPLC to obtain the labeled DNA material in extremely high purity (>99.9%). After successful click chemistry, we purified the biotinylated oligonucleotide again using a second HPLC purification step. The identity of the oligonucleotides was confirmed in all cases by MALDI-TOF mass spectrometry (Scheme 2; Supporting Information).

For the isotope tracing experiment depicted in Figure 1a, we added the DNA duplex **D1** (**ODN1** + **ODN2**) containing the two [$^{15}N_2$]-labeled caC bases and the biotin tag to nuclear extracts of mouse embryonic stem cells (mESC). After 1 h of incubation at 37 °C (for details see Supporting Information), we re-isolated the biotinylated oligonucleotide with the help of streptavidin-bound magnetic beads. The isolated DNA was subsequently totally digested (Supporting Information) and the digest was analyzed by HPLC-MS using a high-resolution mass spectrometer (Orbitrap XL; for details, see the Supporting Information). The obtained data are depicted in Figure 1b–e. We first investigated the results of the control experiments. To this end we treated **D1** with non-mESC nuclear extracts and we incubated **D1** with buffer only. The mass-spectrometric analysis of the control experiments (Figure 1d,e) provided two distinct signals with $m/z = 230.1014$ and 230.0976 in the mass range of the dC nucleoside (Figure 1e). These m/z signals correspond to the naturally occurring isotopologues [$^{13}C_2$]- and [^{18}O]-dC, respectively, and [^{15}N,^{13}C]-dC. The exact mass expected for double ^{15}N-containing [$^{15}N_2$]-dC was not detected in all of these control experiments, even when we digested and analyzed higher amounts of DNA (2.5 nmol dC vs 0.9 nmol dC; Figure 1d,e). We then analyzed the digests obtained from **D1** after incubation with mESC nuclear extract. Surprisingly we noted that after 1 h of incubation with mESC nuclear extracts, the [$^{15}N_2$]-labeled dC species was indeed detected in the total digest of the re-isolated duplex **D1**. The compound [$^{15}N_2$]-dC eluted with a retention time of 7.18 min. The obtained mass



Scheme 2. Synthesis of the [$^{15}N_2$]-labeled caC phosphoramidite and of the DNA strands **ODN1** and **ODN2** containing one [$^{15}N_2$]-labeled caC nucleobases each. The biotin label was inserted using click chemistry. Reagents and conditions: a) polyphosphoric acid, 95 °C, 19 h; b) 1. HMDS, TMS-Cl, 120 °C, 1 h; 2. Hoffer's chlorosugar, $CHCl_3$, RT, 2.5 h; c) CAN, LiI, MeCN, 80 °C, 2 h; d) 1. 1,2,4-triazole, $POCl_3$, TEA, 30 °C, overnight; 2. NH_4OH, 1,4-dioxane, RT, 10 min; e) 1. K_2CO_3, MeOH, RT, 19 h; 2. TBS-Cl, imidazole, DMF, RT, 4 days; f) [$Pd(MeCN)_2Cl_2$], CO, MeOH, 60 °C, 17 h; g) BzCl, pyridine, 0 °C to RT, 19 h; h) HF-pyridine, EtOAc, RT, 14 h; i) DMT-Cl, pyridine, RT, 16.5 h; j) Diisopropylammonium tetrazolide, 2-cyanoethyl-*N*,*N*,*N*′,*N*′-tetraisopropylphosphordiamidite, CH_2Cl_2, RT, 17 h. HMDS = Hexamethyldisilazane, TMS = trimethylsilyl, CAN = cerium ammonium nitrate, TEA = triethylamine, TBS = *tert*-butyldimethylsilyl, DMT = dimethoxytrityl.

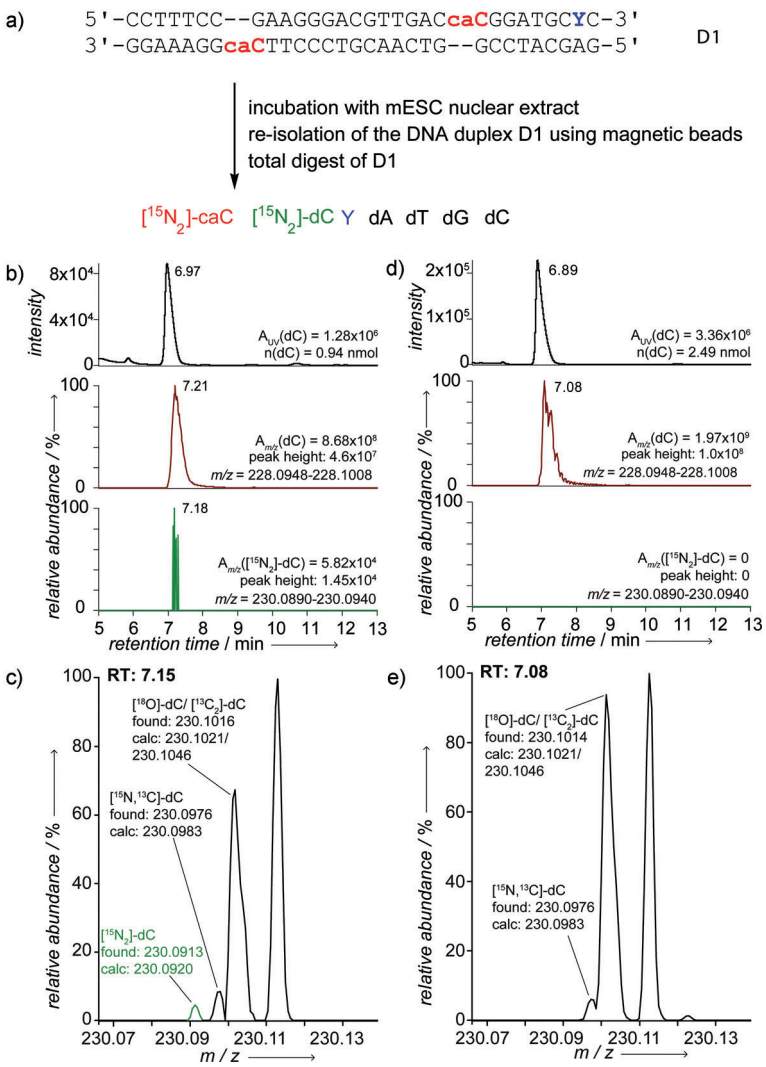


Figure 1. a) Depiction of the nuclear extract experiment. b, d) HPLC-MS analysis of **D1** treated with mESC nuclear extract (b) and non-mESC extract (d). UV traces of HPLC runs at the retention time of dC (top), as well as the mass traces for dC (228.0948–228.1008, middle), and the mass traces for $[^{15}N_2]$-labeled dC (230.0890–230.0940, bottom). c, e) High-resolution mass spectra (m/z = 230.0700–230.1300) of D1 treated with mESC nuclear extract (c) and non-mESC extract (e).

spectrometry signal has the exact mass of $m/z_{found} = 230.0913$ (Figure 1c). The signal is nicely separated from the signals of the other naturally occurring isotopologues $[^{15}N,^{13}C]$-dC (m/z=230.0976) and $[^{18}O/^{13}C_2]$-dC (m/z = 230.1016) and thus is clearly detectable. Furthermore, the measured high-resolution data for the compound $[^{15}N_2]$-dC agrees very well with the calculated exact mass of $[^{15}N_2]$-dC (m/z_{calcd} = 230.0920). The data show all together that mESC nuclear extract has the capability to decarboxylate caC to dC. Whether this activity is caused by a specific enzyme or by other factors present in the extract requires further investigation.

To investigate the reactions that would allow stem cells to decarboxylate caC, we first heated the caC nucleoside and an oligonucleotide containing a caC base. However, under no circumstances was decarboxylation observable. Even when the caC nucleoside was heated to reflux in water for several hours, the mass spectrometric analysis of the reaction solution provided no evidence for decarboxylation (data not shown). This is not surprising because decarboxylation would furnish a high-energy cytosine carbanion. In contrast, saturation of the C5–C6 double bond followed by tautomerization of the C4-amino group would give an imine intermediate with the substructure of a β-iminocarboxylic acid, which would allow efficient decarboxylation (Figure 2).[10,24] To investigate this idea, we reduced the 5-carboxy-dC methyl ester **11** with $NaBH_4$. The reduced compound **12** is stable and was isolated and fully characterized (Figure 2a; Supporting Information). However, upon saponification of the methyl ester, rapid decarboxylation to compound **13** is observed. Subsequent treatment of the dihydro-dC derivative with DDQ furnishes the TBS-protected compound dC. This result shows that saturation of the C5–C6 double bond is critical for efficient decarboxylation, as suggested recently.[10] As reaction with a H^- species is unlikely to occur naturally, we then treated TBS-protected caC with a thiol/imidazole mixture, hoping for a more temporary 1,4-addition followed by decarboxylation and elimination. This chemical cascade is indeed observed (data not shown). Treatment of the caC nucleoside with octane thiol in presence of imidazole gave rise to a new spot on the TLC plate. This spot was not observed in the absence of imidazole, showing that some proton catalysis is required for the reaction. Isolation of the compound and full characterization (Supporting Information) showed that the new compound is indeed the TBS-protected dC. This compound is generated by decarboxylation followed by elimination of the added thiol. To investigate if this chemistry would be compatible with the decarboxylation of caC embedded in an oligonucleotide (Figure 2b,c), we again used the isotope tracing method. For the experiment, we incubated the $[^{15}N_2]$-caC-containing DNA strand **ODN2** with different amino acids. Indeed, when we added the amino acid cysteine or combinations of cysteine with histidine/arginine to an aqueous solution of **ODN2**, decarboxylation was detectable (Figure 2b). To confirm that decarboxylation is occurring, we re-isolated the DNA after reaction overnight, digested the oligonucleotide down to the nucleoside level and analyzed the mixture using high resolution HPLC-MS. Figure 2b (left) shows the data obtained for the cysteine/histidine combination. The data depicted in Figure 2b (right) are the results of the cysteine/arginine experiment. In both cases, the $[^{15}N_2]$-dC signal with an m/z value of 230.0922 was clearly detected. The amount of decarboxylation increases with an increasing concentration of thiol in solution and was in total around 5% (half-saturated solution of amino acids, 50°C, overnight).

In summary, using isotope tracing with a DNA strand in which we incorporated double ^{15}N-labeled caC, we could gain first evidence that stem-cell nuclear extracts have the ability

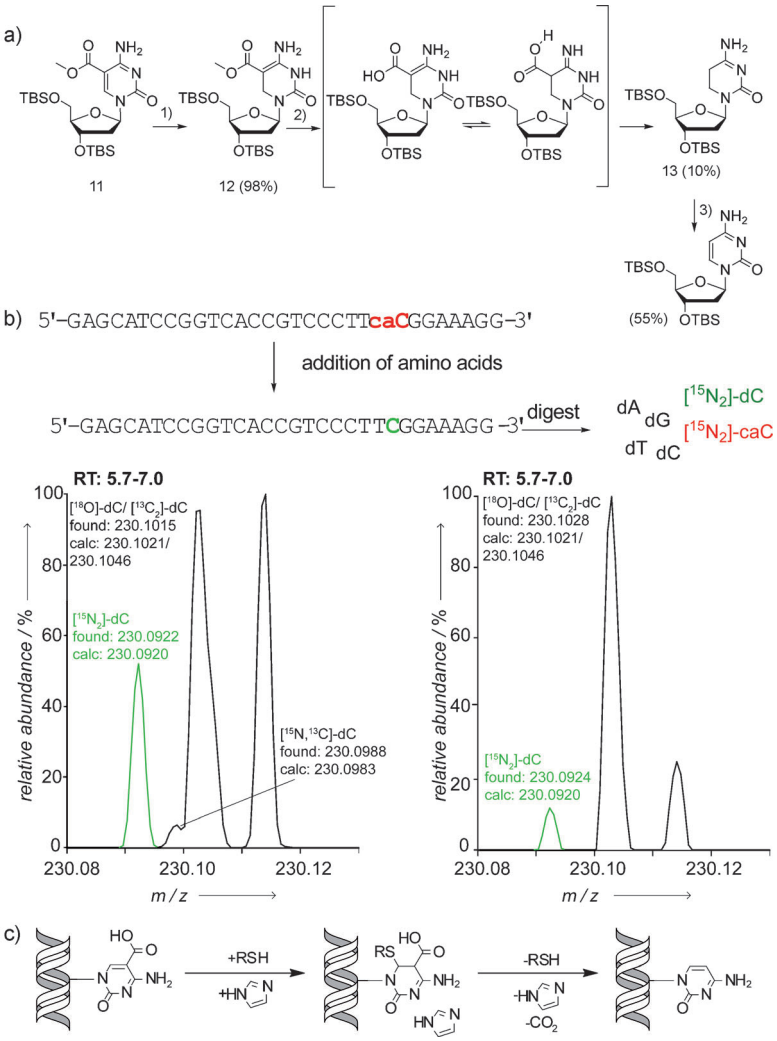


Figure 2. a) Reaction of caC with $NaBH_4$ causes decarboxylation: 1) $NaBH_4$, MeOH, RT, 20 min; 2) LiOH, MeCN/H_2O, RT, 2 h; 3) DDQ, 1,4-dioxane, RT, 25 h. b) Incubation of caC containing DNA with cysteine and with cysteine in combination with histidine or arginine leads to decarboxylation. Analysis was performed after total digest of the treated DNA using high-resolution HPLC-MS. Illustration of the mechanism of decarboxylation triggered by nucleophilic attack of the C6 position.

to decarboxylate caC. Chemical model studies show that this decarboxylation is accelerated by saturation of the C5–C6 double bond of caC. This saturation is for example achieved after attack of the electrophilic C6 position of caC by a thiol. The C6-reacted caC will decarboxylate and re-aromatize. Notably, the methylation of dC to mC by the different DNMT (methyltransferase) enzymes proceeds by a quite similar mechanism. Here the C6 carbon of dC is first attacked by a thiol nucleophile, followed by reaction of the C5 atom with the coenzyme *S*-adenoylmethionine (SAM), which provides a CH_3^+-ion equivalent. These enzymes have already been associated with demethylation activity. For example, they were shown to eliminate formaldehyde from hmC, converting it into dC.[17,25] Another enzymatic reaction on pyrimidine nucleobases, in which the C5–C6 double bound is temporarily saturated is performed by isoorotate decarboxylase (IDCase), which activates water with the help of a Zn^{2+} ion so that an hydroxide anion is generated that is thought to react with the C6 position. This reaction is followed by a decarboxylation that converts isoorotate to uracil.[26,27] The chemical reactivity that we suggest is consequently in agreement with other known enzymatic transformations. We believe that the decarboxylation of caC by stem cells is an important mechanism in the framework of active genome demethylation.

Received: April 3, 2012
Published online: May 29, 2012

Keywords: 5-carboxycytosine · decarboxylation · epigenetics · isotopic labeling · stem cells

[1] H. Cedar, Y. Bergman, *Nat. Rev. Genet.* **2009**, *10*, 295–304.
[2] M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **2008**, *9*, 465–476.
[3] H. Sasaki, Y. Matsui, *Nat. Rev. Genet.* **2008**, *9*, 129–140.
[4] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
[5] 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.
[6] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146–7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.
[7] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
[8] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
[9] K. I. Ladwein, M. Jung, *Angew. Chem.* **2011**, *123*, 12347–12349; *Angew. Chem. Int. Ed.* **2011**, *50*, 12143–12145.
[10] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *ChemBioChem* **2011**, *12*, 206–222.
[11] S. K. Ooi, T. H. Bestor, *Cell* **2008**, *133*, 1145–1148; M. R. Branco, G. Ficz, W. Reik, *Nat. Rev. Genet.* **2012**, *13*, 7–13.
[12] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101–1105.
[13] M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Scholer, J. Walter, *EMBO J.* **2010**, *29*, 1877–1888.
[14] P. Hajkova, S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, M. A. Surani, *Science* **2010**, *329*, 78–82.
[15] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67–79.
[16] K. A. Haushalter, M. W. T. Stukenberg, M. W. Kirschner, G. L. Verdine, *Curr. Biol.* **1999**, *9*, 174–185.
[17] Z. Liutkevičiutė, G. Lukinavicius, V. Masevičius, D. Daujotytė, S. Klimašauskas, *Nat. Chem. Biol.* **2009**, *5*, 400–402.

[18] Z. Liutkeviciute, E. Kriukienė, I. Grigaitytė, V. Masevičius, S. Klimašauskas, *Angew. Chem.* **2011**, *123*, 2138 – 2141; *Angew. Chem. Int. Ed.* **2011**, *50*, 2090 – 2093.
[19] J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes Sr., A. A. Axhemi, *Biochim. Biophys. Acta Gen. Subj.* **2005**, *1723*, 256 – 264.
[20] C. J. LaFrancois, J. Fujimoto, L. C. Sowers, *Chem. Res. Toxicol.* **1998**, *11*, 75 – 83.
[21] G. H. Clever, C. Kaul, T. Carell, *Angew. Chem.* **2007**, *119*, 6340 – 6350; *Angew. Chem. Int. Ed.* **2007**, *46*, 6226 – 6236.
[22] J. Asakura, M. J. Robins, *J. Org. Chem.* **1990**, *55*, 4928 – 4933.
[23] G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *474*, 398 – 402; P. M. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew. Chem.* **2008**, *120*, 8478 – 8487; *Angew. Chem. Int. Ed.* **2008**, *47*, 8350 – 8358.
[24] N. K. Thalji, W. E. Crowe, G. L. Waldrop, *J. Org. Chem.* **2009**, *74*, 144 – 152.
[25] R. Métivier, R. Gallais, C. Tiffoche, C. Le Peron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, *Nature* **2008**, *452*, 45 – 50.
[26] A. Liu, H. Zhang, *Biochemistry* **2006**, *45*, 10407 – 10411.
[27] R. D. Palmatier, R. P. McCroskey, M. T. Abbott, *J. Biol. Chem.* **1970**, *245*, 6706 – 6710.

