

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie  
der Ludwigs-Maximilians Universität München

**Epigenetic dynamics of cytidine modifications**

Benjamin Hackner

aus

München

2014

### **Erklärung**

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

### **Eidesstattliche Versicherung:**

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

Benjamin Hackner, München den 26. Mai 2014

Dissertation eingereicht am:

03.06.2014

1. Gutachter:

Prof. Dr. T. Carell

2. Gutachter:

PD Dr. Dietmar Martin

Mündliche Prüfung am:

03.07.2014



*Für meine Familie*

*„Sprachen sind der Schlüssel zum Weltfrieden.“*

**Batman**



# Danksagung

Ich möchte mich bei zahlreichen Personen für die fachliche und persönliche Unterstützung während meiner Promotion bedanken.

An erster Stelle möchte ich mich bei meinem Doktorvater *Prof. Dr. Thomas Carell* für das spannende Thema und die exzellenten Arbeitsbedingungen bedanken, sowie meinen Aufenthalt in Berlin bedanken.

*Herrn PD Dr. Dietmar Martin* danke ich herzlich für die Übernahme des Koreferats und seine ansteckende Begeisterung für Chemie.

Ebenso danke ich den Mitgliedern der Prüfungskommission: *Prof. Dr. A Vollmar, PD Dr. S. Michalakis, Prof. Dr. P. Knochel* und *Prof. Dr. A. Hoffmann-Röder*

Frau *Slava Gärtner* danke ich für ihre Hilfe in allen organisatorischen Belangen. Außerdem geht mein Dank an Frau *Sabine Voß* für das Organisieren all der großen und kleinen Dinge, die man in der Chemie benötigt sowie der seelisch moralischen Unterstützung.

*Toni Pfaffeneder* möchte ich für die enge Zusammenarbeit die den großen Erfolg unseres Themas erst ermöglichten danken. Ich konnte viel lernen und habe einen guten Freund gewonnen, Danke.

*Dr. Markus Müller* möchte ich für seine stetigen Anregungen, seine biochemischen Hilfestellungen und Diskussionsbereitschaft danken.

Einen großen Dank gehen an die Korrekturleser, *Johannes Harder, Markus Müller, Stefan Schiesser, Jessica Steinbacher, Toni Pfaffeneder* und *Sandra Koch*.

*Johannes Harder* und *Ulrike Lischke* möchte ich für viele gemeinsame Momente danken die stets gute Laune hinterließen.

Meinen Praktikanten und Masterranden, insbesondere *Edris Parsa* möchte ich für die gute Zusammenarbeit danken. Ich wünsche euch noch eine schöne Zeit im AK und anderswo.

Dem ganzen Arbeitskreis *Carell, insbesondere meinen Laboren F4.001 und F.4004* danke ich für die produktive und freundschaftliche Arbeitsatmosphäre, in die ich mich gerne eingebracht habe.

Viele Kollegen konnte ich dieser Zeit als Freunde gewinnen, euch alle zu nennen wäre uferlos. Ihr habt mir geholfen neben den vielen schönen Momenten auch die traurigen und schweren Zeiten zu überstehen. Ohne euch wäre ich nicht so weit gekommen, Danke.

Eine Vielzahl von Personen hat mich im Laufe meiner Doktorarbeit abseits des Labors begleitet und unterstützt, insbesondere *Erwin Wiesenmayer, Jan Schwarz, Mona Calik, Fabian Hanusch, Jan Schwarz* und *Jürgen Eirich*. Ihre Freundschaft, Liebe und Unterstützung haben diese Arbeit erst möglich gemacht. Ihnen gilt mein besonderer Dank.

Batman möchte ich für die vielen Weisheiten danken, Danke Batman.

Meinen Eltern und meinem Bruder möchte ich für ihre Liebe, Unterstützung und ihr Vertrauen danken.

## **Publications and presentations of this thesis:**

Reprinted with permission,

2014: Toni Pfaffeneder, Fabio Spada, Mirko Wagner, Caterina Brandmayr, Silvia Laube, David Eisen, Matthias Truss, Jessica Steinbacher, **Benjamin Hackner**, Olga Kotljarova, David Schuermann, Stylianos Michalakis, Olesea Kosmatchev, Stefan Schiesser, Barbara Steigenberger, Nada Raddaoui, Udo Müller, Heinrich Leonhardt, Primo Schär, Markus Müller and Thomas Carell ; *Nat. Chem. Bio.* **2014**, *accepted*

“Tet enzymes oxidize T to hmU in mouse embryonic stem cells“

P. Bashtrykov, A. Rajavelu, **B. Hackner**, S. Ragozin, T. Carell and A. Jeltsch,

*ChemBioChem*, **2014**, *accepted*, DOI: 10.1002/cbic.201300740

“Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain”

2013: S. Schiesser,\* T. Pfaffeneder,\* K. Sadeghian, **B. Hackner**, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *JACS* **2013**, *135*, 14593-14599. (\* Equal contribution)

“Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine“

B. Steigenberger,\* S. Schiesser,\* **B. Hackner**, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366-369. (\* Equal contribution)

“Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and their Incorporation into Oligonucleotides by Polymerase Chain Reaction”

2012: Poster Präsentation auf der IMB Konferenz “Demethylation, DNA Repair and Beyond” in Mainz: T. Pfaffeneder,\* **B. Hackner**,\* S. Schiesser, M. Truss, M. Muller, T. Carell (\* Equal contribution)

“Discovery of fC in stem cells and mechanistic insights into decarboxylation, deformylation and dehydroxymethylation”

S. Schiesser,\* **B. Hackner**,\* T. Pfaffeneder, M. Muller, C. Hagemeyer, M. Truss, T. Carell,  
*Ang. Ch., Int. Ed.* **2012**, *51*, 6516-6520. (\* Equal contribution)

“Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing”

T. F. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Munzel, M. Muller, T. Pfaffeneder, **B. Hackner**, W. Feiden, U. Schuller, T. Carell, H. A. Kretzschmar, *Int.J. Can.* **2012**, *131*, 1577-1590.

“Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base", are associated with anaplasia in human brain tumors”

2011: T. Pfaffeneder,\* **B. Hackner**,\* M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeyer, T. Carell, *Ang. Ch., Int. Ed.* **2012**, *50*, 7008-7012. (\* Equal contribution)

“The discovery of 5-formylcytosine in embryonic stem cell DNA”

Copyright of the publications belong to the publishers.



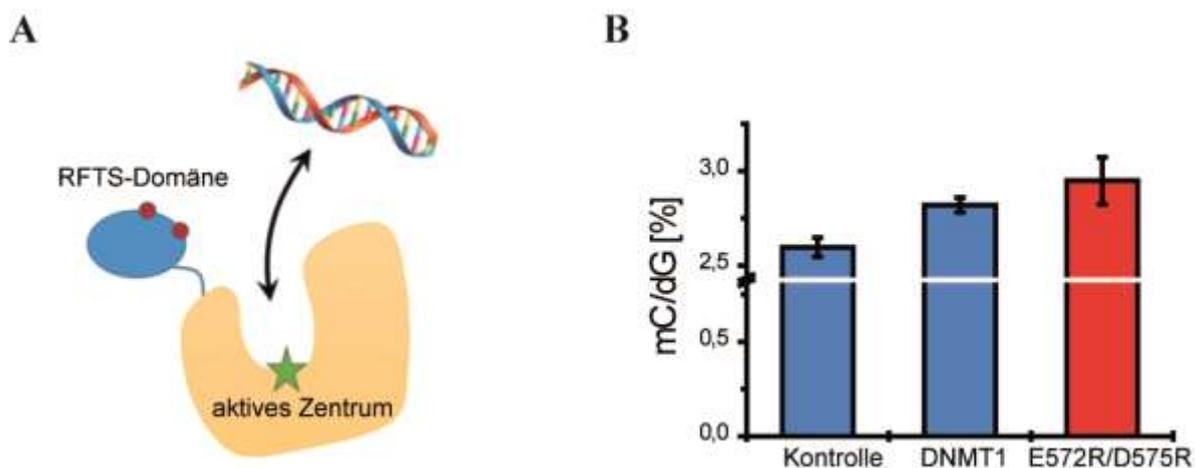
# Table of Contents

ZUSAMMENFASSUNG: .....	II
SUMMARY: .....	VI
<b>I. INTRODUCTION .....</b>	<b>1</b>
1. PLURIPOTENT CELLS IN MODERN MEDICINE .....	1
2. EPIGENETICS.....	3
2.1 <i>Histone modifications</i> .....	4
2.2 <i>DNA methylation</i> .....	6
2.2.1 <i>DNA Methyltransferases</i> .....	8
2.3 <i>DNA Demethylation</i> .....	14
2.3.1 <i>Base excision repair of mC</i> .....	15
2.3.2 <i>Deamination of mC</i> .....	16
2.3.3 <i>Oxidation of mC</i> .....	17
2.3.3.1 <i>TET</i> .....	18
2.3.3.2 <i>Hydroxymethylcytosine</i> .....	21
2.3.3.3 <i>Other oxidized pyrimidine species</i> .....	22
2.3.3.4 <i>Excision of new epigenetic bases by BER</i> .....	24
2.3.3.5 <i>C-C bond cleavage pathway</i> .....	25
<b>II. AIM OF THESIS.....</b>	<b>29</b>
<b>III. CONCLUSION AND OUTLOOK .....</b>	<b>30</b>
<b>IV. LITERATURE .....</b>	<b>32</b>
<b>V. THE DISCOVERY OF 5-FORMYLCYTOSINE IN EMBRYONIC STEM CELL DNA.....</b>	<b>42</b>
<b>VI. MECHANISM AND STEM-CELL ACTIVITY OF 5-CARBOXYCYTOSINE DECARBOXYLATION DETERMINED BY ISOTOPE TRACING.....</b>	<b>48</b>
<b>VII. SYNTHESIS OF 5-HYDROXYMETHYL-, 5-FORMYL-, AND 5-CARBOXYCYTIDINE-TRIPHOSPHATES AND THEIR INCORPORATION INTO OLIGONUCLEOTIDES BY POLYMERASE CHAIN REACTION .....</b>	<b>54</b>
<b>VIII. TARGETED MUTAGENESIS RESULTS IN AN ACTIVATION OF DNA METHYLTRANSFERASE 1 AND CONFIRMS AN AUTOINHIBITORY ROLE OF ITS RFTS DOMAIN .....</b>	<b>59</b>
<b>IX. DEAMINATION, OXIDATION, AND C–C BOND CLEAVAGE REACTIVITY OF 5- HYDROXYMETHYLCYTOSINE, 5-FORMYLCYTOSINE, AND 5-CARBOXYCYTOSINE .....</b>	<b>66</b>
<b>X. LOW VALUES OF 5-HYDROXYMETHYLCYTOSINE (5HMC), THE "SIXTH BASE", ARE ASSOCIATED WITH ANAPLASIA IN HUMAN BRAIN TUMORS.....</b>	<b>74</b>
<b>XI. TET ENZYMES OXIDIZE T TO HMU IN MOUSE EMBRYONIC STEM CELLS.....</b>	<b>90</b>
<b>XII. DNA METHYLATION IN THE CONTEXT OF OXIDIZED CYTIDINE DERIVATIVES.....</b>	<b>102</b>
<b>XIII. DYNAMICS OF PYRIMIDINES DURING METHYLTRANSFERASE TREATMENT .....</b>	<b>111</b>
<b>XIV. ABBREVIATIONS.....</b>	<b>117</b>

## Zusammenfassung:

In dieser Doktorarbeit wurde der 5-Methylcytosin (mC) Gehalt von Zellen, die von *Pavel Bashtrykov* mit verschiedenen DNA-Methyltransferase 1 (DNMT1) Mutanten transfiziert wurden, mittels einer LC-MS/MS Isotopen-Verdünnungsmethode bestimmt. Dabei konnte gezeigt werden, dass eine Mutation der Aminosäuren 572 und 575 von Glutamat bzw Aspartat zu Arginin einen erhöhten Level an mC zeigt. Dies konnte durch den Kooperationspartner mit einer *in vitro*-Aktivitätsbestimmung bestätigt werden, in der die Mutante sogar noch deutlich höhere Aktivitäten als der Wildtyp zeigten.<sup>[1]</sup>

Dadurch konnte eine Hypothese bestätigt werden, nach der das Enzym durch die Bindung einer N-terminalen Proteindomäne über ionische Kontakte an den katalytischen Teil des Proteins im C-Terminus reguliert wird (**Abb. 1**).<sup>[2]</sup>

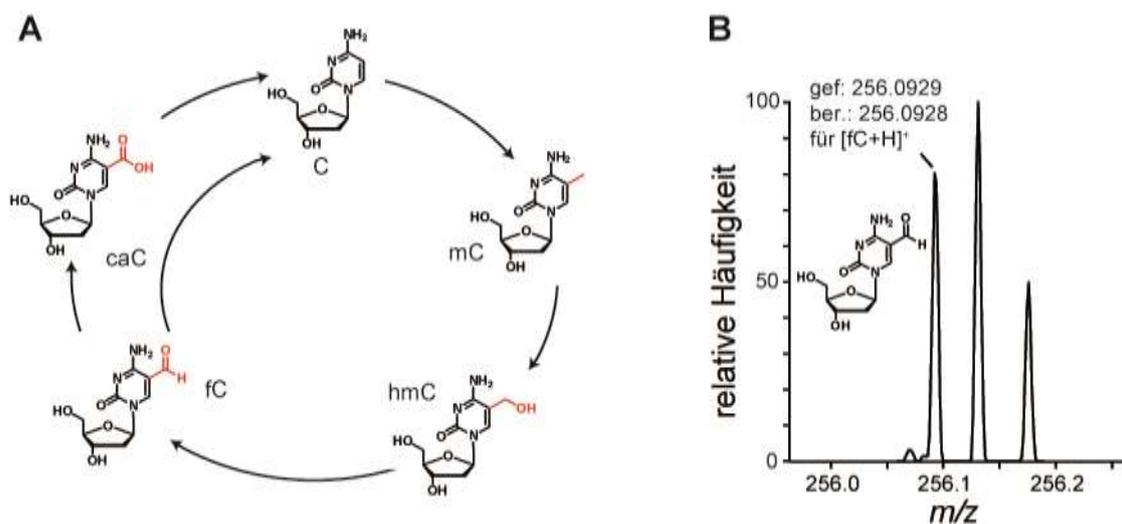


**Abbildung 1:** A) Schematische Sicht der allosterischen Selbstregulation der RFTS Domäne von DNMT1 und der Mutationsseiten (rot) B) Darstellung des 5-Methylcytosin Gehalts in genomischer DNA der Kontrolle und des Wildtyps (blau) und der verwendeten Mutante (rechts, rot). Hierbei sind leicht erhöhte Werte der Mutante erkennbar.

Seit 2009 ist die Oxidation des mCs durch die *ten eleven translocation* (TET) Proteine zu 5-Hydroxymethyldeoxycytidin (hmC) bekannt und ein oxidativer Weg zur aktiven Demethylierung wurde vorgeschlagen.<sup>[3-5]</sup> Mittels einer LC-MS Isotopen-Verdünnungsmethode, die insbesondere von *Daniel Globisch* entwickelt wurde, konnte in dieser Doktorarbeit der Gehalt von hmC in einigen Hirntumorgeweben bestimmt werden.<sup>[6]</sup> Es konnte gezeigt werden, dass diese Modifikation in den untersuchten Tumortypen stark

reduziert ist und ein geringer hmC-level als Tumormarker in Betracht gezogen werden könnte.

In einem anderen Teil dieser Arbeit konnte gemeinsam mit *Toni Pfaffeneder* eine weitere Oxidationsstufe des Cytidins (dC) in der DNA von embryonalen Mäusestammzellen (mESC), das 5-Formyldeoxycytidin (fC), mit Hilfe modernster Massenspektrometrie entdeckt und grob quantifiziert werden (**Abb. 2**).<sup>[7]</sup> Es wurde vorgeschlagen, dass es sich hierbei um ein weiteres Intermediat der oxidativen aktiven Demethylierung handelt und von den TET Proteinen gebildet wird. Darüber hinaus konnte es erfolgreich chemisch modifiziert werden, was unter anderem dazu beitrug die genomische Verteilung der Modifikation zu bestimmen.<sup>[8]</sup>



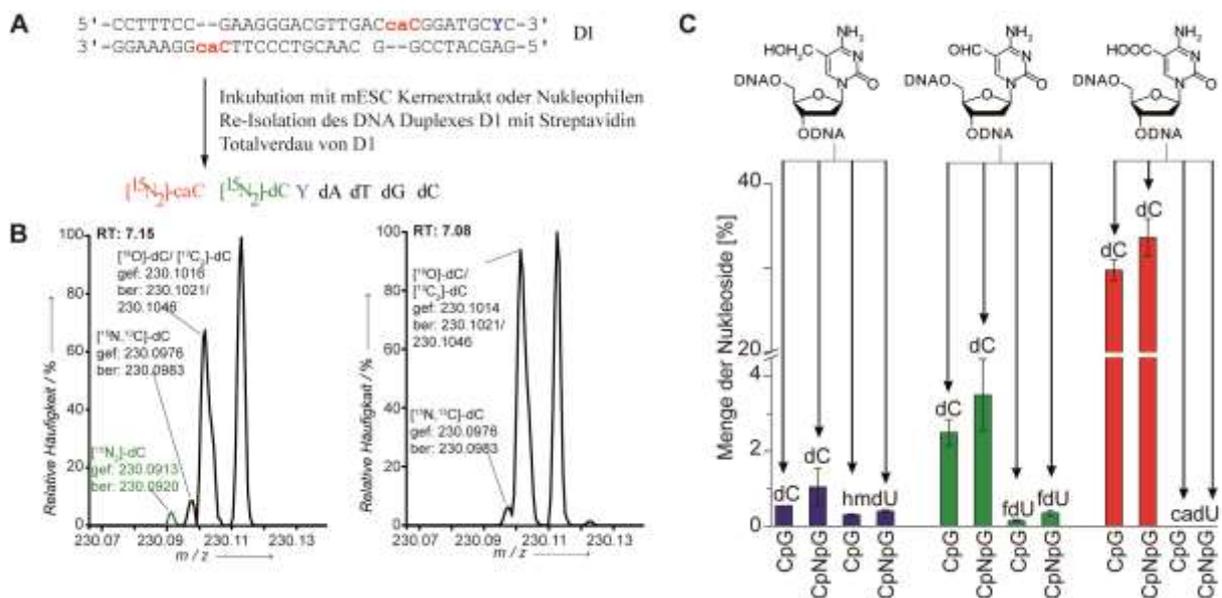
**Abbildung 2:** A) Illustration des oxidativen aktiven Demethylierungsvorgangs. B) Spektrum des in mESC DNA neu entdeckten 5-Formylcytidins.

Diese Entdeckung konnte kurz darauf von einer anderen Gruppe unabhängig bestätigt werden.<sup>[9]</sup>

Darüber hinaus wurde in einem Isotopen-Verfolgungsexperimentes ein Oligonukleotid mit einem stabil markierten Isotopolog des 5-Carboxylcytidins (caC) synthetisiert, mit Kernextrakt von mESCs inkubiert und anschließend durch Affinitätsaufreinigung wiedergewonnen. Dabei konnte markiertes dC massenspektrometrisch und damit auch erste Hinweise auf eine aktive Decarboxylierung nachgewiesen werden (**Abb. 3**).<sup>[10]</sup> Ein in dieser Arbeit synthetisiertes stabil markiertes Isotopolog des dC diente dabei als Standard.

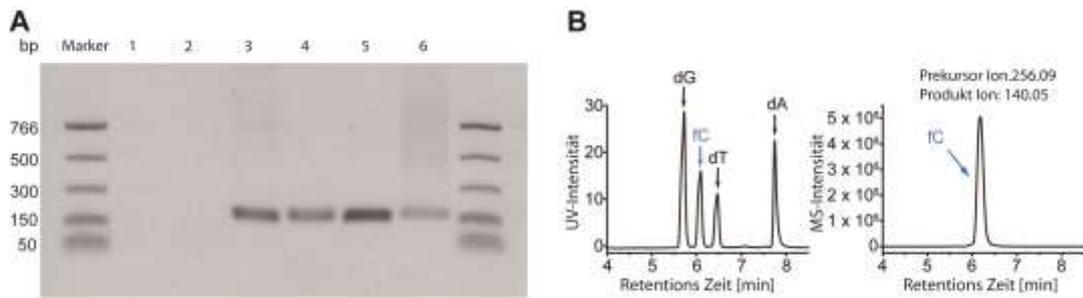
Dieser Standard wurde auch bei weiteren Experimenten genutzt die die Decarboxylierung untermauerten. Hierbei wurden von *Dr. Stefan Schiesser* katalytische Zentren möglicher Enzyme durch die Verwendung von beteiligten Aminosäuren simuliert. Durch hohe Konzentration von Thiolen in Anwesenheit von Katalysatoren konnte eine Decarboxylierung

beobachtet werden. Die exakten Mengen der Decarboxylierung wurden anschließend über eine neue Fragmentierungsmethode an einem Triple-Quadropol LC-MS/MS Gerät, die von *Toni Pfaffeneder* entwickelt und für die im Rahmen dieser Dissertation interne Standards hergestellt wurden, bestimmt. Es konnte gezeigt werden, dass unter den getesteten Bedingungen durch C(5)-C(exo) Bindungsbruch bei caC bis zu 30% dC entstand. Der Bindungsbruch bei fC lag bei nur 1/10 verglichen mit caC und bei hmC war er vernachlässigbar gering.



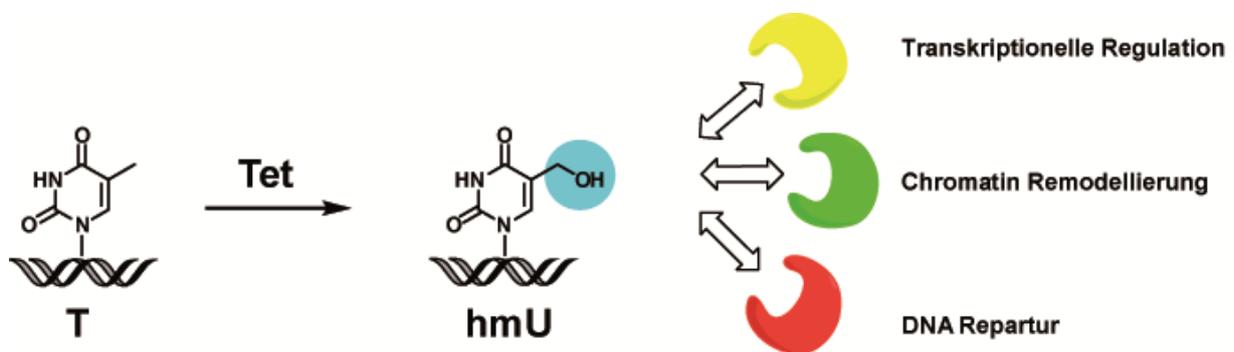
**Abbildung 3:** A) Illustration des Arbeitsschemas zur Verfolgung eines caC markierten Oligonukleotids. Y ist Biotin-desoxy-Uridin, das zur Affinitätsaufreinigung genutzt wurde B) Darstellung des Spektrums von dC nach Inkubation mit Kernextrakt von mESCs (links), oder der Kontrolle (rechts). Stabil markiertes dC in der Anwesenheit von natürlichem dC in der mit Kernisolat behandelten Probe (grün). C) Darstellung der Effizienz des CC Bindungsbruchs der verschiedenen Oxidationsprodukte nach der Inkubation mit einer hohen Thiolkonzentration.

Desweiteren wurde der Einbau von Triphosphaten der neuen Cytidin-Derivate, die von *Dr. Stefan Schiesser und Barbara Steigenberger* synthetisiert wurden, durch diverse Polymerasen etabliert und massenspektrometrisch verifiziert (**Abb. 4**).<sup>[11]</sup> Dadurch stehen nunmehr lange DNA Fragmente mit den neuen epigenetischen Basen zur Untersuchung der verschiedenartigen biologischen Prozesse zur Verfügung.



**Abbildung 4:** A) Darstellung des Gels von PCR Fragmenten nach der Inkorporation der Triphosphate. 1&2: Negativ Kontrolle ohne Templat und ohne dCTP, 3: positiv Kontrolle mit dCTP, 4: hmCTP, 5: fCTP, 6: caCTP. B) LC-MS/MS Analyse von Fragment 5 nach Totalverdau. Klar zu erkennen ist die Detektion von fC sowohl mittels UV-Detektor (links) als auch mittels MS/MS Fragmentierung(rechts).

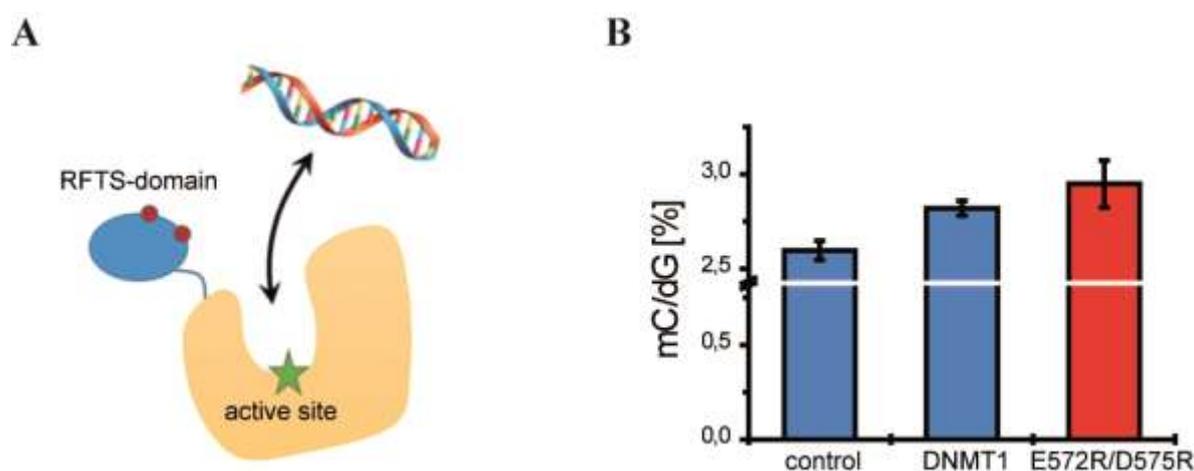
Mit Hilfe der bereits erwähnten LC-MS/MS Methode konnte durch *Toni Pfaffeneder* mit 5-Hydroxymethyluridin (hmU) ein weiteres DNA-Nukleosid als Oxidationsprodukt der TET-Enzyme in Stammzellen entdeckt und quantifiziert werden. Dieser DNA-Baustein wurde im Rahmen der aktiven Demethylierung bisher mit einer Deaminierung von hmC in Verbindung gebracht und galt als ein weiteres mögliches Intermediat, da es von speziellen Glykosylasen erkannt wird. Darüber hinaus konnten einige Proteine indentifiziert werden, die an diese spezielle DNA-Base binden (**Abb. 5**).



**Abbildung 5:** Illustration der Generierung von hmU aus dT durch die TET Proteine und mögliche Beteiligung an diversen biologischen Prozessen.

## Summary:

In this thesis the level of 5-methylcytidine (mC) in cells, that were transfected by *Pavel Bashtrykov* with different DNA-methyltransferase 1 (DNMT1) mutants, was determined using an LC-MS/MS isotope dilution method. It could be shown that the level of this mutant was elevated when the residues 572 and 575 in the protein were mutated from a glutamate and an aspartate to arginine residues. Additional *in vitro* activity studies by the collaboration partner confirmed a higher methylation activity of the DNMT1 mutants. This confirmed a hypothesis of self-regulation.<sup>[1]</sup> In this theory the protein is regulating its own activity by binding of an N-terminal domain via ionic contacts to the catalytic domain at the C-terminus (**Fig. 1**).<sup>[2]</sup>

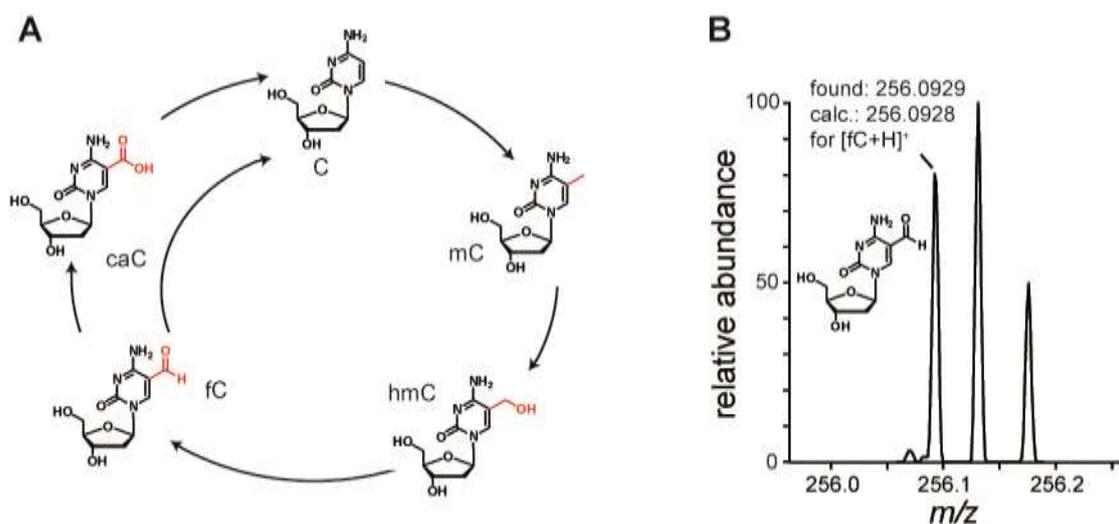


**Figure 1:** A) Schematic view of the allosteric self-regulation of the N-terminal domain (blue) and the mutation sites (red dots). B) Illustration of the 5-mC content in the genomic DNA of the control, the wildtype (blue bars) and the mutant enzyme (red). Slightly enhanced methylation in the mutant version compared to the wild type enzyme.

Since 2009 the oxidation of mC by the ten eleven translocation (TET) proteins to 5-hydroxymethylcytidine (hmC) is known, leading to the hypothesis that an oxidative pathway for active demethylation exist.<sup>[4-5]</sup> Using a stable isotope dilution LC-MS method, developed by *Daniel Globisch*, the content of this modification in the tissue of brain tumors was determined.<sup>[6]</sup> It was shown that hmC was strongly reduced in the investigated tumor types. This result might allow using hmC-levels as tumor markers.

In another part of this thesis another cytidine modification, 5-formylcytidine (fC), was discovered, together with *Toni Pfaffeneder*, in the genomic DNA of mouse embryonic stem cells (mESCs) using highly sophisticated mass spectrometry (**Fig. 2**).<sup>[7]</sup> It is proposed that it is

a further intermediate in the active demethylation cycle and that this modification is also generated from hmC by the TET enzymes. Furthermore a method for chemical modification of fC was developed, that was later used by another group to determine the genomic distribution of this modification.<sup>[8]</sup>



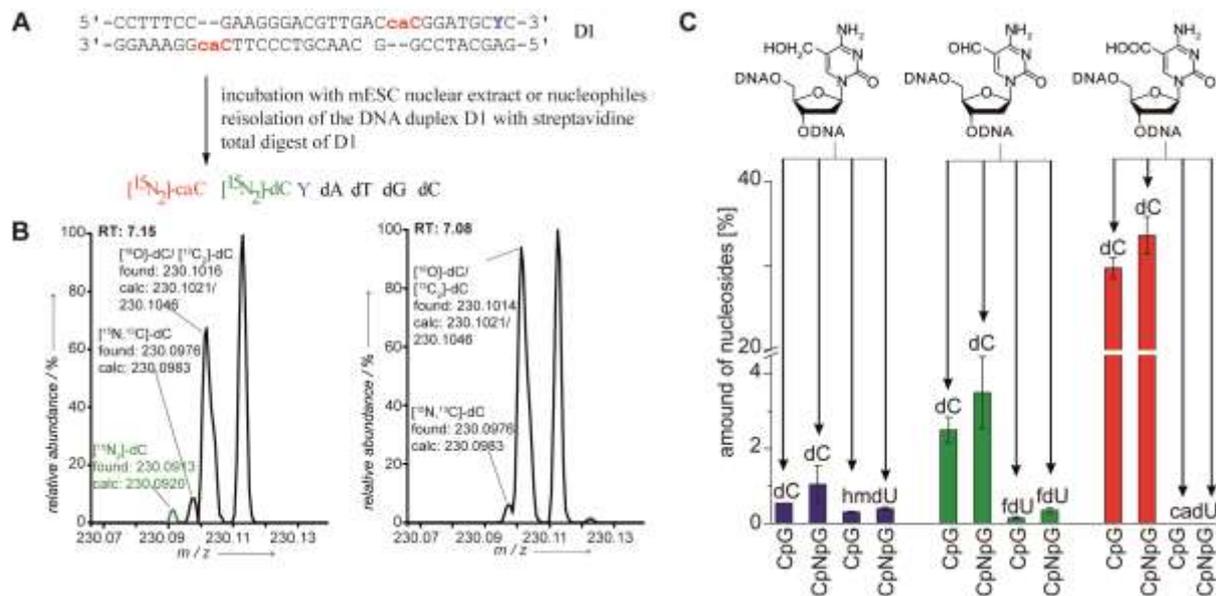
**Figure 2:** A) Illustration of the active demethylation cycle. B) Spectrum of the newly discovered 5-formylcytidine in mESCs.

The discovery of fC in genomic DNA of mESC was confirmed later on by another group. They could also identify the last proposed intermediate of the active demethylation cycle, namely 5-carboxylcytidine, as a product of the TET enzymes in mESCs.<sup>[9]</sup>

Moreover an isotope tracing experiment with an isotopically labeled 5-carboxycytidine (caC) in an oligonucleotide was performed in this thesis. After incubation with nuclear extracts of mESCs the oligonucleotide was re-isolated using affinity purification and enzymatically digested to the nucleoside level. The isotopically labeled dC was discovered, which provided evidence for the existence of a decarboxylase activity (**Fig. 3**).<sup>[10]</sup> A labeled nucleoside standard of dC was synthesized and used as a standard for the MS experiments.

Additionally this standard was used by *Stefan Schiesser* to perform chemical decarboxylation experiments to validate the earlier finding. Active sites of possible decarboxylases were mimicked by using different amino acid combinations. High concentrations of thiols in the presence of catalysts lead to the same signal of labeled dC.

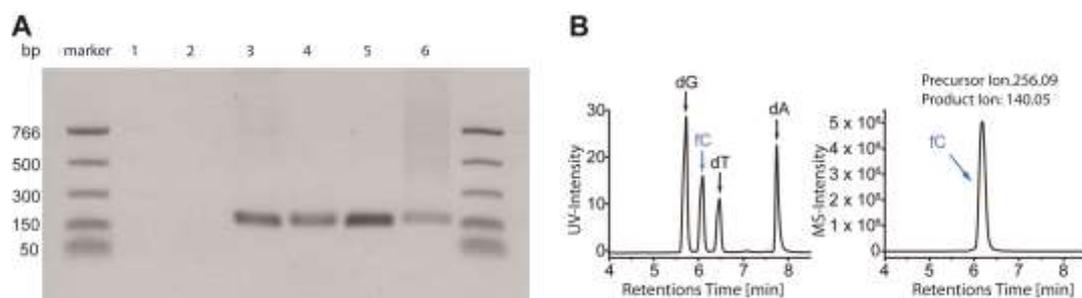
The exact amount of the chemical decarboxylation was then determined using a new low resolution fragmentation method on a triple-quadruple LC-MS/MS machine. This method was



**Figure 3:** A) Illustration of the workflow for the tracing of the oligonucleotide carrying the marked caC. Y is biotin-desoxyuridine, which was used for the affinity purification. B) Illustration of the spectrum of dC after the incubation with nuclear extracts of mESCs (left), or of dC in the control experiment (right). The traced stable isotope of dC is marked in green. C) Illustration of the efficiency of the C-C bond cleavage of the different oxidized mC derivatives after incubation with high concentrations of a thiol.

developed by *Toni Pfaffeneder* and used several standards, that were synthesized or purified in this thesis. It could be shown that the C(5)-C(exo) bond cleavage of caC was the most efficient reaction under the tested conditions, it resulted in up to 30% dC formation.

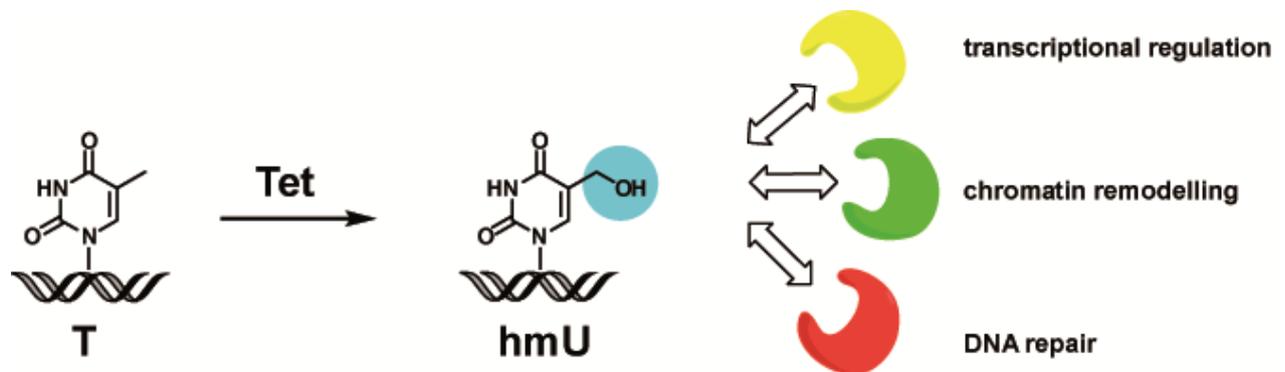
Furthermore the incorporation of the triphosphates of the new cytidine derivatives using polymerase chain reaction was established and confirmed by LC-MS/MS (**Fig.4**).<sup>[11]</sup>



**Figure 4:** A) Illustration of the agarose gel containing the PCR fragments after incorporation of the newly synthesized triphosphates. 1&2: negative controls without template or without dCTP, 3: positive control with dCTP, 4: hmCTP, 5: fCTP, 6: caCTP. B) LC-MS/MS analysis of fragment 5 after total digest. The incorporated fC is clearly detectable using UV light (left) or MS/MS fragmentation (right).

The triphosphates were synthesized by *Stefan Schiesser* and *Barbara Steigenberger*. Now long DNA fragments containing the new epigenetic bases are available for investigations of epigenetic processes.

Using the earlier mentioned LC-MS/MS method, *Toni Pfaffeneder* identified 5-hydroxymethyluridine (hmU) as a product of TET-enzymes in the DNA of mESCs (**Fig. 5**). This nucleotide was connected with the deamination of hmC and was suggested to be another intermediate of active demethylation, as it is recognized by special glycosylases. Furthermore specific readers for this novel DNA-base were identified.



**Figure 5: Illustration of the generation of hmU from dT by TET proteins and of the biological processes it might be involved in.**

# I. Introduction

## 1. Pluripotent cells in modern medicine

Understanding cellular differentiation on the molecular level is one of the main challenges of the 21<sup>st</sup> century. The life of a vertebrate organism starts with the fertilization of an egg by a sperm cell resulting in a zygote.<sup>[12]</sup> Out of this zygote, the whole multicellular organism develops by cell division and cell differentiation. The ability of these cells to found a whole organism is called *totipotency*.<sup>[13]</sup> After a few cell divisions, the blastocyst is formed. It consists of the inner cell mass and the outer trophoblast. The cells of the inner cell mass begin to differentiate and are no longer able to form an individual on their own anymore, however, they can still differentiate into any of the three germ layers (endo-, meso-, ectoderm)-this state is called *pluripotent*.<sup>[14]</sup> In 1981 the first cultured embryonic stem (ES) cells could be isolated from the inner cell mass of a mouse embryo and expanded *in vitro*.<sup>[15-16]</sup>

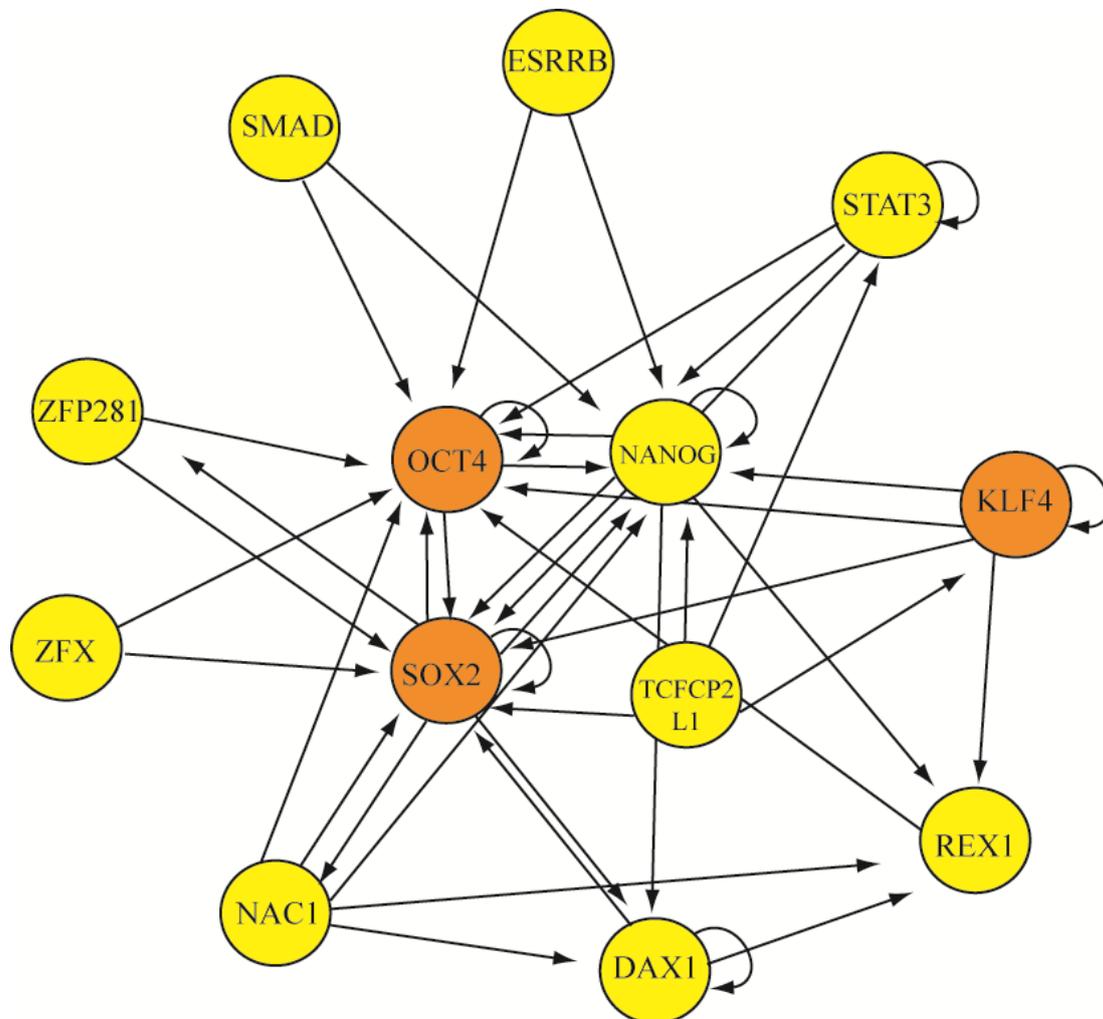
*“Pluripotential embryonic cells appear to give rise to both rapidly differentiating cells and others which like themselves, remain undifferentiated.” – LeRoy Stevans and Clarence Little.<sup>[17]</sup>*

In 1954 this definition of an ES cell was stated by *LeRoy Stevans* and *Clarence Little* and is valid until today. Because of their ability to develop to any kind of cell type and tissue, these cells are of interest for modern medicine. They promise to replace non-healthy tissue by a new tissue generated from stem cells. However stem cell therapy has a number of drawbacks still to overcome, as it is difficult to control the differentiation of stem cells to a certain tissue and many differentiation pathways are to date unknown. The major problems are the availability of stem cells due to the ethical dilemma of the death of an embryo when isolating ES cells and the immune response to foreign cell material. To avoid this it has been tried to gain pluripotent cells from other sources. In 2012 the Nobel Prize was awarded to *John B. Gurdon* and *Shinya Yamanaka*, two pioneers in the area of nuclear transfer and induced *pluripotent* stem cells (iPSC). These cells derive from somatic cells that are reprogrammed by several factors to the *pluripotent* state. These so called Yamanaka factors, *Oct4*, *Sox2*, *Klf4* and *c-Myc*, are part of a complex regulatory network, the so called pluripotency network that is unique to pluripotent cells.<sup>[18-19]</sup> A core network of more than a dozen transcription factors, including three of the Yamanaka factors, which tightly regulate each other's expression is

responsible for the pluripotent state of the cell (**Fig. 6**). Other methods for reprogramming have been shown, most recently a still controversially debated method that induced reprogramming by external stimuli, such as low pH.<sup>[20-22]</sup>

These iPSCs are very similar to ES cells, as they are able to divide indefinitely and are able to differentiate, however they have some disadvantages such as the low efficiency of generation.<sup>[23]</sup>

Understanding the reprogramming processes and the unique genomic properties of *pluripotent* cells on the molecular level, has the potential to contribute to modern medicine. A current example of stem cell therapy is the stand alone growth of live murine teeth.<sup>[24]</sup> This technique has in principle the capability to replace prostheses and give rise to new teeth. *Volponi et al.* have summarized the ongoing research in a review.<sup>[25]</sup>



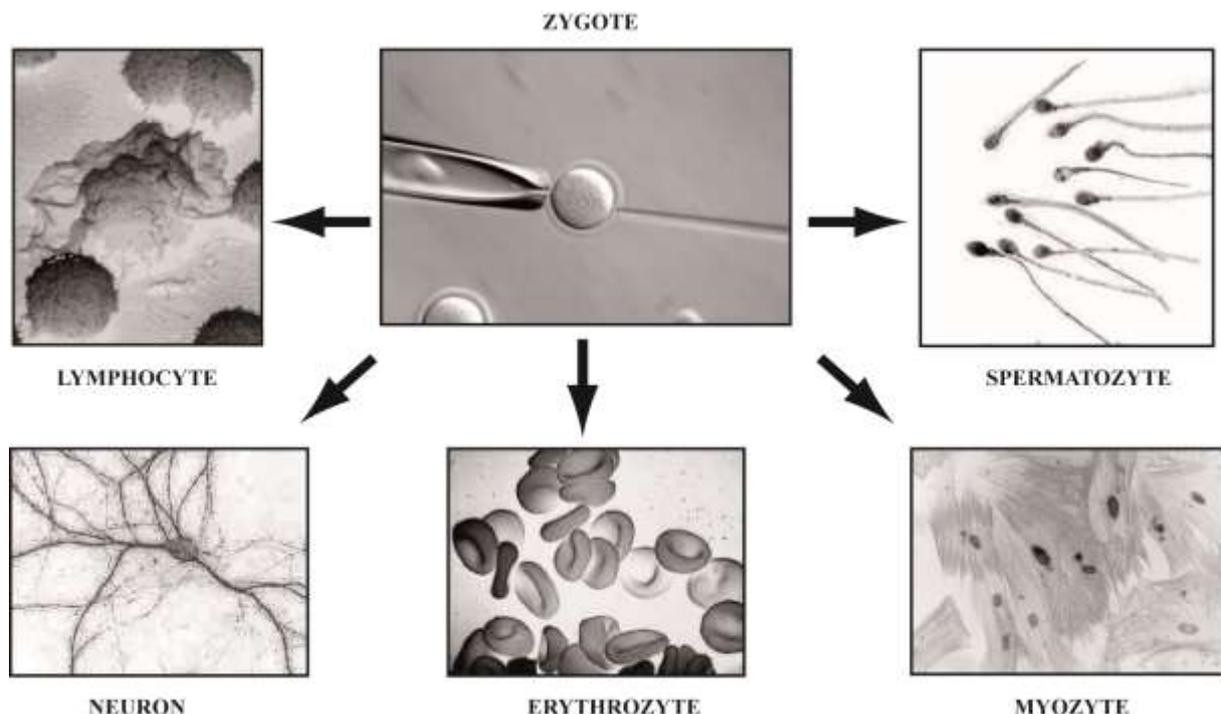
**Figure 6:** Illustration of the core transcription factor network for pluripotency. Yamanaka factors are in orange. C-myc is not a part of the core pluripotency network.

## 2. Epigenetics

In order to understand the process of reprogramming, it is important to understand the reversible process of cell differentiation. As described a neuron and a lymphocyte derive from the very same *totipotent* cell (**Fig. 7**). The diversity of cell types is thus dependent on regulations beyond the genetic code. These regulations are taken together in the scientific field of epi-genetics (greek: *epi- beyond*).<sup>[26]</sup>

Necessary for the diversity is a “switch” on of the transcription of cell specific genes and hence different expression patterns. These genes include the pluripotency network, which is downregulated upon environmental signals and furthermore tissue type specific genes that are upregulated. Consequently the cell is epigenetically programmed to differentiate. Not only tissue cells differ in their epigenetic status. Comparing ES cells cultured with different signal pathway inhibitors, *Nichols* and *Smith* established in 2009 the concept of “naïve” and “primed” cells.<sup>[27]</sup> They differ on the molecular level in their epigenetic status and are therefore developmentally and functionally different.

The molecular level of these regulations is established by specific promoters and enhancers on the DNA-level, together with a set of special transcription factors on the protein-level. These levels are so called *cis*- and *trans*-elements. In the environment of our genetic code, the chromatin, these elements control gene expression.



**Figure 7: Possible differentiation of stem cells to distinct cell types.**

The genetic information is converted into messenger RNA (mRNA) and ultimately translated into proteins. The different repertoire of proteins at a certain location and time point, the *proteome*, is controlled by several processes such as proteosomal degradation, however only epigenetic marks can, just like the DNA sequence, be stably inherited.<sup>[28]</sup> Epigenetic mechanisms play a major role in the chromatin environment and work, in part, via its physical properties. One of two major epigenetic mechanisms is the dynamic regulation of the nucleosome structure.

## 2.1 Histone modifications

DNA is wrapped around histones, forming the nucleosomes, to gain the extreme compaction necessary to store the huge genetic information of our genome into the nucleus. There exist four core-histones, H2A, H2B, H3 and H4 as well as a huge number of special histone variants.<sup>[29-30]</sup>

The Nucleosomes are built up out of one H3/H4 tetrameric complex and two H2A/H2B dimers forming a histone octamer.<sup>[31]</sup> These octamers have several positive charges on their outside and therefore attract DNA. They act as spools on which the DNA is wrapped around and therefore stored in a very compact form. At certain amino acid side chains in their long N-terminal ends, the so called tails, histones can be covalently modified. As these modifications have been investigated for several decades we have been able to gain a broad knowledge about them. More than 100 different modifications have been described, which range from well-known ones such as lysine acetylation, lysine methylation and serine/threonine phosphorylation to less understood ones such as crotonylation or poly-ADP ribosylation.<sup>[32-33]</sup>

In 1961 histone acetylation has been discovered and investigated ever since.<sup>[34]</sup> Lysine residues can be acetylated by histone-acetyltransferases (HATs) using acetyl-CoA and deacetylated by histone-deacetylases (HDACs).<sup>[35-36]</sup> HATs, HDACs and histone hyperacetylation have been reported early to associate with high transcriptional activity.<sup>[37]</sup> *In vitro* and *in vivo* studies have supported the idea that neutralization of the positive charge of lysine residues by acetylation leads to weakened DNA binding.<sup>[38-39]</sup> Thus promoter- and enhancer-DNA is physically better available for the transcription machinery and other biological processes such as DNA repair. The half-life of acetylation marks is in the range of a few minutes and thus acetylation and deacetylation is a highly dynamic process, to ensure quick recovery of a more densely packed chromosome.<sup>[40]</sup>

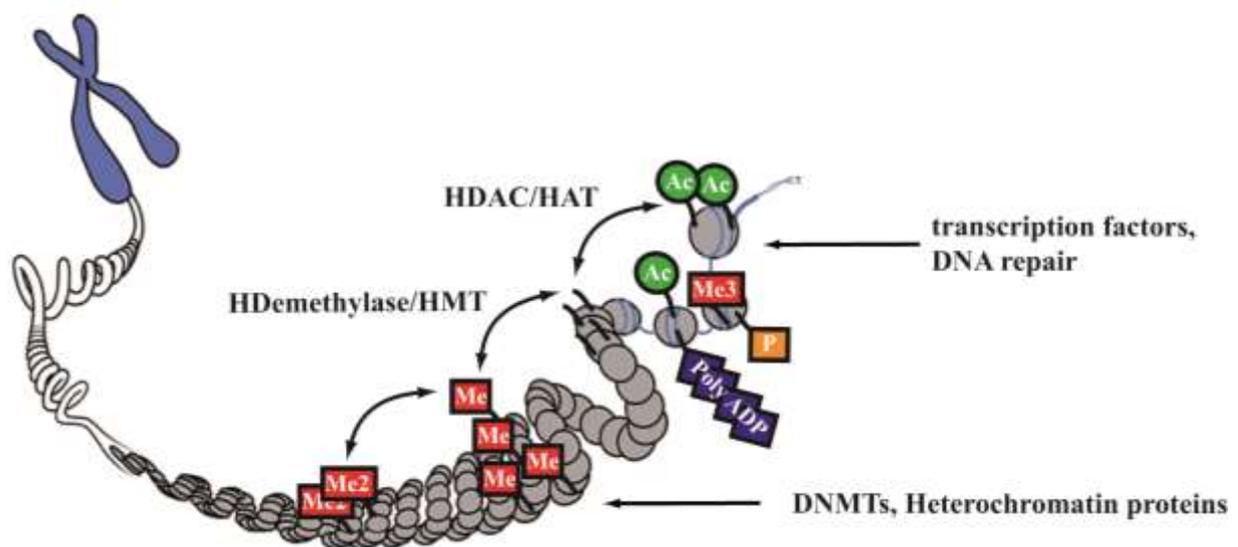
Histones can also be phosphorylated by distinct kinases or poly-ADP ribosylated by poly-ADP ribosylation polymerase (PARPs) on serine and threonine side chains, leading to a negative charge on histone tails and therefore presumably to charge repulsion with the DNA backbone.<sup>[41-42]</sup> Similar to acetylation, the resulting looser packed DNA has enhanced availability. The major processes associated with poly-ADP ribosylation are DNA repair and heat shock response, whereas phosphorylation is mainly involved in cellular development and regulation of other histone modifications.<sup>[43-45]</sup>

Another well investigated modification is mono-, di- and tri-methylation of lysines.<sup>[46-48]</sup> It is carried out by histone methyltransferases (HMTs) using *S*-adenosyl methionine (SAM) as cofactor.<sup>[46,49-50]</sup> About 60 different HMTs have been annotated in the human genome and most of them harbor a Su(var)3-9, Enhancer-of-zeste, Trithorax (SET) domain, the catalytic active domain.<sup>[51]</sup> The exposure time of lysine residues to SET domains is important for the methylation state and is regulated by chromatin remodeling and other histone modifiers.<sup>[52]</sup> It has been suggested that methylation has indirect effects on nucleosome dynamics by interaction with different partners, rather than a general direct mechanistic effect.<sup>[53]</sup> Some of the recognizing elements of the three lysine methylation states are the Chromo-, PWWP-, plant homeo- (PHD) and Tudor-domains frequently found in *cis*-regulatory elements and chromatin associated factors.<sup>[54-55]</sup>

Methylation of different lysine residues therefore has different effects on transcriptional regulation. A mark generally associated with active transcription is H3K4 trimethylation, which is found at transcription starting sites (TSS) in eukaryotes. Its presence correlates positively with active transcription.<sup>[56]</sup> However there are also bivalent marks such as H3K27 which depending on its methylation status and surrounding histone marks can either activate transcription or repress it. Mono- and di-methylation marks on H3K27 are only observed in combination with phosphorylation of H3S28 and are associated with active transcription, whereas trimethylated H3K27 is only observed in the absence of H3S28P and associated with repression.<sup>[56-57]</sup>

In the younger history of histone research, conserved histone demethylases have been identified in yeast and humans.<sup>[58-59]</sup> They contain a jumonji C (JmiC) domain that is able to oxidize the methyl groups using Fe<sup>II</sup> and 2-oxoglutarate (2-OG) to hydroxymethyl-lysines. The oxidized methylgroup is then released as formaldehyde. Demethylases are reported to act strictly dependent on the methylation status, as JHDM1A is only able to demethylate di-methylated H3K9 whereas JHDM3A is able to demethylate tri-methylated H3K9.<sup>[59-60]</sup>

The heavy intra- and inter-crosstalk of histone modifications and modifiers indicates the tight regulation and strong dynamics of these processes (**Fig. 8**). Their improper activity can lead to several diseases, especially to cancer, as the expression of several oncogenes can be altered.<sup>[61]</sup> Correct dynamic histone modifications are also important for the developmental status and pluripotency of cells. Bivalent marked genes, so called “poised” genes are associated with pluripotent cells and are necessary for their rapid differentiation.<sup>[62]</sup> Another very important crosstalk of histone modifications for pluripotency, is the connection of histone modifications and modifiers to the second major mechanism of epigenetics, DNA methylation.<sup>[63-64]</sup>

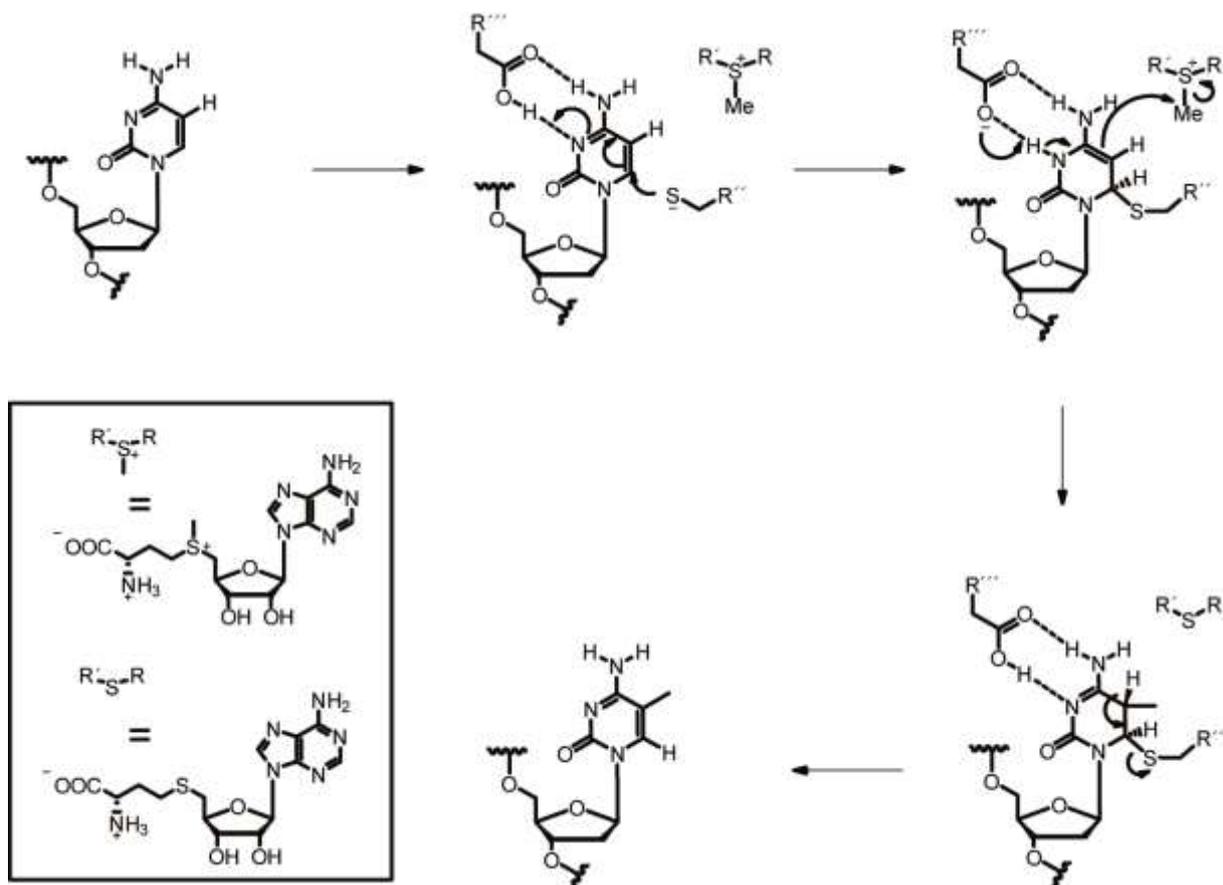


**Figure 8: Illustration of chromosome arrangement and dynamic regulation of histone modifications. Recognizing elements of different biological processes are indicated with arrows.**

## 2.2 DNA methylation

Histone modifiers and modifications can recruit DNA-Methyltransferases (DNMTs) to specific gene loci.<sup>[65]</sup> DNA methylation in eukaryotes was first discovered in 1948 by *Hotchkiss* and is involved in several cellular events, such as cellular development, gene expression, chromatin remodeling and regulation of parental imprinting as well as stabilization of X chromosome inactivation.<sup>[66-71]</sup> DNA methylation occurs at several DNA-sites, however the most important and abundant site is the so called CpG site. Of CpG sites about 60-80% are methylated, which corresponds to about 4-6% of all cytosines.<sup>[72]</sup> If the observed frequency of a CpG site on 200 base pair length is higher than statistically expected, these regions are defined as CpG islands (CGI). The CGIs occur mainly in promoter

regions and can be found in about 70% of all human genes.<sup>[73-74]</sup> CGIs are classified in high-, medium- and low-density CGIs, depending on the frequency of CpG sites. In promoters high density CGIs are usually unmethylated, whereas especially low density CGIs seem to be more frequently methylated.<sup>[75-76]</sup> The CG dinucleotides are in comparison to the other dinucleotides statistically underrepresented, because deamination of 5-methyldeoxycytidine (mC) to deoxythymidine (T) is about 4 times faster than deamination of C to T and its repair is not as efficient as for other mutations.<sup>[77-78]</sup> Therefore during evolution, it is thought that many CG sites were lost.<sup>[72]</sup> Several enzymes can bind methylated cytosine via a methyl-CpG binding domain (MBD) because the methyl group is pointing into the major groove of the DNA.<sup>[79]</sup> These proteins can either make promoters unavailable for transcription factors or directly recruit inhibitors of transcription.<sup>[80]</sup>



**Figure 9: Conserved mechanisms of DNA methyltransferases: Briefly an attack of a cysteine residue on the C(6) of cytosine leads to formation of a covalent enzyme-substrate intermediate. After attack of the formed enamine on the methyl group of S-adenosyl-methionine (left box; up), *syn*-elimination leads to the product 5-methyldeoxycytidine and S-adenosyl-homocysteine (left box; bottom)**

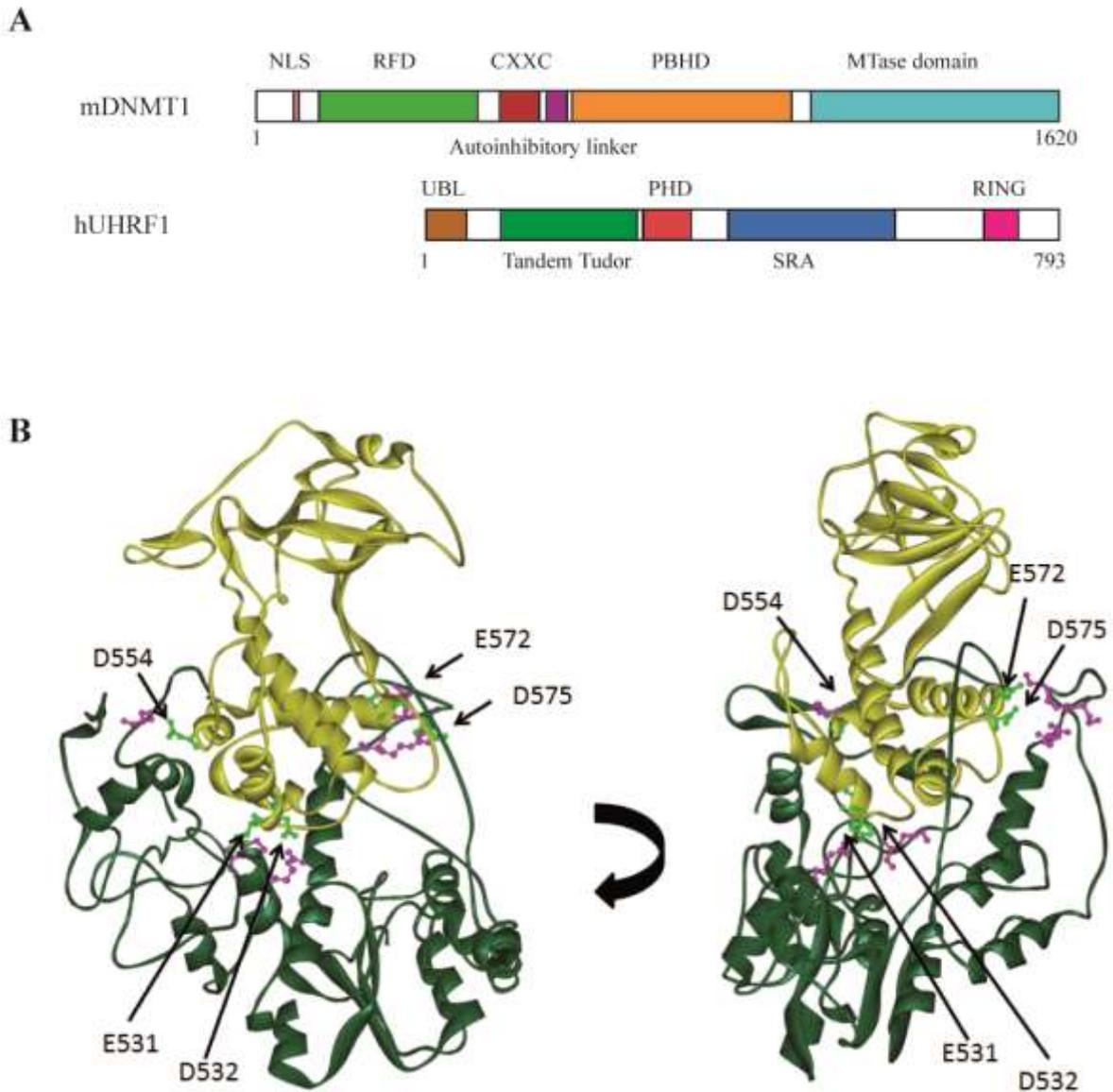
Therefore mC is strongly correlated with transcriptional repression. As described DNMTs can be recruited by histone modifiers, however, also the inverse recruitment is known and leads to

chromatin remodeling.<sup>[81]</sup> Aberrant mC levels can lead to serious diseases, the best investigated of them being cancer. Wrong methylation patterns can, just like aberrant histone modifications, lead to upregulation of oncogenes and downregulation of tumoursuppressors, which may lead to unregulated cell growth and tumorigenesis.<sup>[82]</sup> Methylation patterns are also important for cell differentiation as each cell type is methylated at specific sites.<sup>[83]</sup> Methylation is carried out by DNMTs using a highly conserved mechanism. The catalysis is a nucleophilic attack of a cysteine residue in the active site at the C(6) position of cytosine, after flipping the target base into the active site.<sup>[84]</sup> This leads to a covalently bound DNA-protein intermediate and a double bond at the C(4)-C(5) position forming an enamine. This enamine attacks the methyl group of the co-substrate SAM in an S<sub>N</sub>-type reaction. After *syn*-elimination of the cysteine residue 5-methylcytosine (mC) is released (**Fig. 9**).

### 2.2.1 DNA Methyltransferases

There are three types of mammalian methyltransferases sharing a similar structure and catalyze the same reaction, the maintenance methyltransferase DNMT1 which is responsible of copying methylation marks to newly synthesized DNA together with its interaction partner ubiquitin-like plant homeodomain and RING finger domain (UHRF1) and two *de novo* methyltransferases DNMT3a and DNMT3b.

Mouse DNMT1 comprises 1620 amino acids and its N-terminal region harbours several interaction and regulation domains.<sup>[72]</sup> It contains the nuclear localisation signal (NLS) and the CXXC zinc finger domain, which is a common DNA binding motif.<sup>[85]</sup> Also a sequence to target the enzyme to replication foci (RFD) and a PCNA binding domain (PBD), both guiding the enzyme to the newly synthesized DNA substrates at the replication fork, are located there.<sup>[86-88]</sup> To achieve localisation there, UHRF1 interacts with the RFD domain and acts as a molecular chaperone of DNMT1, guiding it to CpGs at the replication fork by binding hemimethylated DNA over its SET and RING finger-associated (SAR) domain (**Fig. 10**).<sup>[89-90]</sup> It can discriminate between fully methylated DNA and hemimethylated DNA by steric repulsion and flips out mC into its active site, where specific contacts are made.<sup>[91-92]</sup> At the hemimethylated sites DNMT1 has a high processivity and methylates long stretches of DNA without dissociating or swapping the DNA strand.<sup>[93]</sup> The polybromo homology domain of DNMT1, whose function is to date unknown, as well as an autoinhibitory domain, that has recently been discovered and shown to prevent *de novo* methylation, are harboured in the N-terminus.<sup>[94]</sup> The C-terminal region contains the MTase domain and is separated from the N-terminal region by GK repeats (**Fig. 10**).



**Figure 10:** A) Schematic view of the structure of mouse DNMT1 and human UHRF1. B) Crystal structure of mDNMT1. The binding of the RFTS domain (yellow) to the DNA binding pocket of DNMT1 that is located in catalytic part (dark green) is stabilized by several hydrogen bonds and ionic contacts (indicated by arrows) and leads to intrinsic suppression. Important contacts of the RFTS domain are labeled (light green). Residues of the catalytic domain and the CXXC domain involved in interaction with the RFTS domain are shown in pink. These images have been prepared using the crystal structure (PDB code: 3AV4) published by *Takeshita et al.* (2011) and was taken from *Bashtrykov et al.*<sup>[1]</sup>

UHRF1 comprises of 793 amino acids and can enhance DNMT1's activity by about 5-fold.<sup>[95]</sup> Additionally, UHRF1 connects DNA methylation directly to histone modifications, as its PHD and Tandem Tudor domain recognize H3K9-trimethylation.<sup>[95-96]</sup> Other binding partners are histone deacetylase 1, histone acetyltransferase and PCNA.<sup>[90,97-98]</sup> Due to its interaction

partners, it is therefore proposed that it is part of a multiple enzyme complex that is responsible for the epigenetic code replication machinery.<sup>[99]</sup>

It is the most important regulator of DNMT1 and is also involved in proteasomal degradation of DNMT1 by acting as E3 Ligase.<sup>[100]</sup> In UHRF1 double knock-out mES cells and mouse embryos the global methylation drops as much as in a DNMT1 double knock-outs.<sup>[90,101]</sup> Thus as it stands UHRF1 and DNMT1 together are responsible for the correct inheritance of methylation patterns.

While the C-terminal catalytic part of DNMT3a/b are very homologous, the regulatory N-terminal part contains different domains. A cysteine-rich plant homeo domain (PHD), also known as ATRX-DNMT3-DNMT3L (ADD) domain, binds zinc and is known to take part in various protein-protein interactions, such as with the nucleosome associated proteins myc, SETDB1, BRg1 and HDAC1 (**Fig. 11**).<sup>[65,102-104]</sup>

A conserved proline-tryptophane motif (PWWP domain) is also located at the DNMT3a/b N-terminus. It is part of the royal domain super-family that is known to bind to histone tails with various marks.<sup>[55]</sup> These interactions lead to pericentromeric chromatin localisation and enhanced activity.<sup>[105-106]</sup>

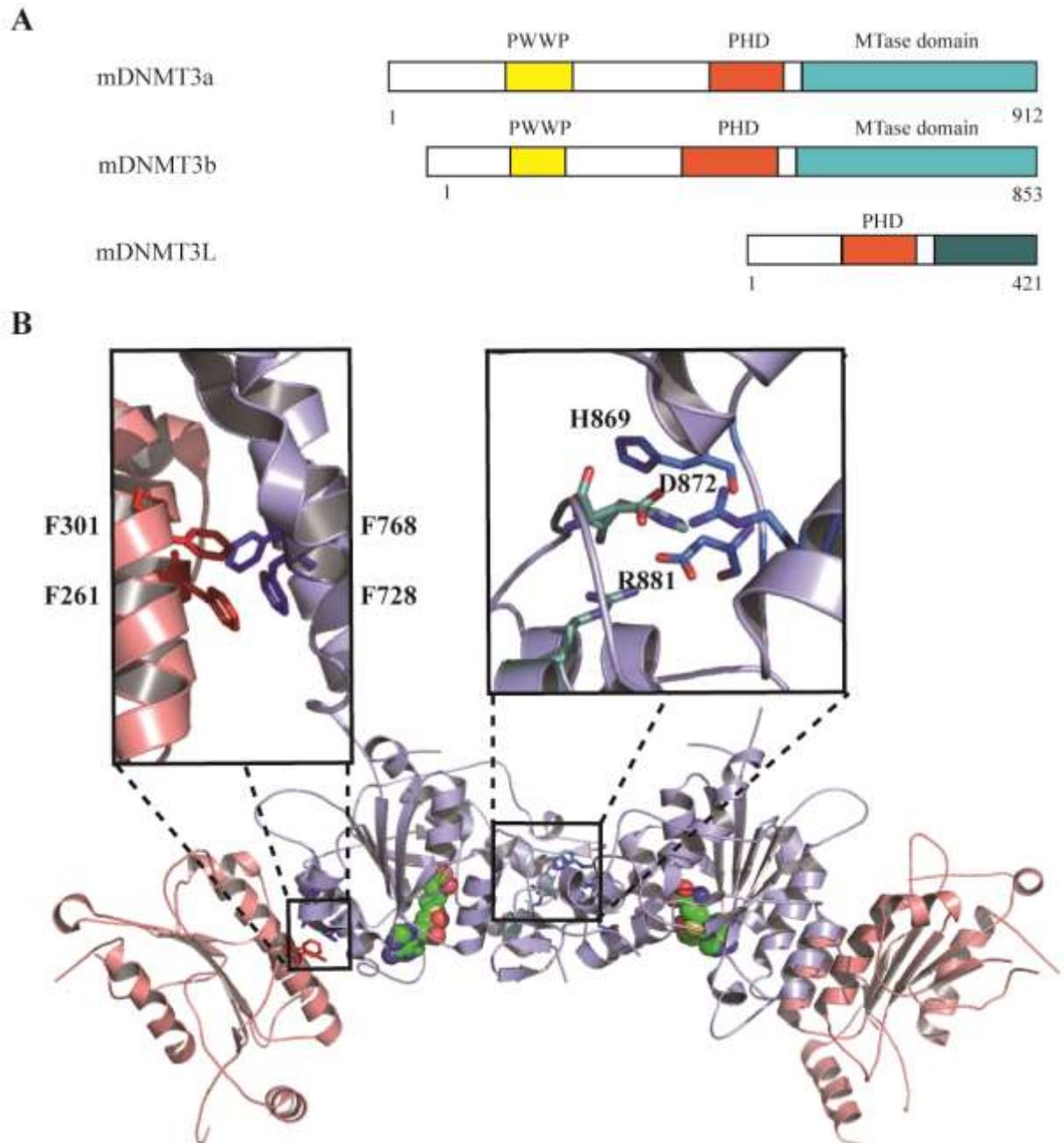
DNMT3L, another member of the DNMT-family in mammals, is catalytically not active as the C-terminal part lacks several motifs of the methyltransferase domain and cannot bind the cofactor SAM. The N-terminal part just like DNMT3a/b contains the PHD domain which is necessary for the interaction of the DNMT3s.<sup>[107-108]</sup> Via these interactions it plays despite its inactivity an important role in de novo methylation, which is described later in this chapter.

Genetic studies on DNMT knock-out mice prove their significant role in cellular development, as all DNMT knock-outs are lethal and mutations in these genes lead to severe diseases. But they also show the differences in the contribution to DNA methylation.

DNMT1 knock-out mice die shortly after gastrulation, show loss of genomic imprinting and X-chromosome inactivation.<sup>[109-111]</sup> Dnmt3a knock-out mice develop to term, but die shortly after birth, while DNMT3b deficient mice die *in utero* at embryonic day E9.5. These phenotypes indicate that the enzymes despite their structural and biochemical similarities cannot complement each other.<sup>[112]</sup> The indispensable role of DNMTs in mammalian development makes it necessary to control their activity and location tightly.

Maintenance methyltransferase DNMT1, that is highly expressed in dividing cells, has a preference to bind to double stranded hemi-methylated CG-sites in a CG rich context, however it also binds to poly-(dA), -(dT), -(dG), -(dG/dC) DNA, poly (ADP-ribose), fully methylated double stranded DNA (dsDNA) as well as methylated single stranded DNA

(ssDNA).<sup>[113-118]</sup> However binding does not correlate with the kinetics of methylation, as most of the binding partners show slow methylation kinetics or are even unsuitable as substrates.<sup>[119]</sup>



**Figure 11:** A) Schematic view of DNMT3 proteins. B) Crystal structure of the DNMT3a/L tetrameric complex. Zoom into the interaction sites of DNMT3a (purple) with another DNMT3a molecule over polar contacts of aminoacid residues H869, H872 and R881. Interaction site of DNMT3a with DNMT3L (both salmon) is comprised of residues F768, F728 of DNMT3a and F301 and F261 of DNMT3L. SAH in spheric illustration. PDB code: 2QRV.<sup>[120]</sup>

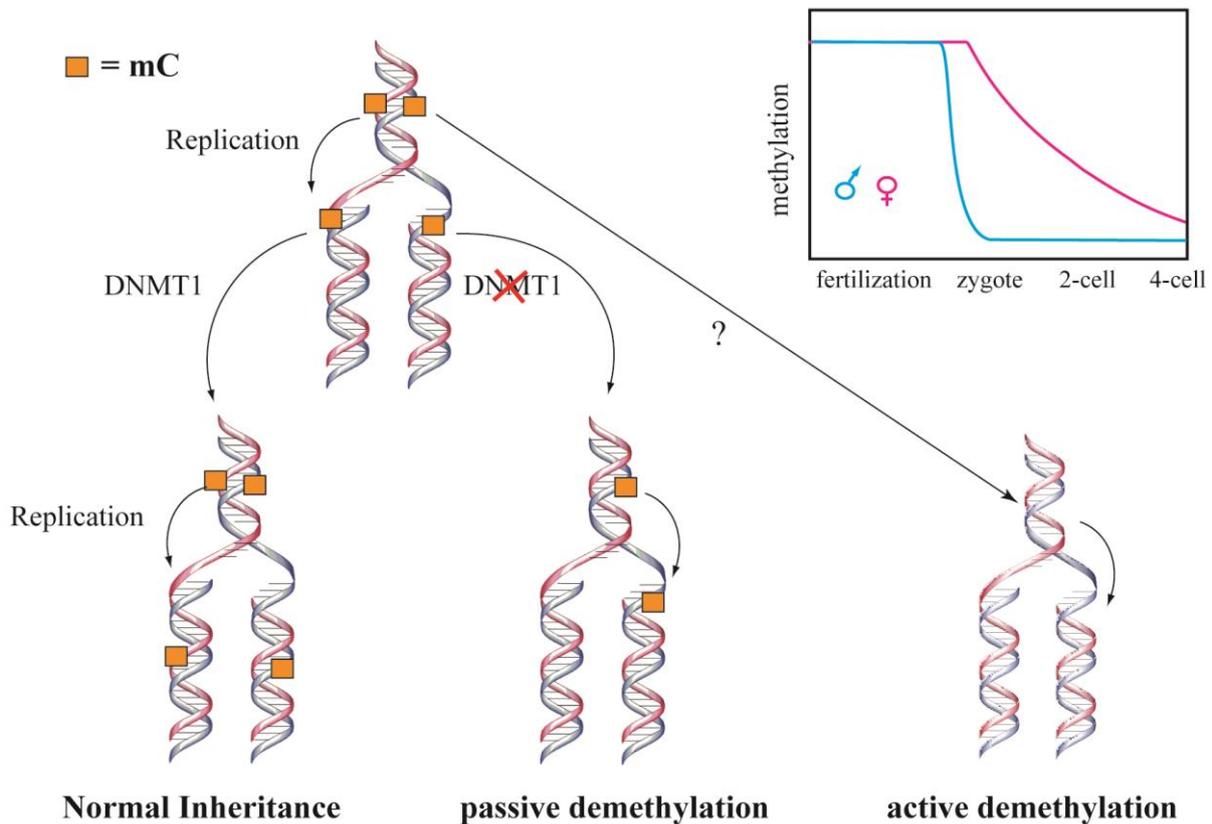
through a network of hydrogen bonds and form the DNA binding site (**Fig. 11**).<sup>[139]</sup> Interaction over hydrophobic contacts with DNMT3L seems to introduce a conformational change that increases binding to SAM, and hence its activity.

Additionally, the complexity of the regulation of DNA-methylation is greatly enhanced by post-translational modifications, such as sumoylation or phosphorylation, by the existence of different isoforms of DNMTs and distinct localisation during the cell cycle and embryonic development.<sup>[72]</sup>

DNA methylation, however, is not only tightly controlled, as evidence occurs that it is reversible too, and thus like histone modification a dynamic process during mammalian development.

## 2.3 DNA Demethylation

As the transcriptional requirements upon cell specification change very fast, a dynamic and flexible epigenome is necessary. In contrast to that DNA methylation was thought to be a stable epigenetic mark and the only possibility to remove it was by diluting it during DNA-replication (**Fig. 12**), recent studies suggest a more dynamic picture. The loss of methylation marks has been observed genome-wide and locus specific upon different stimuli. Comparing the methylation levels during embryonic development they appear strongly altered. Only 4-8 hours after fertilization the mC-levels of the paternal pronucleus drop rapidly.<sup>[140-141]</sup> As the first cell division is not thought to take place at that time and this change was shown to happen in the presence of a replication inhibitor, a passive demethylation by impaired DNMT1 activity is unlikely.<sup>[141]</sup>



**Figure 12:** Schematic view of the mC levels during early cell development (box); Overview of possible proceedings with methylation marks: maintenance of methylation (orange mark) after replication on the left site, passive demethylation after replication and inhibited DNMT1 activity in the middle and active removal of mC on the right.<sup>[142]</sup>

Another wave of active genome-wide demethylation occurs in primordial germ cells. Here almost all methylation patterns including the one of imprinted genes are removed.<sup>[143]</sup>

Additionally locus specific demethylation has been reported in non-dividing neurons, in T-lymphocytes and during hormone-regulated gene activation.<sup>[144-146]</sup>

It is hard to imagine that one single mechanism is responsible for these very different demethylation events. Therefore a number of pathways and mechanisms have been proposed and investigated.

### 2.3.1 Base excision repair of mC

In plants base excision repair (BER) of mC is thought to be the major pathway of active demethylation and it is carried out by the family of Demeter (Dme) nucleases.<sup>[147]</sup>

BER has two mechanisms:

1. The short patch mechanism of BER starts with the recognition of the damaged base. It is then flipped out of the DNA into the active site of the enzyme and the nucleobase is removed via a  $S_N1$ -type reaction after the attack of a water molecule. Then an endonuclease cuts the resulting abasic site (AP-site) moiety out of the DNA leading to a single strand break. In case of a bivalent nuclease the endonuclease is not necessary, as this step is carried out by the nuclease.<sup>[148-149]</sup> PARPs bind to these nicks and recruit a DNA polymerase and a ligase to fill in the gap.<sup>[150]</sup>
2. The alternative long patch mechanism involves the excision of 3-5 bases upstream by strand displacement of the polymerase followed by ligation.<sup>[148-149]</sup>

Both pathways result in the replacement of mC by C.

DNA-methylation in plants is not restricted to CpG sites as CpNpG and CpNpN (N = A, T or C) can be methylated as well.<sup>[66]</sup> The Dme family members are able to recognize and remove mC in this context *in vitro*.<sup>[151]</sup> A mutation in one of the family's genes leads to hypermethylation *in vivo* in all sequence contexts, however, at distinct genomic loci, suggesting an individual function for each of them.<sup>[151-154]</sup>

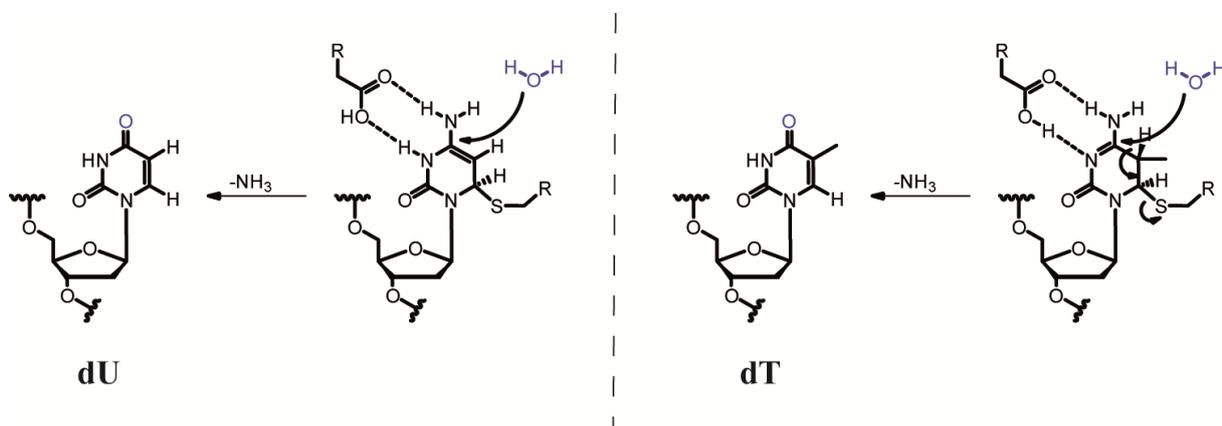
Despite the lack of Dme orthologes in mammals, the same pathway has been reported here as well. The first reports proposed Thymine DNA Glycosylase (TDG) as a glycosylase for mC, however, its activity on mC was 30-40 fold lower than on T-G mismatches.<sup>[155-157]</sup> Another enzyme proposed to excise mC is the methyl-CpG-binding protein MBD4.<sup>[158]</sup> Yet again its activity is 30-40 fold slower than on T:G mismatches. In addition MBD4 null mice are vital and exhibit normal demethylation of the zygotic paternal pronucleus, however, they show an increased number of C to T mutations.<sup>[159-160]</sup> These facts show that the biological relevance

of the excision of mC is presumably very low as impaired active demethylation should lead to developmental defects and no glycosylase with high activity for mC excision in mammals has been found to date.

### 2.3.2 Deamination of mC

Another pathway thought to play a role in active demethylation was deamination of mC and removal of the resulting T-G mismatch by the earlier mentioned TDG and MBD4. Cytosine deaminases as well as DNMTs have been proposed to play a role in the first part of this pathway. The most prominent members of cytosine deaminases are activation-induced deaminase (AID) and the apolipoprotein B pre-mRNA editing enzyme, catalytic polypeptide (APOBEC).<sup>[161-162]</sup> They have several important biological functions, such as antibody recombination in B-Cells, RNA editing and retroviral defense.<sup>[163]</sup> They are both capable of deaminating mC to T in single stranded DNA *in vitro* and in an *E.coli* assay.<sup>[164]</sup> As a consequence of their genomic location in a pluripotency gene cluster, AID and APOBEC are expressed in mouse oocytes, ES cells and primordial germ cells (PGCs).<sup>[164]</sup> While a large scale bisulfite sequencing study indicates a small contribution to active demethylation in PGCs, the global methylation level shows that the main factors of demethylation are yet to be identified.<sup>[165]</sup> Studies in zebra fish embryos could reveal that coexpression of AID and MBD4 leads to T-G mismatches that are repaired so rapidly that they could only be observed using a catalytic mutant of MBD4. When decreasing the expression of AID and APOBEC to a level which does not allow demethylation, the addition of growth arrest and DNA-damage-inducible 45 $\alpha$  (Gadd45a) protein was able to restore demethylation, indicating cooperative effects of the three proteins.<sup>[166]</sup> However, AID and APOBEC1 null mice do not show developmental defects, are vital and fertile.<sup>[167-170]</sup> Additionally, dsDNA cannot be deaminated *in vitro* by AID which is surprising if this mechanism was responsible for global DNA demethylation.<sup>[164]</sup> Also DNMTs have been reported to deaminate mC *in vitro* and *in vivo* on the promoter level. For rapid dynamics of methylation and demethylation using the same enzyme with two opposite activities would make perfect sense.<sup>[146]</sup> Chemically the saturation of the C(5)-C(6) bond in the covalent DNMT-DNA intermediate would facilitate the nucleophilic attack of water at the C(4) position. The stabilized amino group would act as a leaving group and its substitution would lead to either desoxyuridine or T depending on the presence of SAM (**Fig. 13**). DNMTs are known to associate with TDG and to stimulate each

other's activity.<sup>[171]</sup> Methylation activity of DNMT3a is reduced upon TDG interaction, whereas TDGs activity is enhanced.



**Figure 13: Proposed deamination mechanism in DNMT-DNA intermediates with (right) and without (left) SAM; nucleophilic attack of water on the C(4)-atom leading to the substitution of the stabilized amino group.<sup>[119]</sup>**

However, in this report deamination was only proved indirectly by a Ligation-mediated-PCR assay after potential deamination and AP-site cleavage and took place at very low SAM concentrations.<sup>[146]</sup> Therefore the biological relevance of this setup remains unclear, as SAM is ubiquitous in the cell and necessary for various biological processes, while this process would require rapid changes in the SAM-level.<sup>[4]</sup>

### 2.3.3 Oxidation of mC

Demethylation of histones is carried out by oxidation of the methyl group and subsequent release of formaldehyde.<sup>[60]</sup> A similar mechanism for DNA was proposed early as AlkB, a member of the DNA and 2-OG dependent dioxygenases family, has been shown to be involved in bacterial response to alkylation damage.<sup>[4]</sup> They are able to carry out the demethylation of 3-mC and 1-methyladenine using 2-OG, Fe<sup>II</sup> and molecular oxygen and release formaldehyde by a similar mechanism as histone demethylation.<sup>[172]</sup>

Oxidation of thymidine in various organisms is well known and studied. In trypanosomes a glucose moiety is attached to 5-hydroxymethyluridine (hmU), the oxidation product of T, to produce base J. The enzymes J-binding protein 1 and 2 (JBP1 and JBP2) were identified as the dioxygenases responsible for its formation.<sup>[173]</sup> The *Rao group* was able to identify homologues of these enzymes in mammals, when they found the ten eleven translocation (TET) family of proteins in 2009.<sup>[3]</sup> This discovery had a huge impact in the field of active

DNA demethylation as it identified a long time searched enzyme to perform the putative oxidative demethylation reaction.

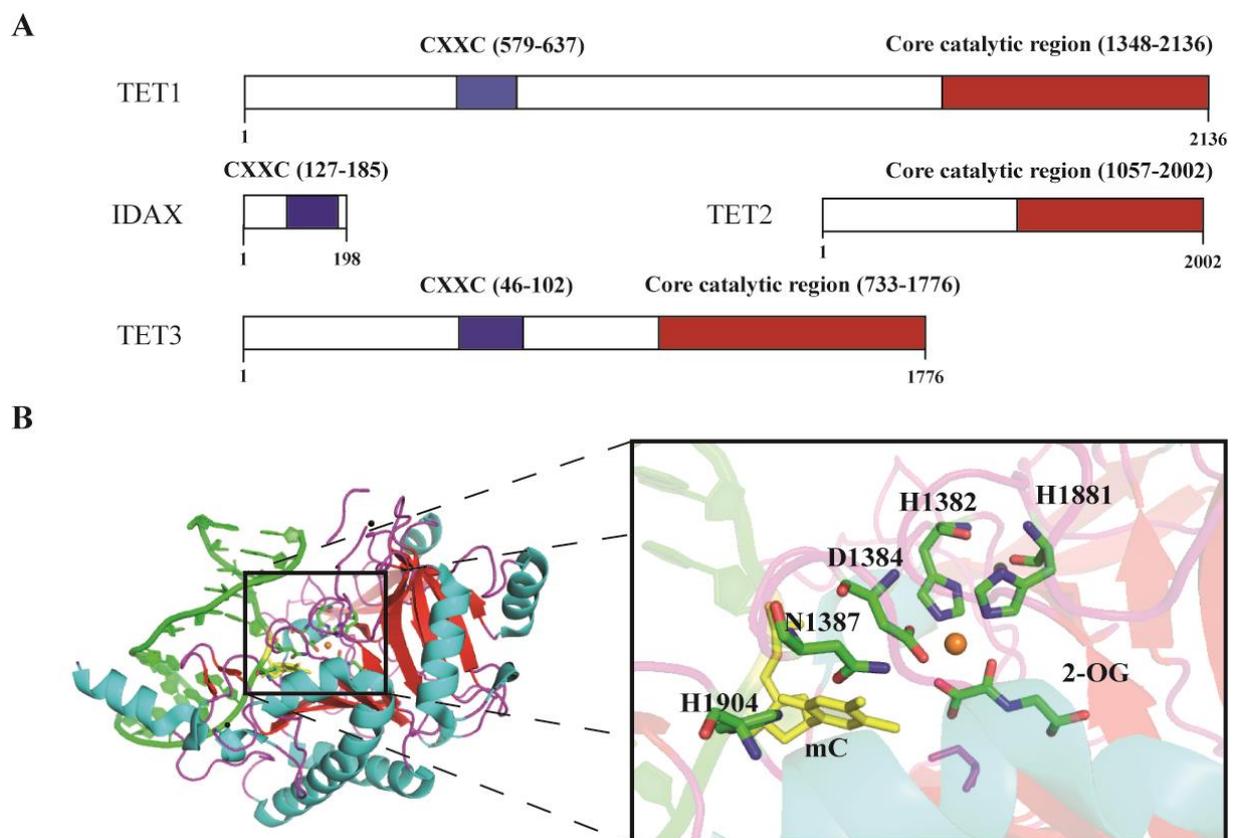
### 2.3.3.1 TET

Three TET-family proteins were discovered in the mouse genome, TET1, 2 and 3.<sup>[9]</sup> The transcription of these proteins is tightly controlled as their expression is crucial for the development status of the cell.<sup>[174]</sup> TET1 and TET2 are positively regulated by Oct4 and their mRNA as well as their protein levels decrease during differentiation. On the other hand TET3 is upregulated during that process.<sup>[175-176]</sup> Thus TET proteins are directly connected to the pluripotency network and are crucial to maintain ES cell identity, as demonstrated by *Ito et al.* in knock-down approaches.<sup>[174]</sup> A recent study suggests miRNA-22 to downregulate TET2.<sup>[177]</sup> Also recently *Zhang and coworkers* were able to show that Calpain-proteases are responsible for the degradation of TET proteins.<sup>[176]</sup> Several proteins and also metabolites have been found to regulate TET activity. UHRF2 and also L-ascorbate show positive stimulation of TET proteins either by direct interaction or indirect effects such as substrate targeting.<sup>[178-179]</sup> On the other hand CXXC5 and inhibition of the Dvl and Axin complex (IDAX) negatively stimulate TET activity.<sup>[180]</sup>

Computational screens revealed that IDAX was detached from TET2 by a chromosomal inversion.<sup>[181]</sup> It contains, like TET1 and TET3 a CXXC domain, a common DNA binding motif with three subfamilies.<sup>[181]</sup> TET proteins originated from a JBP-like gene that fused with a gene containing a CXXC domain. CXXC subfamily 1 recognizes specifically unmodified CpGs in DNA and targets the enzymes, such as DNMT1, MLL and KDMA, to it.<sup>[182-185]</sup> Subfamily 2 has yet only been identified in the first two CXXC domains of MDB1 and has not been shown to bind DNA or possess any other function.<sup>[182]</sup> Subfamily 3 includes the domains of IDAX, TET 1 and TET3.<sup>[181]</sup> Interestingly the TET1 CXXC domain is reported to not bind DNA at all, while the TET3 CXXC domain is reported to preferentially bind unmodified CpGs.<sup>[186-187]</sup> The CXXC domain of IDAX binds preferentially to unmethylated CpGs *in vitro* and is enriched at CpG islands and CpG-sites *in vivo*.<sup>[180]</sup>

Even though IDAX was detached during evolution it is targeting TET2 for caspase-dependent degradation. If key DNA binding residues are mutated in the CXXC domain, this negative regulation cannot occur. This finding supports the idea that IDAX recruits TET2 to certain gene loci before degradation occur.<sup>[180]</sup> While TET1 is reported to act mainly at TSS but also in the gene body, TET2 oxidizes mC in the gene body. The activity on the TSS has been

reported to negatively correlate with gene expression, while enzymatic activity in the gene body is positively correlated.<sup>[188]</sup> Consistently, TET1 has also been proposed to play a role in poised promoters, as its depletion leads to both the activation as well as the repression of some of its target. While the activation is proposed to depend on its catalytic activity, presumably in the gene body, the repression may be related to its interaction with the SIN3A repressor complex, but in any case not to its catalytic activity.<sup>[189-190]</sup> TET2 and also TET3 are able to recruit O-linked  $\beta$ -N-acetylglucosamine (GlcNAc) transferase (OGT) to the chromatin by a strong direct interaction. The formed complex leads to increased glycosylation of histones at TSS and also leads to increased methylation of H3K4.<sup>[191-193]</sup>

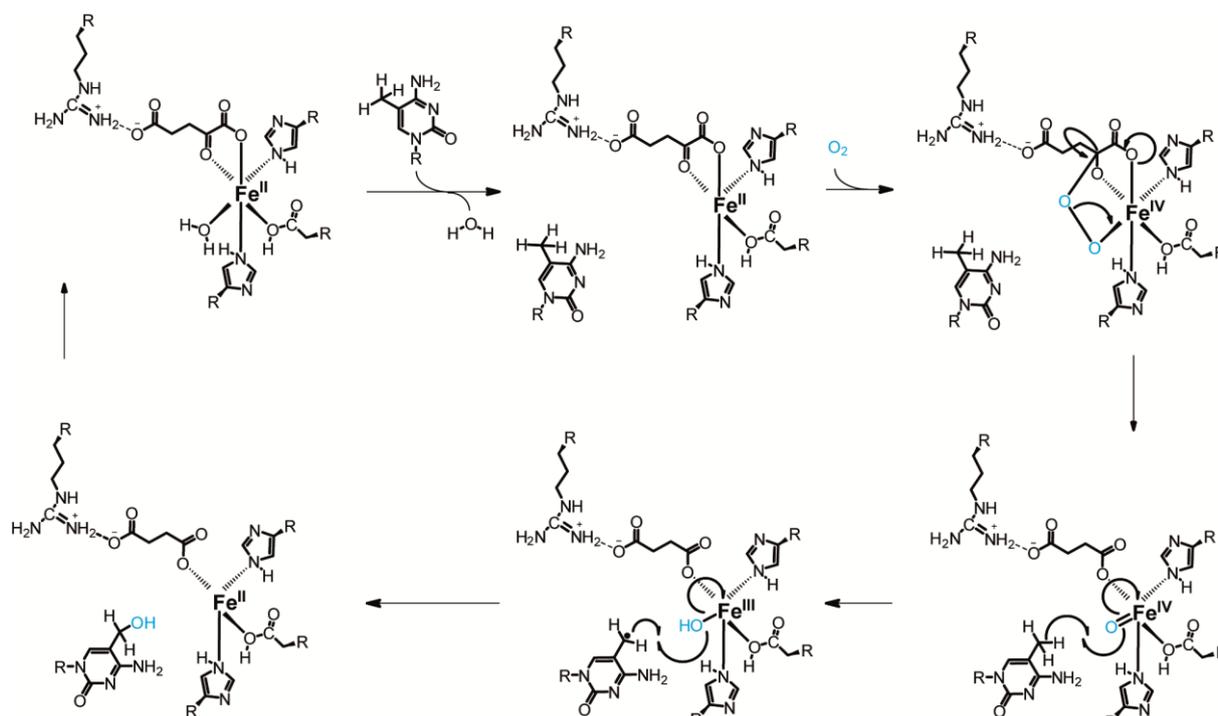


**Figure 14: A schematic view of the TET proteins; B crystal structure of TET2 catalytic domain and zoom into the active site. The substrate (yellow) is flipped out of the DNA and recognized by H1904 and D1384. The Fe<sup>II</sup> (orange sphere) is coordinated by D1384, H1382, H1881 and cofactor 2-oxoglutarate. Zinc ions in black sphere. PDB code: 4NM6.<sup>[194]</sup>**

In the core catalytic domain at their C-terminus all proteins possess a Cys-rich motif that is coordinating zinc ions and allows DNA binding.<sup>[194]</sup> Also the C-terminus contains a double stranded  $\beta$ -helix and a less conserved low complexity insert.<sup>[3,181]</sup> The active site consists of a conserved His-X-Asp/Glu motif specific for dioxygenases, an additional conserved

His-residue is also involved in coordinating  $\text{Fe}^{\text{II}}$  together with the cofactor 2-OG (**Fig. 14**). The cofactor is stabilized over ionic contacts to an Arg residue. Recently, the crystal structure of TET2 catalytic domain in complex with dsDNA has been solved and gave insight into the mechanism of action (**Fig. 15**).<sup>[194-195]</sup>

Base recognition of mC in the dsDNA by TET2 is achieved by two hydrogen bonds from H1904 and N1387 to the endocyclic N-atoms of mC and by base stacking to Y1902. The target base is flipped into the active site, which brings it near the iron cofactor which leads to a DNA torsion of  $40^\circ$ .



**Figure 15: Reaction mechanism of 5-methylcytosine oxidation by TET proteins. Upon binding of molecular oxygen (blue), 2-oxoglutarate is decarboxylated and an oxo-ferryl species is generated. Stepwise single electron transfer mechanisms lead to the oxidation of the C-H bond and regeneration of the  $\text{Fe}^{\text{II}}$  center.**<sup>[194-195]</sup>

In this structural study, mutations in the aminoacids responsible for base recognition lead, as expected, to decreased mC oxidation but not to significant DNA interaction changes. Investigation using oligonucleotides with dC in different sequence contexts shows that while mCpA DNA had less than 2%, mCpC DNA showed up to 10% and mCpG DNA at least 85% oxidized mC thus indicating a strong preference for mCpG-sites. However as the methyl group does not show any contacts to the catalytic site the preference cannot not be expanded to binding, conclusively mCpG and CpG have similar binding affinity. *Hu et al.* speculate that

this might be due to the different characteristics of the oxidation products compared to mC. Stabilization of the methyl group by a hydrophobic contact would block the interaction of the oxidation products that contain hydrophilic groups.<sup>[194]</sup>

The products are generated by decarboxylation of the cofactor upon binding of molecular oxygen and oxidation of Fe<sup>II</sup> to Fe<sup>IV</sup> (**Fig. 15**). The resulting oxo-ferryl species is able to oxidize the nearby C-H bond by a single electron transfer generating a methyl radical and an Fe<sup>III</sup> specie as intermediates. The O-atom is transferred with another single electron transfer to the methyl radical and the Fe<sup>III</sup> is accepting the electron to recover the Fe<sup>II</sup> center and close the enzymatic cycle by the release of the product, 5-hydroxymethylcytosine (hmC).<sup>[196-197]</sup>

### 2.3.3.2 Hydroxymethylcytosine

This oxidized mC was already discovered in 1952 in phages and 1972, at suspicious high levels, in mammals and was later believed to be an oxidative damage.<sup>[198-200]</sup> However, as its (re-)discovery revealed it as a product of dioxygenases, it was opening the road to an active demethylation cycle.<sup>[4]</sup> The level of hmC is dependent on the expression-level of TET proteins. It was shown that the already mentioned genome wide loss of mC in the male pronucleus after fertilization is not a conversion to dC but is going hand in hand with TET3 expression and the oxidation to hmC.<sup>[201]</sup> The highest levels of hmC, with 0.7% of all cytosines, have been detected in the brain, which corresponds to about 12% of all CpGs.<sup>[202]</sup> In mESC the level is about 0.2% of all cytosines and upon differentiation the level is decreasing.<sup>[7]</sup> Aberrant levels of hmC are correlating with several diseases, especially in the brain.<sup>[203]</sup> In astroglomas hmC is significantly reduced and might be used as a marker for classification of tumors into the World Health Organization (WHO) grades.<sup>[6]</sup> In psychotic disorders hmC and TET1 levels are significantly higher.<sup>[204]</sup>

To determine the genomic distribution, different genome-wide mapping approaches of hmC have been performed and revealed its accumulation in gene bodies, promoters and TSS.<sup>[189-190,205-206]</sup> The location of hmC suggests a role in transcription, however, this role is so far not completely understood. It is possible that it activates transcription by blocking the binding of MBPs.<sup>[207]</sup> On the other hand it is reported that in patients with psychotic disorders, increased hmC-levels at the TSS of the GAD67 promoter, leads to reduced mRNA levels.<sup>[204]</sup> Consistent with that is the earlier mentioned finding of enhanced transcription after depletion of TET1, which is mainly generating hmC at TSS.<sup>[188]</sup> Further research effort has to be done to reveal the exact role of hmC in transcription.

Recently, two oxidative variants of bisulfite sequencing have been established for base resolution sequencing of hmC. In principle they rely on computational comparison of bisulfite sequencing before and after an oxidation step, carried out by either TET (TET-assisted BS) or chemically with a  $\text{KRuO}_4$  complex.<sup>[208-209]</sup> This is possible due to the different chemical behavior of further oxidized cytosines and hmC during sequencing.

With a proteomic approach, proteins that bind hmC have been identified, among them are the glycosylases NEIL1 and MBD2, the factor necessary for maintenance methylation UHRF1 and several proteins with a yet unknown function such as UHRF2, Wdr76 and C3orf37. In this study UHRF2 was shown to have a positive effect on TET-activity and to rise the hmC level by a factor of 1.2.<sup>[178]</sup> It was shown that UHRF1 binds hmC over its SET and RING finger associated (SRA) domain with a similar affinity to mC, opening a possible way for hmC inheritance.<sup>[210]</sup> However evidence from pre-implantational embryos suggests that hmC is passively removed through replication. Unlike DNMT1, DNMT3a and DNMT3b have been shown to methylate opposite hemi-hmC DNA than unmethylated DNA in a similar rate.<sup>[211]</sup> Conclusive with that *de novo* methylation has been shown to play the major role in hmC production.<sup>[212]</sup> Even though hmC inheritance therefore is in principle possible, there is no direct evidence for it.

It has been reported that hmC can be deaminated to hmU, which would lead similar to the earlier mentioned deamination of mC to a mismatch and DNA repair.<sup>[213]</sup> In the study AID and TET1 overexpression lead to the detection of hmU, however, recent work in our group could show that TET1 catalytic activity is not restricted to mC but can also take place on dT. This explains the detection of hmU even though deamination is still possible.

### **2.3.3.3 Other oxidized pyrimidine species**

During this work we were able to discover the further oxidation product of hmC, 5-formyldeoxycytidine (fC) in the genomic DNA of mESC, and estimated a level of 0.02% of all Cytosine derivatives.<sup>[7]</sup> Shortly thereafter, the discovery and the connection to TET enzymes could be confirmed by the *Zhang group*, who together with the *Xu group* were also able to detect 5-carboxyldeoxycytidine (caC) as the last product of mC oxidation cascade.<sup>[9,214]</sup> The reaction rate to these 2 products are significantly lower than the production of hmC and the levels of fC are only 1/10 respectively 1/100 for caC of the hmC level.<sup>[9]</sup> This is about 3.5 caC per  $10^6$  nucleotides and about 1000 molecules per genome, in comparison 8-oxo-guanosine, a known marker for oxidative damage is reported to occur in 1 per  $10^{6-7}$

nucleotides.<sup>[215]</sup> Despite their low abundance, they might play an important role in active demethylation and epigenetic regulation, because these modifications occur mainly at CpG islands and are considered transient intermediates.

The hydrophilic properties of these new epigenetic bases are similar to those of hmC and contrary to mC. However the carbonic acid of caC which should be deprotonated under physiological conditions, allows ionic contacts and therefore, theoretically tighter binding or repulsion.

By LC-MS based approaches several selective readers for these new DNA modifications have been identified, however, in comparison to caC much more is known about fC, which might be due to its higher abundance.<sup>[178,216]</sup> Interestingly some of the readers, such as FOXK1 and 2 are known as transcriptional regulators.<sup>[216]</sup> The genomic distribution of fC was initially determined by a chemical ligation of the aldehyde with an aldehyde reactive probe (ARP), a technique that was first introduced by us (*Pfaffeneder et al.*).<sup>[7]</sup> The fragments containing the ligated fC-ARP product were enriched using streptavidin beads. Sequencing revealed that like hmC fC is enriched around the TSS, H3K4 trimethylated histones and is frequently bound to RNA polymerase II.<sup>[8]</sup> The findings of this study strongly correlate fC to transcription, however, the function is not 100% clear. When fC and caC are located in the gene body *in vitro*, the fidelity of RNA polymerase II is impaired and the polymerase is stalled.<sup>[217]</sup> *In vivo* this would lead to pre-terminated RNA fragments. The authors of the study therefore suggest a role in transcriptional regulation and mRNA processing, as e.g. splicing. It has been also recently reported that caC is enriched in euchromatic regions in follicular cells, also suggesting a role in transcriptional regulation.<sup>[218]</sup>

Recently, two chemical modification variants of bisulfite sequencing (BS) were established for base resolution sequencing of fC and caC. During BS, dC, fC and caC are deaminated to dT while mC and hmC are not. Like the base resolution sequencing of hmC they rely on computational comparison of bisulfite sequencing before and after the chemical modification of fC/caC, as the modification prevents fC and caC from deamination to dT.<sup>[219-220]</sup> The different chemical behavior of dC derivatives during bisulfite sequencing, was investigated by *Jin et al.* in the detail.<sup>[221]</sup> These new technique might give insights on the exact genomic location and function of the newly discovered bases.

An antibody study in mouse embryos could show that fC and caC are generated in pre-implantational mouse embryos and are then diluted through replication.<sup>[222]</sup> The authors also state that due to the behavior of caC in BS, earlier studies reporting demethylation have to be reinvestigated.<sup>[222]</sup> The levels of these modifications during early mESC differentiation

have also been studied, and the kinetics of the nucleosides have been compared in a cluster analysis. This revealed that fC and caC behave similar, with and peak early in their levels. In comparison hmC behaves differently and could not be correlated with them. This finding supports the idea of hmC as a stable epigenetic base and fC/caC as intermediates in an active demethylation cycle. In this study also hmU as a product of T oxidation by TET enzymes has been discovered, and some direct and indirect binders have been identified using a proteomic approach. However, further studies have to be made to elucidate biological functions of this new modification.

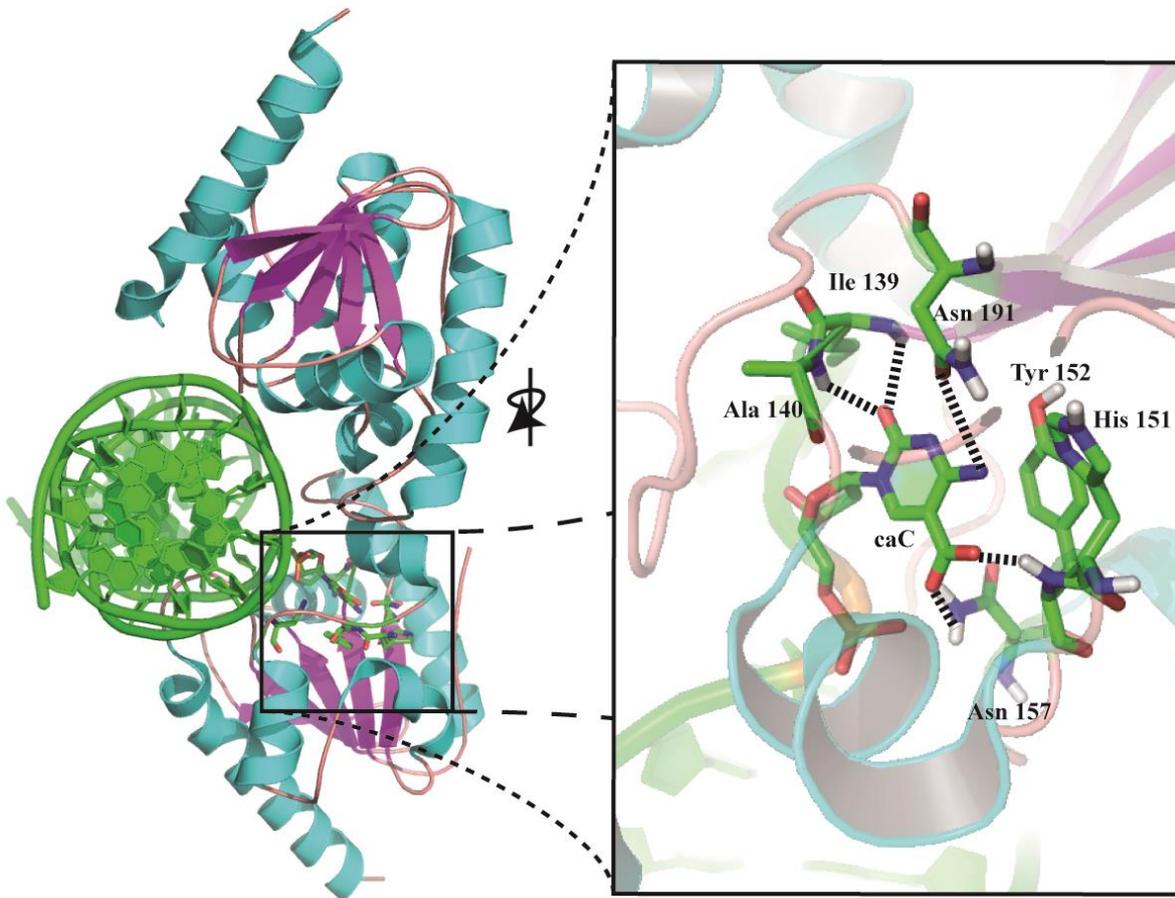
#### **2.3.3.4 Excision of new epigenetic bases by BER**

Shortly after the discovery of fC and caC, it was reported that they are substrates for TDG and it was proposed that active demethylation would occur by BER.<sup>[214,223]</sup> In fact TDG knock-out cells show an increase of around 40% in fC and of about the same increase in caC. The activity of the enzyme on fC is 1.5 fold faster than on T:G and 4 fold slower on caC, hmC was found not to be a substrate for TDG.<sup>[223]</sup> It was also reported that hmU is excised by TDG with significant but lower activity.<sup>[224-225]</sup> A crystal structure of caC and hmU containing DNA was obtained by different groups and gave insights into the molecular mechanism and discrimination of the different substrates (**Fig. 16**).<sup>[225-226]</sup> Several amino acids, such as Ser271, Ser272, His151, Asn157, Asn191 and Asn230, interact with the accommodated base in the cage of the active site and are able to act as hydrogen-donors or acceptors.

The catalytic activity is established by the activation of a water molecule by Asn140, which leads to a putative oxacarbenium ion intermediate.<sup>[227-228]</sup> The conserved Tyrosine residue 152 stacks with the AP-site in the postreactive complex. For caC Asn191 seems to be the most important contact, where as it is not necessary for TDG activity on fC (**Fig. 16**). It is proposed that it stabilized the N3-protonated species of caC, whose excision is unlike fC pH-dependent.<sup>[229]</sup> A mutation of Asn230 or Asn157 results in greatly reduced activity on hmU, but does not affect caC excision significantly. While mutations in both abolish excision of caC but not of hmU.<sup>[230]</sup>

While fC and caC seem to be exclusively excised by TDG, hmU is recognized and removed by 3 distinct glycosylases. Next to the earlier mentioned activity of TDG, MBD4 and SMUG1 are able to catalyze the hydrolysis to an AP-site, with a 6-10 fold reduced activity on hmU:G compared to TDG.<sup>[224,231]</sup>

Due to the high activity TDG has on the newly discovered fC and caC bases *in vitro* and *in vivo*, as well as the fact that TDG knock-out mice are not viable and show developmental



**Figure 16:** Crystal structure of human TDG catalytic domain mutant N140A in complex with A:caC mismatch dsDNA. Zoom into the active site showing caC coordinating amino acids and hydrogen bonds. PDB code: 3UO7.<sup>[226]</sup>

defects, it is conclusive to see this as a likely pathway for DNA demethylation.<sup>[232-233]</sup> However the generation of abasic sites (naturally occurring 1 per  $10^5$  nucleotides) in all fC and caC-sites would lead to an about 4-fold increased number of AP-sites and therefore to a greater genomic instability, which is in respect to embryonic development very unfavorable.<sup>[234]</sup>

### 2.3.3.5 C-C bond cleavage pathway

An attractive alternative for the BER pathway is the cleavage of the C(5)-C(exo) bond of the oxidized dC-derivatives by an enzyme. A mutant of the bacterial DNA methyltransferase HhaI has been reported to reversibly add/remove aldehydes using different substrates, to

result among others in hmC.<sup>[235]</sup> In 2012 DNMT3a and 3b but not DNMT1 were reported to have the same enzymatic reactivity depending on their redox state and the presence of SAM. A catalytically inactive enzyme did not show this activity and thus DNMTs were suggested to act as bidirectional enzymes.<sup>[236]</sup> Shortly thereafter the same group reported that all three human DNMTs are able to directly demethylate mC dependent on Ca<sup>II</sup>.<sup>[237]</sup> Very recently *Klimasauskas and coworkers* were able to show similar effects of methyltransferases on hmC and caC.<sup>[238]</sup> Interestingly in this thesis, it was possible to break the C-C bond by using thiols and basic amino acids, mimicking active sites which contain these amino acids like DNMTs.<sup>[10,239]</sup> It was shown, that incubation of hmC with high concentrations of  $\beta$ -mercaptoethanol and imidazole leads to breakage of the C-C bond and formation of dC. However, the cleavage of hmC is very slow and resulted in a yield below 1%, which is far below the yield of dehydroxymethylation reported by DNMTs or photohydration.

Higher yields were detected for the cleavage of fC, resulting in formic acid (2.5%) and for cleavage of caC which leads to the formation of carbon dioxide (28%). Mechanistically, the acid catalyzed addition of a thiol to the C(6) atom of the cytosine derivatives results in a saturated C(5)-C(6) bond and facilitates the elimination of the leaving groups. For fC the aldehyde has to be attacked nucleophilically by a thiol or water before acting as a leaving group. Interestingly, during a proteomic approach a specific binder for fC was discovered bearing a formyltransferase activity, which might be a putative deformylase by transferring the formyl group on metabolites, as for an example tetrahydrofolate.<sup>[216]</sup> Quantum chemical calculations showed, that saturation of the C(5)-C(6) double bond results in a far lower energy of a potential vinyl anion intermediate after the release of carbon dioxide from the molecule. According to the calculations the energy barrier would drop by about 40 kcal/mol, which could enable a very fast decarboxylation process.<sup>[239]</sup>

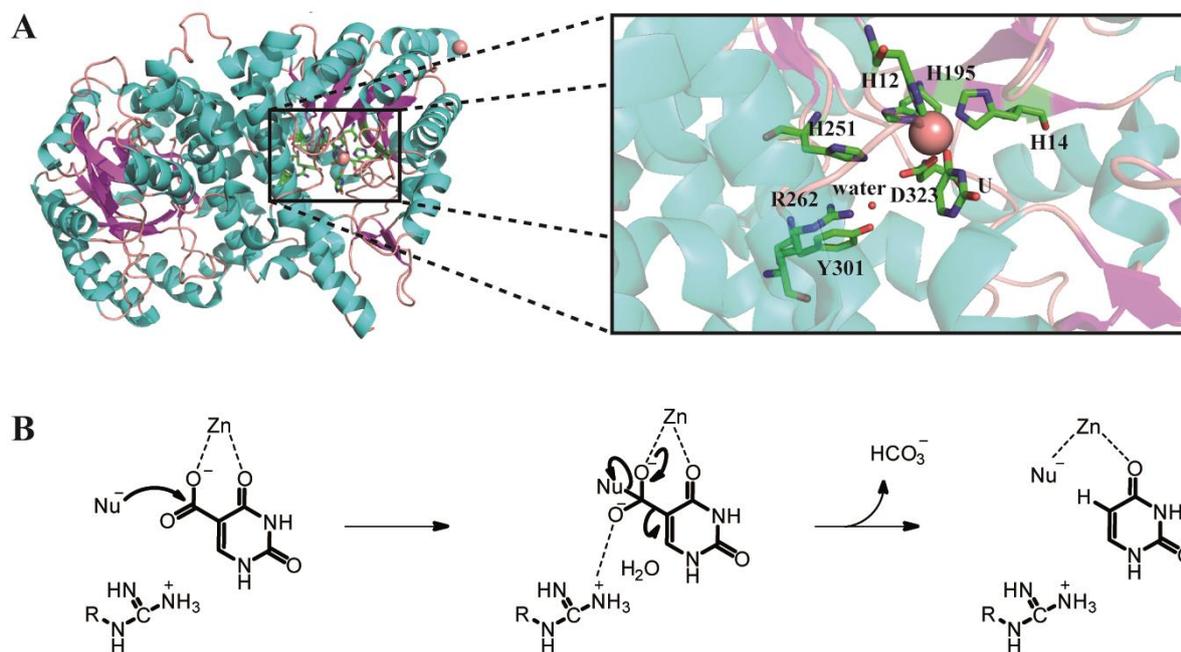
Decarboxylation is a very frequent process in nature and very well understood, therefore a quite convincing candidate-pathway for demethylation of mC. It takes places in various metabolic pathways, in which organic acids need to be converted such as fatty acid synthesis or the thymidine salvage pathway. This is managed by decarboxylases of which at least 90 different classes according to IUPAC are known. A great variety of mechanisms involve different cofactors such as organic biotin or pyridoxalphosphate (PLP) or inorganic cofactors such as metal complexes.<sup>[240]</sup>

However, some decarboxylases do not need any cofactors. Interestingly one of these is isoorotate-decarboxylase in the thymidine salvage pathway, an enzyme the proposed caC-

decarboxylase could be very similar to. IDcase is the enzyme with the highest rate enhancement known to date, accelerating the reaction  $10^{17}$  fold.<sup>[241]</sup>

The crystal structure of IDcase has recently been solved and subsequently more insight into the mechanism is gained (**Fig. 17**).<sup>[242]</sup> The active site consists of three histidines and one aspartate coordinating a zinc ion and several other conserved residues coordinating water. Two models for the mechanism have been proposed upon mutational studies:<sup>[242]</sup>

1. In the first model the aspartate acts as a nucleophile to attack the C(exo) and form an unstable tetrahedral intermediate (**Fig. 17**). Upon elimination from the aromatic ring a vinyl anion that is quickly protonated is formed. This is followed by elimination of hydrogen carbonate after attack of a water molecule on the aspartate-carboxyl intermediate. Then the product is released and the enzymatic cycle can be restarted.
2. In the second model, the nucleophilic attack is carried out by a water molecule after activation by the aspartate. Again a tetrahedral intermediate is formed and the C-C(exo) bond is polarized and broken. After quick protonation of the vinyl intermediate and release of hydrogen carbonate, the product is released.



**Figure 17:**A) Crystal structure of IDcase with U in the active site; zoom into active site and aminoacids involved in the catalytic activity indicated with stick illustration and zinc in light orange as sphere illustration. PDB code: 4HJW.<sup>[242]</sup> B) Proposed mechanism of IDcase with a tetrahedral intermediate, stabilized by N262.

In each of the model mechanisms the N262 residue is stabilizing the water molecules and intermediates during the process. Interestingly the authors of the study could show that the enzyme was capable of forming cytidine from carboxycytidine. They suggest that a putative caC-decarboxylase might consist of a similar active site but would interact differently with metal ions as the amino group is a stronger electron donor. Additionally a DNA binding motif must exist as caC is only formed in dsDNA *in vivo*, resulting in a larger enzyme. Bioinformatic research could help to identify such an enzyme.<sup>[242]</sup>

In this thesis the decarboxylation activity of nuclear extracts of mESC was studied using MS-based isotope tracing in dsDNA. An oligo containing double <sup>15</sup>N labeled caC was incubated, reisolated and analyzed after total digestion, resulting in a signal for the labeled dC. Even though the study could not identify a decarboxylase, it shows that the existence of such an enzyme can be proven by analysis on the molecular level.<sup>[10]</sup>

## **II. Aim of thesis**

DNA methylation and active DNA demethylation are processes with high relevance to modern medicine. While DNA methylation has been in the focus of research since decades, active demethylation is a relatively young scientific field. Several pathways of demethylation have been investigated, but with the discovery of hmC as a product of TET enzymes, the oxidative demethylation pathways have been in the focus.

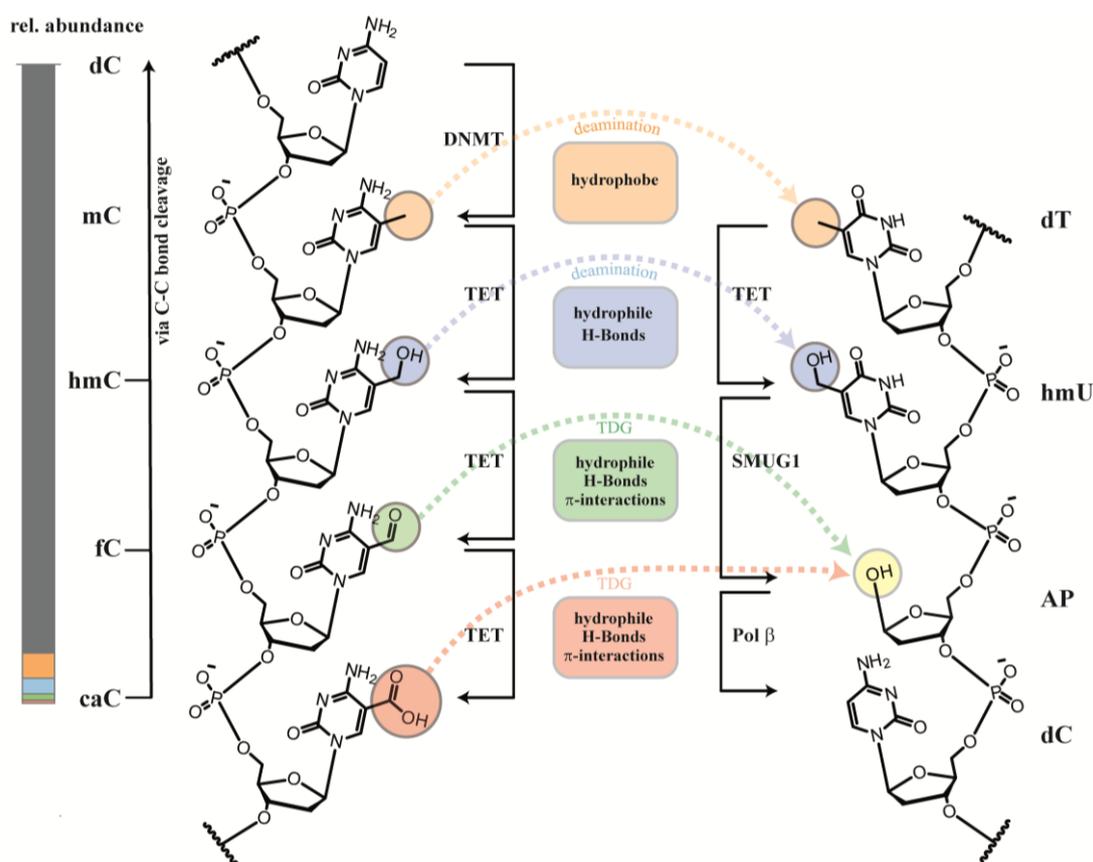
The search for intermediates of this DNA-demethylation pathway in the brain did not lead to the identification of fC or caC. The main goal of this thesis was to determine the fate of hmC in the cell, to continue the search for the intermediates of DNA demethylation and to investigate the process of DNA methylation/demethylation. This process is thought to take place mainly in pluripotent cells, such as mESC. Therefore they are a reasonable biological system for the investigation. The study was designed to give more insights into the pathways of active demethylation and their functions. To investigate the dynamics of DNA-methylation/demethylation the development of biochemical tools and analytical methods was necessary.

### III. Conclusion and Outlook

The investigation of cytosine dynamics is still ongoing and after the discovery of fC and caC more and more insights are gained. Even though the formation of mC has been investigated for decades, the very complex methyltransferase enzymes are still not fully understood. In this thesis a recently stated self-regulation mechanism of DNMT1 could be confirmed.

The pathway of combined oxidation and excision of mC-derivatives (**Fig.18**) was convincingly proven *in vitro* and *in vivo*, but the selectivity of genomic location still needs to be addressed. The C-C bond cleavage pathway needs further validation, as especially decarboxylation seems to be very attractive. First hints for this process could be found in this thesis but yet no decarboxylase has been identified. The bivalent function of DNMTs are yet to be confirmed on the molecular level, however quick dynamics and tight regulation could be a challenge *in vitro* as well as *in vivo*.

Additionally deamination of mC and hmC might play a role for the dynamics of cytidines and might couple them to those of uridines (**Fig. 18**).



**Figure 18:** Overview of active demethylation pathways with chemical properties (middle) and relative abundancies (left bar).

Nature invests energy into the generation of fC, caC and hmU and removes them indirectly under the pressure of genomic instability. Different biological processes that involve the different cytosine modifications might make the variety of intermediates and pathways necessary, but yet we are not at the point of understanding them completely.

These processes include gene regulation, in which the role of the demethylation intermediates is yet not fully understood. However, hints for bivalent roles of modified CpGs arise. In this context also the crosstalk of histone modifications and new cytosine modifications would be interesting. Also the effects on *trans*-elements is yet not investigated, however the different chemical behavior of caC and mC could lead to different DNA-histone/protein interactions upon modification change and could trigger gene regulation (**Fig. 18**).

The metabolism of the newly discovered bases is also unknown, therefore it might be interesting to include urine samples and metabolized cytosine modifications, for example glucose-conjugated caC, in the investigations. The negative charge would need to be neutralized in order to pass the cell membrane and degrade caC after its excision, therefore a metabolization is very likely.

Due to the different chemical behavior in bisulfite sequencing older data has to be reconsidered and confirmed.

Further research using modern tools in biochemistry, such as next generation sequencing, LC-MS and isotope tracing as well as established methods as genetic manipulations should be able to shed light into cytosine dynamics and its effects on biological processes. Understanding the pathways and functions could help us to gain more insights in cellular development, reprogramming and altered expression levels of oncogenes and ultimately to contribute modern medicine.

## IV. Literature

- [1] P. Bashtrykov, *et al.*, *Chembiochem* **2014**, 10.1002/cbic.201300740, n/a-n/a.  
*Targeted Mutagenesis Results in an Activation of DNA Methyltransferase 1 and Confirms an Autoinhibitory Role of its RFTS Domain*
- [2] K. Takeshita, *et al.*, *P. Natl. Acad. Sci. USA* **2011**, 108, 9055-9059.  
*Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1)*
- [3] M. Tahiliani, *et al.*, *Science* **2009**, 324, 930-935.  
*Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*
- [4] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, 11, 607-620.  
*Active DNA demethylation: many roads lead to Rome*
- [5] D. Globisch, *et al.*, *PLoS One* **2010**, 5, e15367.  
*Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates*
- [6] T. F. Kraus, *et al.*, *Int. J. Cancer* **2012**, 131, 1577-1590.  
*Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors*
- [7] T. Pfaffeneder, *et al.*, *Angew. Chem., Int. Ed. Engl.* **2011**, 50, 7008-7012.  
*The discovery of 5-formylcytosine in embryonic stem cell DNA*
- [8] E.-A. Raiber, *et al.*, *Genome Biol.* **2012**, 13, R69.  
*Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase*
- [9] S. Ito, *et al.*, *Science* **2011**, 333, 1300-1303.  
*Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine*
- [10] S. Schiesser, *et al.*, *Angew. Chem., Int. Ed. Engl.* **2012**, 51, 6516-6520.  
*Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing*
- [11] B. Steigenberger, *et al.*, *Org. Lett.* **2013**, 15, 366-369.  
*Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction*
- [12] R. Asch, C. Simerly, T. Ord, V. A. Ord, G. Schatten, *Hum. Reprod.* **1995**, 10, 1897-1906.  
*Fertilization and development: The stages at which human fertilization arrests: microtubule and chromosome configurations in inseminated oocytes which failed to complete fertilization and development in humans*
- [13] S. Mitalipov, D. Wolf, in *Engineering of Stem Cells, Vol. 114* (Ed.: U. Martin), Springer Berlin Heidelberg, **2009**, pp. 185-199.
- [14] F. Ulloa-Montoya, C. M. Verfaillie, W.-S. Hu, *J. Biosci. Bioeng.* **2005**, 100, 12-27.  
*Culture systems for pluripotent stem cells*
- [15] G. R. Martin, *P. Natl. Acad. Sci. USA* **1981**, 78, 7634-7638.  
*Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells*
- [16] M. J. Evans, M. H. Kaufman, *Nature* **1981**, 292, 154-156.  
*Establishment in culture of pluripotential cells from mouse embryos*
- [17] L. C. Stevens, C. C. Little, *P. Natl. Acad. Sci. USA* **1954**, 40, 1080-1087.  
*Spontaneous Testicular Teratomas in an Inbred Strain of Mice*
- [18] K. Takahashi, S. Yamanaka, *Cell* **2006**, 126, 663-676.  
*Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors*
- [19] Stuart H. Orkin, K. Hochedlinger, *Cell* **2011**, 145, 835-850.  
*Chromatin Connections to Pluripotency and Cellular Reprogramming*
- [20] H. Obokata, *et al.*, *Nature* **2014**, 505, 676-680.  
*Bidirectional developmental potential in reprogrammed cells with acquired pluripotency*
- [21] H. Obokata, *et al.*, *Nature* **2014**, 505, 641-647.  
*Stimulus-triggered fate conversion of somatic cells into pluripotency*
- [22] N. Malik, M. Rao, in *Pluripotent Stem Cells, Vol. 997* (Eds.: U. Lakshmipathy, M. C. Vemuri), Humana Press, **2013**, pp. 23-33.
- [23] S. Cao, K. Loh, Y. Pei, W. Zhang, J. Han, *Protein & Cell* **2012**, 3, 834-845.  
*Overcoming barriers to the clinical utilization of iPSCs: reprogramming efficiency, safety and quality*
- [24] A. Ohazama, S. A. C. Modino, I. Miletich, P. T. Sharpe, *Journal of Dental Research* **2004**, 83, 518-522.  
*Stem-cell-based Tissue Engineering of Murine Teeth*
- [25] A. A. Volponi, Y. Pang, P. T. Sharpe, *Trends Cell Biol* **2010**, 20, 715-722.

- Stem cell-based biological tooth repair and regeneration*  
 [26] R. Holliday, *Biol. Rev.* **1990**, 65, 431-471.
- Mechanisms for the control of gene activity during development*  
 [27] J. Nichols, A. Smith, *Cell Stem Cell* **2009**, 4, 487-492.
- Naive and Primed Pluripotent States*  
 [28] D. Van Hoof, C. L. Mummery, A. J. Heck, J. Krijgsveld, *Expert Rev. Proteomics* **2006**, 3, 427-437.
- Embryonic stem cell proteomics*  
 [29] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, 389, 251-260.
- Crystal structure of the nucleosome core particle at 2.8[thinsp]Å resolution*  
 [30] P. B. Talbert, S. Henikoff, *Nat. Rev. Mol. Cell Biol.* **2010**, 11, 264-275.
- Histone variants [mdash] ancient wrap artists of the epigenome*  
 [31] J. D. McGhee, G. Felsenfeld, *Annu. Rev. Biochem.* **1980**, 49, 1115-1156.
- Nucleosome Structure*  
 [32] T. Kouzarides, *Cell* **2007**, 128, 693-705.
- Chromatin Modifications and Their Function*  
 [33] M. Tan, *et al.*, *Cell* **2011**, 146, 1016-1028.
- Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification*  
 [34] D. M. Phillips, *Biochem. J.* **1963**, 87, 258-263.
- The presence of acetyl groups of histones*  
 [35] R. Marmorstein, S. Y. Roth, *Curr. Opin. Gen. Dev.* **2001**, 11, 155-161.
- Histone acetyltransferases: function, structure, and catalysis*  
 [36] S. A. M. Thiagalingam, *et al.*, *Ann. N. Y. Acad. Sci.* **2003**, 983, 84-100.
- Histone Deacetylases: Unique Players in Shaping the Epigenetic Histone Code*  
 [37] V. G. Allfrey, R. Faulkner, A. E. Mirsky, *P. Natl. Acad. Sci. USA* **1964**, 51, 786-794.
- Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis*  
 [38] L. Hong, G. P. Schroth, H. R. Matthews, P. Yau, E. M. Bradbury, *J. Biol. Chem.* **1993**, 268, 305-314.
- Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA*  
 [39] P. C. Megee, B. A. Morgan, M. M. Smith, *Genes Dev.* **1995**, 9, 1716-1727.
- Histone H4 and the maintenance of genome integrity*  
 [40] T. K. Barth, A. Imhof, *Trends Biochem. Sci.* **2010**, 35, 618-626.
- Fast signals and slow marks: the dynamics of histone modifications*  
 [41] T. Banerjee, D. Chakravarti, *Mol. Cell. Biol.* **2011**, 31, 4858-4873.
- A Peek into the Complex Realm of Histone Phosphorylation*  
 [42] P. O. Hassa, S. S. Haenni, M. Elser, M. O. Hottiger, *Microbiol. Mol. Biol. Rev.* **2006**, 70, 789-829.
- Nuclear ADP-Ribosylation Reactions in Mammalian Cells: Where Are We Today and Where Are We Going?*  
 [43] S. Messner, M. O. Hottiger, *Trends Cell Biol.* **2011**, 21, 534-542.
- Histone ADP-ribosylation in DNA repair, replication and transcription*  
 [44] Steven J. Petesch, John T. Lis, *Mol. Cell* **2012**, 45, 64-74.
- Activator-Induced Spread of Poly(ADP-Ribose) Polymerase Promotes Nucleosome Loss at Hsp70*  
 [45] G. R. Green, D. L. Poccia, *Dev. Biol.* **1985**, 108, 235-245.
- Phosphorylation of sea urchin sperm H1 and H2B histones precedes chromatin decondensation and H1 exchange during pronuclear formation*  
 [46] K. Murray, *Biochemistry* **1964**, 3, 10-15.
- The Occurrence of  $\epsilon$ -N-Methyl Lysine in Histones*  
 [47] W. K. Paik, S. Kim, *Biochem. Biophys. Res. Comm.* **1967**, 27, 479-483.
- $\epsilon$ -N-dimethyllysine in histones*  
 [48] K. Hempel, H. W. Lange, L. Birkofer, *Naturwissenschaften* **1968**, 55, 37-37.
- $\epsilon$ -N-Trimethyllysin, eine neue Aminosäure in Histonen*  
 [49] S. Rea, *et al.*, *Nature* **2000**, 406, 593-599.
- Regulation of chromatin structure by site-specific histone H3 methyltransferases*  
 [50] Q. Feng, *et al.*, *Curr. Biol.* **2002**, 12, 1052-1058.
- Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases without a SET Domain*  
 [51] T. C. Petrossian, S. G. Clarke, *Mol. Cell. Prot.* **2011**, 10, M110 000976.
- Uncovering the human methyltransferasome*  
 [52] D. Patnaik, *et al.*, *J. Biol. Chem.* **2004**, 279, 53248-53258.
- Substrate Specificity and Kinetic Mechanism of Mammalian G9a Histone H3 Methyltransferase*  
 [53] G. E. Zentner, S. Henikoff, *Nat. Struct. Mol. Biol.* **2013**, 20, 259-266.
- Regulation of nucleosome dynamics by histone modifications*

- [54] L. Zeng, M.-M. Zhou, *FEBS Lett.* **2002**, *513*, 124-128.  
*Bromodomain: an acetyl-lysine binding domain*
- [55] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025-1040.  
*How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers*
- [56] A. Barski, *et al.*, *Cell* **2007**, *129*, 823-837.  
*High-Resolution Profiling of Histone Methylations in the Human Genome*
- [57] D. Bonenfant, *et al.*, *Mol. Cell. Prot.* **2007**, *6*, 1917-1932.  
*Analysis of Dynamic Changes in Post-translational Modifications of Human Histones during Cell Cycle by Mass Spectrometry*
- [58] Y. Shi, *et al.*, *Cell* **2004**, *119*, 941-953.  
*Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1*
- [59] Y.-i. Tsukada, *et al.*, *Nature* **2006**, *439*, 811-816.  
*Histone demethylation by a family of JmjC domain-containing proteins*
- [60] R. J. Klose, *et al.*, *Nature* **2006**, *442*, 312-316.  
*The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine[thinsp]9 and lysine[thinsp]36*
- [61] M. Albert, K. Helin, *Semin. Cell Dev. Biol.* **2010**, *21*, 209-220.  
*Histone methyltransferases in cancer*
- [62] U. Ben-David, O. Kopper, N. Benvenisty, *Cell Stem Cell* **2012**, *10*, 666-677.  
*Expanding the Boundaries of Embryonic Stem Cells*
- [63] H. Hashimoto, P. M. Vertino, X. Cheng, *Epigenomics* **2010**, *2*, 657-669.  
*Molecular coupling of DNA methylation and histone methylation*
- [64] E. Vire, *et al.*, *Nature* **2006**, *439*, 871-874.  
*The Polycomb group protein EZH2 directly controls DNA methylation*
- [65] F. Fuks, W. A. Burgers, N. Godin, M. Kasai, T. Kouzarides, *EMBO J.* **2001**, *20*, 2536-2544.  
*Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription*
- [66] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204-220.  
*Establishing, maintaining and modifying DNA methylation patterns in plants and animals*
- [67] C. Sapienza, *Sci. Am.* **1990**, *263*, 52-60.  
*Parental imprinting of genes*
- [68] C. Sapienza, *Mol. Carcinog.* **1990**, *3*, 118-121.  
*Genome imprinting, cellular mosaicism and carcinogenesis*
- [69] R. D. Hotchkiss, *J Biol Chem* **1948**, *175*, 315-332.  
*The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography*
- [70] S. Augui, E. P. Nora, E. Heard, *Nat. Rev. Genet.* **2011**, *12*, 429-442.  
*Regulation of X-chromosome inactivation by the X-inactivation centre*
- [71] M. Ehrlich, *Journal of Cellular Biochemistry* **2003**, *88*, 899-910.  
*Expression of various genes is controlled by DNA methylation during mammalian development*
- [72] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *ChemBiochem* **2011**, *12*, 206-222.  
*Structure and Function of Mammalian DNA Methyltransferases*
- [73] R. Lister, *et al.*, *Nature* **2009**, *462*, 315-322.  
*Human DNA methylomes at base resolution show widespread epigenomic differences*
- [74] L. Laurent, *et al.*, *Genome Res.* **2010**, *20*, 320-331.  
*Dynamic changes in the human methylome during differentiation*
- [75] F. Lienert, *et al.*, *Nat. Genet.* **2011**, *43*, 1091-1097.  
*Identification of genetic elements that autonomously determine DNA methylation states*
- [76] A. M. Deaton, A. Bird, *Genes Dev.* **2011**, *25*, 1010-1022.  
*CpG islands and the regulation of transcription*
- [77] G. P. Pfeifer, M. Tang, M. F. Denissenko, *Curr. Top. Microbiol. Immunol.* **2000**, *249*, 1-19.  
*Mutation hotspots and DNA methylation*
- [78] J.-C. Shen, W. M. Rideout, P. A. Jones, *Nucleic Acids Res.* **1994**, *22*, 972-976.  
*The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA*
- [79] B. Hendrich, A. Bird, *Mol. Cell. Biol.* **1998**, *18*, 6538-6547.  
*Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins*
- [80] A. Kuroda, *et al.*, *PLoS One* **2009**, *4*, e6953.  
*Insulin Gene Expression Is Regulated by DNA Methylation*
- [81] J. P. Thomson, *et al.*, *Nature* **2010**, *464*, 1082-1086.  
*CpG islands influence chromatin structure via the CpG-binding protein Cfp1*
- [82] A. Portela, M. Esteller, *Nat. Biotech.* **2010**, *28*, 1057-1068.

- Epigenetic modifications and human disease*
- [83] S. Maegawa, *et al.*, *Genome Res.* **2010**, *20*, 332-340.  
*Widespread and tissue specific age-related DNA methylation changes in mice*
- [84] S. Klimasauskas, S. Kumar, R. J. Roberts, X. Cheng, *Cell* **1994**, *76*, 357-369.  
*HhaI methyltransferase flips its target base out of the DNA helix*
- [85] M. C. Cardoso, H. Leonhardt, *J. Cell Biol.* **1999**, *147*, 25-32.  
*DNA Methyltransferase Is Actively Retained in the Cytoplasm during Early Development*
- [86] L. S.-H. Chuang, *et al.*, *Science* **1997**, *277*, 1996-2000.  
*Human DNA-(Cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1*
- [87] H. P. Easwaran, L. Schermelleh, H. Leonhardt, M. C. Cardoso, *EMBO Rep.* **2004**, *5*, 1181-1186.  
*Replication-independent chromatin loading of Dnmt1 during G2 and M phases*
- [88] H. Leonhardt, A. W. Page, H.-U. Weier, T. H. Bestor, *Cell* **1992**, *71*, 865-873.  
*A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei*
- [89] M. Bostick, *et al.*, *Science* **2007**, *317*, 1760-1764.  
*UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells*
- [90] J. Sharif, *et al.*, *Nature* **2007**, *450*, 908-912.  
*The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA*
- [91] H. Hashimoto, *et al.*, *Nature* **2008**, *455*, 826-829.  
*The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix*
- [92] G. V. Avvakumov, *et al.*, *Nature* **2008**, *455*, 822-825.  
*Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1*
- [93] A. Hermann, R. Goyal, A. Jeltsch, *J. Biol. Chem.* **2004**, *279*, 48350-48359.  
*The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites*
- [94] J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, *Science* **2011**, *331*, 1036-1040.  
*Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation*
- [95] J. Cheng, *et al.*, *J. Biol. Chem.* **2013**, *288*, 1329-1339.  
*Structural Insight into Coordinated Recognition of Trimethylated Histone H3 Lysine 9 (H3K9me3) by the Plant Homeodomain (PHD) and Tandem Tudor Domain (TTD) of UHRF1 (Ubiquitin-like, Containing PHD and RING Finger Domains, 1) Protein*
- [96] N. Nady, *et al.*, *J. Biol. Chem.* **2011**, *286*, 24300-24311.  
*Recognition of Multivalent Histone States Associated with Heterochromatin by UHRF1 Protein*
- [97] M. Achour, *et al.*, *Biochem. Biophys. Res. Comm.* **2009**, *390*, 523-528.  
*UHRF1 recruits the histone acetyltransferase Tip60 and controls its expression and activity*
- [98] M. Unoki, T. Nishidate, Y. Nakamura, *Oncogene* **2004**, *23*, 7601-7610.  
*ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain*
- [99] M. Alhosin, *et al.*, *Journal of Experimental & Clinical Cancer Research* **2011**, *30*, 41.  
*Down-regulation of UHRF1, associated with re-expression of tumor suppressor genes, is a common feature of natural compounds exhibiting anti-cancer properties*
- [100] E. Citterio, *et al.*, *Mol. Cell. Biol.* **2004**, *24*, 2526-2535.  
*Np95 Is a Histone-Binding Protein Endowed with Ubiquitin Ligase Activity*
- [101] P. Bashtrykov, G. Jankevicius, R. Z. Jurkowska, S. Ragozin, A. Jeltsch, *J. Biol. Chem.* **2013**, *10.1074/jbc.M113.528893*.  
*The Uhrf1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase Dnmt1 by an allosteric mechanism*
- [102] C. Brenner, *et al.*, *EMBO J.* **2004**, *24*, 336-346.  
*Myc represses transcription through recruitment of DNA methyltransferase corepressor*
- [103] H. Li, *et al.*, *J. Biol. Chem.* **2006**, *281*, 19489-19500.  
*The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells*
- [104] J. Datta, *et al.*, *Cancer Res.* **2005**, *65*, 10891-10900.  
*Physical and Functional Interaction of DNA Methyltransferase 3A with Mbd3 and Brg1 in Mouse Lymphosarcoma Cells*
- [105] T. Chen, N. Tsujimoto, E. Li, *Mol. Cell. Biol.* **2004**, *24*, 9048-9058.  
*The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin*
- [106] Y.-Z. Ge, *et al.*, *J. Biol. Chem.* **2004**, *279*, 25447-25454.  
*Chromatin Targeting of de Novo DNA Methyltransferases by the PWWP Domain*
- [107] H. Gowher, K. Liebert, A. Hermann, G. Xu, A. Jeltsch, *J. Biol. Chem.* **2005**, *280*, 13341-13348.

- Mechanism of Stimulation of Catalytic Activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L*
- [108] M. S. Karetz, Z. M. Botello, J. J. Ennis, C. Chou, F. Chédin, *J. Biol. Chem.* **2006**, *281*, 25893-25902.
- [109] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915-926.
- [110] E. Li, C. Beard, R. Jaenisch, *Nature* **1993**, *366*, 362-365.
- [111] C. Y. Howell, *et al.*, *Cell* **2001**, *104*, 829-838.
- [112] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257.
- [113] J. Flynn, J. F. Glickman, N. O. Reich, *Biochemistry* **1996**, *35*, 7308-7315.
- [114] A. Reale, G. D. Matteis, G. Galleazzi, M. Zampieri, P. Caiafa, *Oncogene* **2005**, *24*, 13-19.
- [115] A. Bolden, C. Ward, J. A. Siedlecki, A. Weissbach, *J. Biol. Chem.* **1984**, *259*, 12437-12443.
- [116] Ž. M. Svedružić, N. O. Reich, *Biochemistry* **2005**, *44*, 14977-14988.
- [117] G. Vilkaitis, I. Suetake, S. Klimašauskas, S. Tajima, *J. Biol. Chem.* **2005**, *280*, 64-72.
- [118] R. Goyal, R. Reinhardt, A. Jeltsch, *Nucleic Acids Res.* **2006**, *34*, 1182-1188.
- [119] Z. M. Svedruzic, *Curr. Med. Chem.* **2008**, *15*, 92-106.
- [120] Y. Zhang, *et al.*, *Nucleic Acids Res.* **2010**, *38*, 4246-4253.
- [121] M. Fatemi, A. Hermann, S. Pradhan, A. Jeltsch, *J. Mol. Biol.* **2001**, *309*, 1189-1199.
- [122] A. Bacolla, S. Pradhan, R. J. Roberts, R. D. Wells, *J. Biol. Chem.* **1999**, *274*, 33011-33019.
- [123] Ž. M. Svedružić, N. O. Reich, *Biochemistry* **2005**, *44*, 9472-9485.
- [124] J. Flynn, J.-Y. Fang, J. A. Mikovits, N. O. Reich, *J. Biol. Chem.* **2003**, *278*, 8238-8243.
- [125] A. Meissner, *et al.*, *Nature* **2008**, *454*, 766-770.
- [126] F. Eckhardt, *et al.*, *Nat Genet* **2006**, *38*, 1378-1385.
- [127] Y. Zhang, *et al.*, *PLoS Genet.* **2009**, *5*, e1000438.
- [128] K. D. Robertson, *et al.*, *Nat Genet* **2000**, *25*, 338-342.
- [129] F. Fuks, W. A. Burgers, A. Brehm, L. Hughes-Davies, T. Kouzarides, *Nat Genet* **2000**, *24*, 88-91.
- [130] F. Syeda, *et al.*, *J. Biol. Chem.* **2011**, *286*, 15344-15351.
- [131] B. H. Ramsahoye, *et al.*, *P. Natl. Acad. Sci. USA* **2000**, *97*, 5237-5242.

- [132] H. Gowher, A. Jeltsch, *J. Mol. Biol.* **2001**, 309, 1201-1208.  
*Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites*
- [133] J. U. Guo, *et al.*, *Nat. Neurosci.* **2014**, 17, 215-222.  
*Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain*
- [134] V. Handa, A. Jeltsch, *J. Mol. Biol.* **2005**, 348, 1103-1112.  
*Profound Flanking Sequence Preference of Dnmt3a and Dnmt3b Mammalian DNA Methyltransferases Shape the Human Epigenome*
- [135] I. G. Lin, L. Han, A. Taghva, L. E. O'Brien, C.-L. Hsieh, *Mol. Cell. Biol.* **2002**, 22, 704-723.  
*Murine De Novo Methyltransferase Dnmt3a Demonstrates Strand Asymmetry and Site Preference in the Methylation of DNA In Vitro*
- [136] K. E. Bachman, M. R. Rountree, S. B. Baylin, *J. Biol. Chem.* **2001**, 276, 32282-32287.  
*Dnmt3a and Dnmt3b Are Transcriptional Repressors That Exhibit Unique Localization Properties to Heterochromatin*
- [137] F. Chédin, M. R. Lieber, C.-L. Hsieh, *P. Natl. Acad. Sci. USA* **2002**, 99, 16916-16921.  
*The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a*
- [138] K. Hata, M. Okano, H. Lei, E. Li, *Development* **2002**, 129, 1983-1993.  
*Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice*
- [139] D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch, X. Cheng, *Nature* **2007**, 449, 248-251.  
*Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation*
- [140] J. Oswald, *et al.*, *Curr. Biol.* **2000**, 10, 475-478.  
*Active demethylation of the paternal genome in the mouse zygote*
- [141] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, 403, 501-502.  
*Embryogenesis: Demethylation of the zygotic paternal genome*
- [142] K. Iwan, master thesis, **2014**.
- [143] J. Lee, *et al.*, *Development* **2002**, 129, 1807-1817.  
*Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells*
- [144] K. Martinowich, *et al.*, *Science* **2003**, 302, 890-893.  
*DNA Methylation-Related Chromatin Remodeling in Activity-Dependent Bdnf Gene Regulation*
- [145] S. Kangaspeska, *et al.*, *Nature* **2008**, 452, 112-115.  
*Transient cyclical methylation of promoter DNA*
- [146] R. Metivier, *et al.*, *Nature* **2008**, 452, 45-50.  
*Cyclical DNA methylation of a transcriptionally active promoter*
- [147] J.-K. Zhu, *Annu. Rev. Genet.* **2009**, 43, 143-166.  
*Active DNA Demethylation Mediated by DNA Glycosylases*
- [148] M. Dizdaroğlu, *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2005**, 591, 45-59.  
*Base-excision repair of oxidative DNA damage by DNA glycosylases*
- [149] G. Xu, M. Herzig, V. Rotrekl, C. A. Walter, *Mech. Ageing Dev.* **2008**, 129, 366-382.  
*Base excision repair, aging and health span*
- [150] A. Bürkle, C. Brabeck, J. Diefenbach, S. Beneke, *Int. J. Biochem. Cell Biol.* **2005**, 37, 1043-1053.  
*The emerging role of poly(ADP-ribose) polymerase-1 in longevity*
- [151] J. Penterman, *et al.*, *P. Natl. Acad. Sci. USA* **2007**, 104, 6752-6757.  
*DNA demethylation in the Arabidopsis genome*
- [152] M. Gehring, K. L. Bubb, S. Henikoff, *Science* **2009**, 324, 1447-1451.  
*Extensive Demethylation of Repetitive Elements During Seed Development Underlies Gene Imprinting*
- [153] M. Gehring, *et al.*, *Cell* **2006**, 124, 495-506.  
*DEMETER DNA Glycosylase Establishes MEDEA Polycomb Gene Self-Imprinting by Allele-Specific Demethylation*
- [154] A. Ortega-Galisteo, T. Morales-Ruiz, R. Ariza, T. Roldán-Arjona, *Plant Mol. Biol.* **2008**, 67, 671-681.  
*Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks*
- [155] M. Frémont, *et al.*, *Nucleic Acids Res.* **1997**, 25, 2375-2380.  
*Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA*
- [156] J.-P. Jost, *et al.*, *Nucleic Acids Res.* **1999**, 27, 3245-3252.  
*A chicken embryo protein related to the mammalian DEAD box protein p68 is tightly associated with the highly purified protein-RNA complex of 5-MeC-DNA glycosylase*
- [157] B. Zhu, *et al.*, *P. Natl. Acad. Sci. USA* **2000**, 97, 5135-5139.

- 5-Methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex*
- [158] B. Zhu, *et al.*, *Nucleic Acids Res.* **2000**, 28, 4157-4165.
- 5-Methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence*
- [159] F. Santos, W. Dean, *Reproduction* **2004**, 127, 643-651.
- Epigenetic reprogramming during early development in mammals*
- [160] C. B. Millar, *et al.*, *Science* **2002**, 297, 403-405.
- Enhanced CpG Mutability and Tumorigenesis in MBD4-Deficient Mice*
- [161] N. Navaratnam, *et al.*, *J. Biol. Chem.* **1993**, 268, 20709-20712.
- The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase*
- [162] B. Teng, C. Burant, N. Davidson, *Science* **1993**, 260, 1816-1819.
- Molecular cloning of an apolipoprotein B messenger RNA editing protein*
- [163] S. G. Conticello, *Genome Biol.* **2008**, 9, 229.
- The AID/APOBEC family of nucleic acid mutators*
- [164] H. D. Morgan, W. Dean, H. A. Coker, W. Reik, S. K. Petersen-Mahrt, *J. Biol. Chem.* **2004**, 279, 52353-52360.
- Activation-induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues: IMPLICATIONS FOR EPIGENETIC REPROGRAMMING*
- [165] C. Popp, *et al.*, *Nature* **2010**, 463, 1101-1105.
- Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency*
- [166] K. Rai, *et al.*, *Cell* **2008**, 135, 1201-1212.
- DNA Demethylation in Zebrafish Involves the Coupling of a Deaminase, a Glycosylase, and Gadd45*
- [167] P. Revy, *et al.*, *Cell* **2000**, 102, 565-575.
- Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2)*
- [168] M. Muramatsu, *et al.*, *Cell* **2000**, 102, 553-563.
- Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme*
- [169] K.-I. Hirano, *et al.*, *J. Biol. Chem.* **1996**, 271, 9887-9890.
- Targeted Disruption of the Mouse apobec-1 Gene Abolishes Apolipoprotein B mRNA Editing and Eliminates Apolipoprotein B48*
- [170] J. R. Morrison, *et al.*, *P. Natl. Acad. Sci. USA* **1996**, 93, 7154-7159.
- Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism*
- [171] Y.-Q. Li, P.-Z. Zhou, X.-D. Zheng, C. P. Walsh, G.-L. Xu, *Nucleic Acids Res.* **2007**, 35, 390-400.
- Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair*
- [172] J. Nieminuszczy, E. Grzesiuk, *Acta Biochim. Pol. (Engl. Transl.)* **2007**, 54, 459-468.
- Bacterial DNA repair genes and their eukaryotic homologues: 3. AlkB dioxygenase and Ada methyltransferase in the direct repair of alkylated DNA*
- [173] Z. Yu, *et al.*, *Nucleic Acids Res.* **2007**, 35, 2107-2115.
- The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase*
- [174] S. Ito, *et al.*, *Nature* **2010**, 466, 1129-1133.
- Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification*
- [175] K. P. Koh, *et al.*, *Cell Stem Cell* **2011**, 8, 200-213.
- Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells*
- [176] Y. Wang, Y. Zhang, *Cell Rep.* **2014**, 6, 278-284.
- Regulation of TET Protein Stability by Calpains*
- [177] Su J. Song, *et al.*, *Cell Stem Cell* **2013**, 13, 87-101.
- The Oncogenic MicroRNA miR-22 Targets the TET2 Tumor Suppressor to Promote Hematopoietic Stem Cell Self-Renewal and Transformation*
- [178] Cornelia G. Spruijt, *et al.*, *Cell* **2013**, 152, 1146-1159.
- Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives*
- [179] E. A. Minor, B. L. Court, J. I. Young, G. Wang, *J. Biol. Chem.* **2013**, 288, 13669-13674.
- Ascorbate Induces Ten-Eleven Translocation (Tet) Methylcytosine Dioxygenase-mediated Generation of 5-Hydroxymethylcytosine*
- [180] M. Ko, *et al.*, *Nature* **2013**, 497, 122-126.
- Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX*
- [181] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, *Cell Cycle* **2009**, 8, 1698-1710.

- Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids*
- [182] H. F. Jorgensen, I. Ben-Porath, A. P. Bird, *Mol. Cell. Biol.* **2004**, *24*, 3387-3395.  
*Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains*
- [183] J. H. Lee, K. S. Voo, D. G. Skalnik, *J. Biol. Chem.* **2001**, *276*, 44669-44676.  
*Identification and characterization of the DNA binding domain of CpG-binding protein*
- [184] N. P. Blackledge, *et al.*, *Mol. Cell* **2010**, *38*, 179-190.  
*CpG Islands Recruit a Histone H3 Lysine 36 Demethylase*
- [185] M. Birke, *et al.*, *Nucleic Acids Res.* **2002**, *30*, 958-965.  
*The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation*
- [186] C. Frauer, *et al.*, *PLoS One* **2011**, *6*, e16627.  
*Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1*
- [187] Y. Xu, *et al.*, *Cell* **2012**, *151*, 1200-1213.  
*Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development*
- [188] Y. Huang, *et al.*, *P. Natl. Acad. Sci. USA* **2014**, *111*, 1361-1366.  
*Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells*
- [189] K. Williams, *et al.*, *Nature* **2011**, *473*, 343-348.  
*TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity*
- [190] G. Ficiz, *et al.*, *Nature* **2011**, *473*, 398-402.  
*Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation*
- [191] R. Deplus, *et al.*, *EMBO J.* **2013**, *32*, 645-655.  
*TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS*
- [192] P. Vella, *et al.*, *Mol. Cell* **2013**, *49*, 645-656.  
*Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells*
- [193] Q. Chen, Y. Chen, C. Bian, R. Fujiki, X. Yu, *Nature* **2013**, *493*, 561-564.  
*TET2 promotes histone O-GlcNAcylation during gene transcription*
- [194] L. Hu, *et al.*, *Cell* **2013**, *155*, 1545-1555.  
*Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation*
- [195] C. Loenarz, C. J. Schofield, *Nat. Chem. Biol.* **2008**, *4*, 152-156.  
*Expanding chemical biology of 2-oxoglutarate oxygenases*
- [196] E. G. Kovaleva, J. D. Lipscomb, *Nat. Chem. Biol.* **2008**, *4*, 186-193.  
*Versatility of biological non-heme Fe(II) centers in oxygen activation reactions*
- [197] V. K. C. Ponnaluri, J. P. Maciejewski, M. Mukherji, *Biochem. Biophys. Res. Comm.* **2013**, *436*, 115-120.  
*A mechanistic overview of TET-mediated 5-methylcytosine oxidation*
- [198] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781-790.  
*The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid*
- [199] G. R. Wyatt, S. S. Cohen, *Nature* **1952**, *170*, 1072-1073.  
*A new pyrimidine base from bacteriophage nucleic acids*
- [200] E. Privat, L. C. Sowers, *Chem. Res. Toxicol.* **1996**, *9*, 745-750.  
*Photochemical Deamination and Demethylation of 5-Methylcytosine*
- [201] T. P. Gu, *et al.*, *Nature* **2011**, *477*, 606-610.  
*The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes*
- [202] M. Munzel, *et al.*, *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 5375-5377.  
*Quantification of the sixth DNA base hydroxymethylcytosine in the brain*
- [203] K. E. Szulwach, *et al.*, *Nat. Neurosci.* **2011**, *14*, 1607-1616.  
*5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging*
- [204] E. Dong, D. P. Gavin, Y. Chen, J. Davis, *Transl. Psychiatry* **2012**, *2*, e159.  
*Upregulation of TET1 and downregulation of APOBEC3A and APOBEC3C in the parietal cortex of psychotic patients*
- [205] H. Wu, *et al.*, *Genes Dev.* **2011**, *25*, 679-684.  
*Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells*
- [206] C.-X. Song, *et al.*, *Nat. Biotech.* **2011**, *29*, 68-72.  
*Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine*
- [207] V. Valinluck, *et al.*, *Nucleic Acids Res.* **2004**, *32*, 4100-4108.  
*Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2)*

- [208] M. J. Booth, *et al.*, *Nat. Protocols* **2013**, 8, 1841-1851.  
*Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine*
- [209] M. Yu, *et al.*, *Nat. Protocols* **2012**, 7, 2159-2170.  
*Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine*
- [210] C. Frauer, *et al.*, *PLoS One* **2011**, 6, e21306.  
*Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain*
- [211] H. Hashimoto, *et al.*, *Nucleic Acids Res.* **2012**, 10.1093/nar/gks155.  
*Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation*
- [212] A. Inoue, Y. Zhang, *Science* **2011**, 334, 194.  
*Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos*
- [213] J. U. Guo, Y. Su, C. Zhong, G. L. Ming, H. Song, *Cell* **2011**, 145, 423-434.  
*Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain*
- [214] Y. F. He, *et al.*, *Science* **2011**, 333, 1303-1307.  
*Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA*
- [215] C. J. Burrows, *ACS Symp. Ser. Am. Chem. Soc.* **2009**, 2009, 147-156.  
*Surviving an Oxygen Atmosphere: DNA Damage and Repair*
- [216] M. Iurlaro, *et al.*, *Genome Biol.* **2013**, 14, R119.  
*A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation*
- [217] M. W. Kellinger, *et al.*, *Nat. Struct. Mol. Biol.* **2012**, 19, 831-833.  
*5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription*
- [218] A. Alioui, *et al.*, *Nucleus* **2012**, 3, 565-569.  
*5-Carboxylcytosine is localized to euchromatic regions in the nuclei of follicular cells in axolotl ovary*
- [219] X. Lu, *et al.*, *JACS* **2013**, 135, 9315-9317.  
*Chemical Modification-Assisted Bisulfite Sequencing (CAB-Seq) for 5-Carboxylcytosine Detection in DNA*
- [220] C.-X. Song, *et al.*, *Cell* **2013**, 153, 678-691.  
*Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming*
- [221] L. Jin, W. Wang, D. Hu, J. Lu, *Phys. Chem. Chem. Phys.* **2014**, 16, 3573-3585.  
*A new insight into the 5-carboxycytosine and 5-formylcytosine under typical bisulfite conditions: a deamination mechanism study*
- [222] A. Inoue, L. Shen, Q. Dai, C. He, Y. Zhang, *Cell Res.* **2011**, 21, 1670-1676.  
*Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development*
- [223] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, 286, 35334-35338.  
*Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES*
- [224] S. Moréra, *et al.*, *Nucleic Acids Res.* **2012**, 40, 9917-9926.  
*Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyluracil-containing DNA*
- [225] H. Hashimoto, S. Hong, A. S. Bhagwat, X. Zhang, X. Cheng, *Nucleic Acids Res.* **2012**, 40, 10203-10214.  
*Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation*
- [226] L. Zhang, *et al.*, *Nat. Chem. Biol.* **2012**, 8, 328-330.  
*Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA*
- [227] P. J. Berti, J. A. B. McCann, *Chem. Rev. (Washington, DC, U. S.)* **2006**, 106, 506-555.  
*Toward a Detailed Understanding of Base Excision Repair Enzymes: Transition State and Mechanistic Analyses of N-Glycoside Hydrolysis and N-Glycoside Transfer*
- [228] L. Zhang, *et al.*, *Nat. Chem. Biol.* **2012**, 8, 328-330.  
*Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA*
- [229] A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, A. C. Drohat, *JACS* **2013**, 135, 15813-15822.  
*Divergent Mechanisms for Enzymatic Excision of 5-Formylcytosine and 5-Carboxylcytosine from DNA*
- [230] H. Hashimoto, X. Zhang, X. Cheng, *DNA Repair* **2013**, 12, 535-540.  
*Activity and crystal structure of human thymine DNA glycosylase mutant N140A with 5-carboxylcytosine DNA at low pH*
- [231] R. J. Boorstein, *et al.*, *J. Biol. Chem.* **2001**, 276, 41991-41997.  
*Definitive Identification of Mammalian 5-Hydroxymethyluracil DNAN-Glycosylase Activity as SMUG1*

- [232] S. Cortellino, *et al.*, *Cell* **2011**, *146*, 67-79.  
*Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair*
- [233] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472-479.  
*TET enzymes, TDG and the dynamics of DNA demethylation*
- [234] J. Nakamura, J. A. Swenberg, *Cancer Res.* **1999**, *59*, 2522-2526.  
*Endogenous Apurinic/Apyrimidinic Sites in Genomic DNA of Mammalian Tissues*
- [235] Z. Liutkeviciute, G. Lukinavicius, V. Masevicius, D. Daujotyte, S. Klimasauskas, *Nat. Chem. Biol.* **2009**, *5*, 400-402.  
*Cytosine-5-methyltransferases add aldehydes to DNA*
- [236] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, *J. Biol. Chem.* **2012**, *287*, 33116-33121.  
*The Mammalian de Novo DNA Methyltransferases DNMT3A and DNMT3B Are Also DNA 5-Hydroxymethylcytosine Dehydroxymethylases*
- [237] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, *J. Biol. Chem.* **2013**, 10.1074/jbc.M112.445585.  
*DNA 5-Methylcytosine Demethylation Activities of the Mammalian DNA Methyltransferases*
- [238] Z. Liutkevičiūtė, *et al.*, *JACS* **2014**, *136*, 5884-5887.  
*Direct Decarboxylation of 5-Carboxylcytosine by DNA C5- Methyltransferases*
- [239] S. Schiesser, *et al.*, *JACS* **2013**, *135*, 14593-14599.  
*Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine*
- [240] T. Li, L. Huo, C. Pulley, A. Liu, *Bioorg. Chem.* **2012**, *43*, 2-14.  
*Decarboxylation mechanisms in biological system*
- [241] A. Radzicka, R. Wolfenden, *Science* **1995**, *267*, 90-93.  
*A proficient enzyme*
- [242] S. Xu, *et al.*, *Cell Res.* **2013**, *23*, 1296-1309.  
*Crystal structures of isoortate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase*

## V. The discovery of 5-formylcytosine in embryonic stem cell DNA

T. Pfaffeneder\*, **B. Hackner\***, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeyer, T. Carell, *Angewandte Chemie International Edition* **2012**, 50, 7008-7012. (\* Equal contribution)

### **Introduction:**

As described in Chapter I.2.3.3 the discovery of hmC as a product of the TET enzymes had a huge impact in the field of active demethylation. The proposed cycle of active demethylation however included the existence of the higher oxidized cytidine derivatives, fC and caC. These intermediates could not be identified in mouse tissue in an earlier study, despite the high levels of the first putative intermediate hmC. As DNA demethylation was however thought to take place especially in early cell development, the search was successfully extended to mESCs in this manuscript.

### **Declaration of contribution:**

Based on the LC-MS method established by *Dr. Daniel Globisch* using nucleoside standards prepared by *Dr. Martin Münzel*, I analyzed in tight collaboration with *Toni Pfaffeneder* the genomic DNA of mESCs and knock out cells, provided by *Dr. Matthias Truss*. Together we discovered and characterized fC using MS/MS and chemical modifications. We also determined the values of mC and hmC, and estimated the amount of fC by a coinjection study.

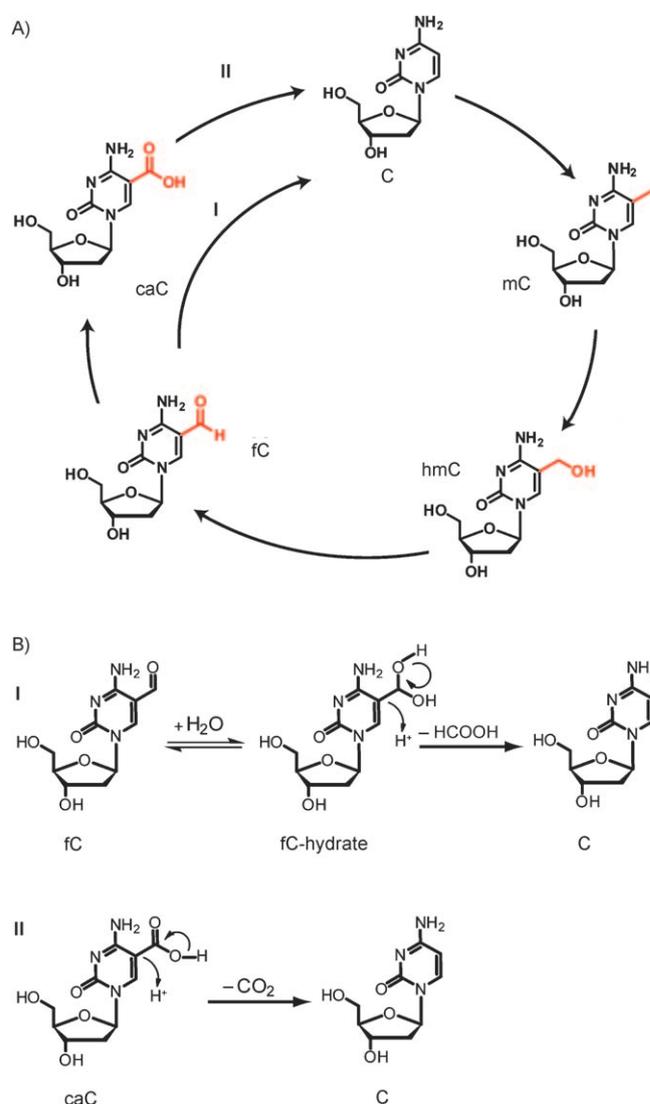
## The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA\*\*

Toni Pfaffeneder, Benjamin Hackner, Matthias Truß,\* Martin Münzel, Markus Müller, Christian A. Deiml, Christian Hagemeyer, and Thomas Carell\*

Cellular development requires the silencing and activation of specific gene sequences in a well-orchestrated fashion. Transcriptional gene silencing is associated with the clustered methylation of cytosine bases (C) in CpG units of promoters. The methylation occurs at position C5 of cytosine to give 5-methylcytosine (mC) with the help of special DNA methyltransferases (DNMT).<sup>[1]</sup> The DNA methylome is significantly reprogrammed at various stages during early development,<sup>[2]</sup> during the development of primordial germ cells,<sup>[2c,3]</sup> or later in a locus-specific way at postdevelopmental stages.<sup>[4]</sup> Decreasing levels of mC can be established passively by successive rounds of DNA replication in the absence of methyltransferases. Active demethylation, in contrast, is proposed to be a process in which the mC bases are directly converted back into unmodified cytosines in the genome.<sup>[5]</sup> The recent discovery that mC can be further oxidized to hydroxymethylcytosine (hmC) with the help of TET enzymes<sup>[6]</sup> has led to the idea that hmC is connected to epigenetic reprogramming,<sup>[7]</sup> maybe as an intermediate in an, as yet controversial, active demethylation process.<sup>[4,5,8]</sup> Indeed recent data suggest that active demethylation in postdevelopmental phases may proceed through deamination of hmC to give 5-hydroxymethyluridine (hmU), which is then removed from the genome with the help of the base excision repair (BER) system.<sup>[9]</sup> Chemically, an attractive alternative mechanism for a more global active demethylation could be envisioned through further oxidation of hmC to give either 5-formylcytosine (fC) or 5-carboxylcytosine (caC) followed by elimination of a formyl or carboxyl group, respectively

(Scheme 1).<sup>[5a,10]</sup> Although such an oxidative active demethylation pathway with hmC as the starting point has been frequently postulated,<sup>[5a,10]</sup> none of the further oxidized bases (fC, caC) have so far been detected.<sup>[10a]</sup>

To examine the question of whether hmC is the only oxidized base present in genomic DNA or if other higher oxidized species may be present as well, we performed an HPLC-MS study using synthetic fC and caC material as



**Scheme 1.** A) Putative cycle of methylation and oxidative demethylation of cytidine derivatives. B) Details of the demethylation reaction via vinyl carbanions.<sup>[11]</sup> I: Deformylation of fC to C. II: decarboxylation of caC to C.

[\*] M. Sc. T. Pfaffeneder,<sup>[†]</sup> M. Sc. B. Hackner,<sup>[†]</sup> Dipl.-Chem. M. Münzel, Dr. M. Müller, Dipl.-Biochem. C. A. Deiml, Prof. Dr. T. Carell CIPSM, Fakultät für Chemie und Pharmazie Ludwig-Maximilians-Universität München Butenandtstrasse 5–13, 81377 München (Germany) E-mail: thomas.carell@lmu.de Homepage: <http://www.carellgroup.de>

Dr. M. Truß,<sup>[†]</sup> Prof. Dr. C. Hagemeyer Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder- und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie Ziegelstrasse 5–9, 10098 Berlin (Germany) E-mail: matthias.truss@charite.de

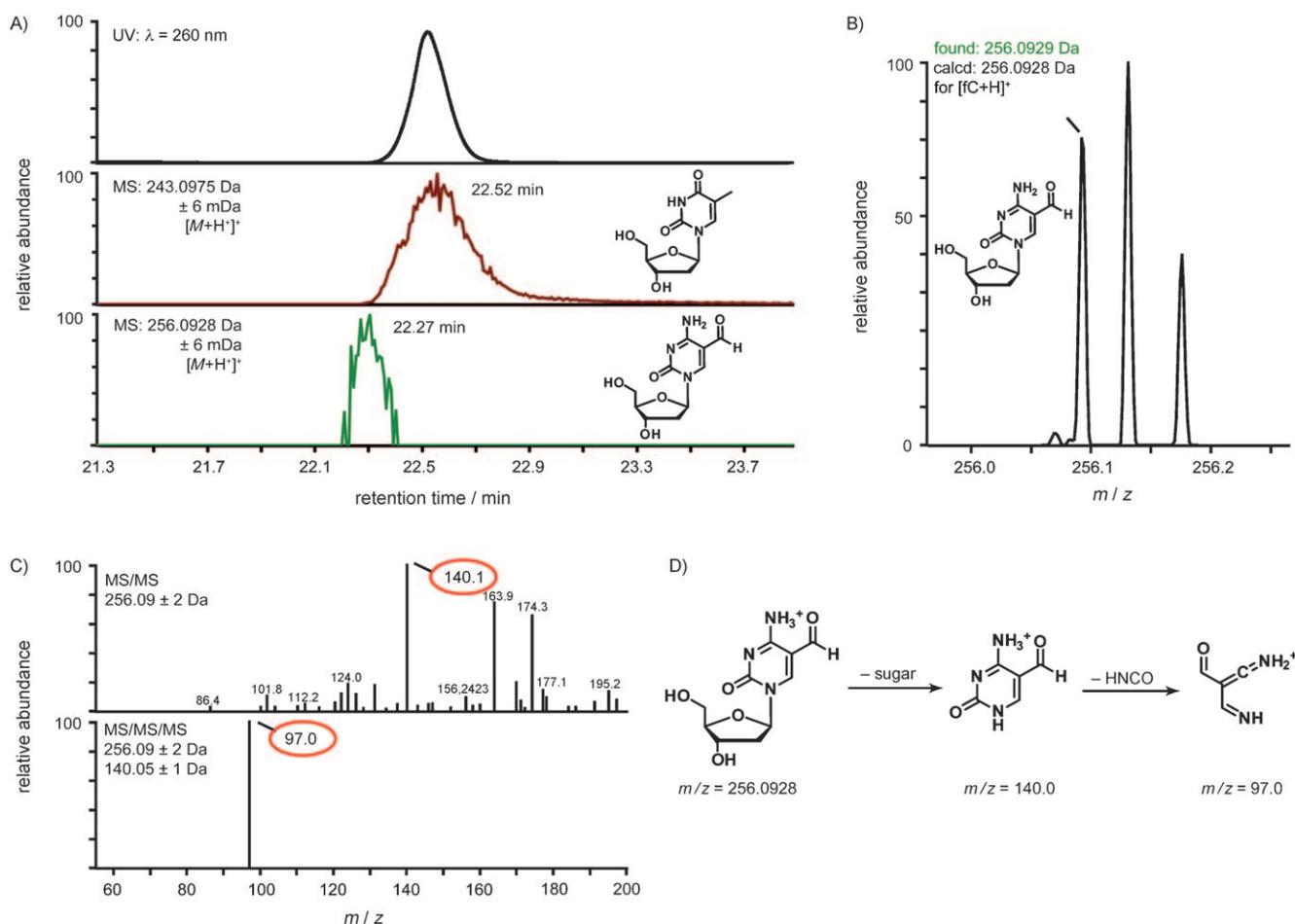
[†] These authors contributed equally to this work.

[\*\*] We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support through the SFB program (SFB 749, 646, and TRR54) and by grant CA275 8/4. Further support was obtained from the Excellence Cluster (Center for Integrative Protein Science, CIPSM) and NGFNplus (01GS0870).

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201103899>.

standards.<sup>[10a,12]</sup> Specifically, we used DNA isolated from mouse embryonic stem cells and embryoid bodies (mEBs) after two and three days of differentiation for the study. In mES cells, CpG-rich gene promoters are actively maintained in a hypomethylated state and differentiation of mES cells induces a wave of gene-specific de novo methylation that involves repression of TET1 and TET2 expression as well as reduction of global hmC and an increase in the global mC levels.<sup>[7d]</sup> In addition, we analyzed DNA isolated from nerve cell tissue, which features the highest levels of hmC in mice. Indeed, hippocampus and cortex tissue have hmC levels of up to approximately 0.7%/G,<sup>[12a]</sup> which is about twofold higher than the hmC values measured in mES cells (0.39%/G). The chromatogram obtained of a fully digested DNA sample from mES cells shows the signals for the four canonical bases A, C, G, and T plus the signal for mC (Figure 3A).<sup>[10a,12a]</sup> If detection is performed by mass spectrometry, the correct high-resolution mass ( $m/z$  values) for these five compounds and additionally the mass signal for hmC can be clearly detected. To our surprise, we detected in addition to these six signals one more signal of a compound that eluted with a retention time very close to T (Figure 1A). This signal was initially only detectable in the DNA material isolated from mES cells. The

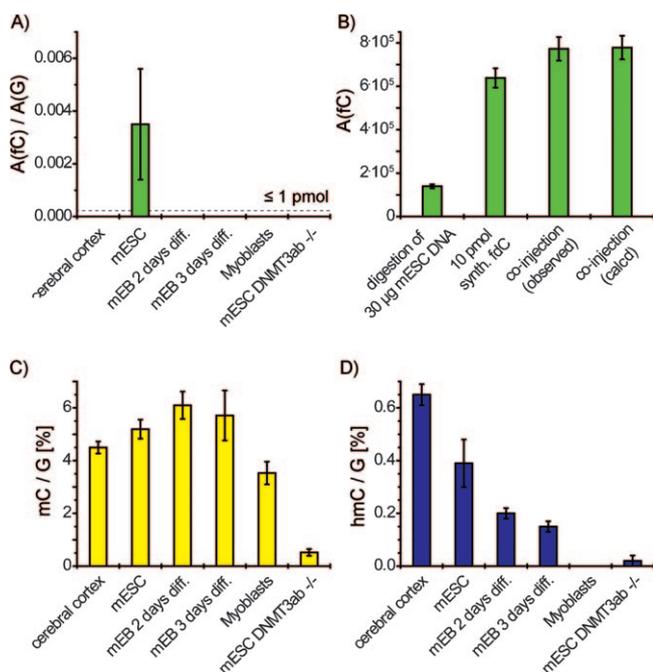
new signal could not be directly detected in DNA isolated from mEBs, but unknown fragment ions were seen in further MS<sup>n</sup> studies. The unknown compound detected in the mES cell DNA had a high-resolution mass signal ( $m/z_{\text{found}} = 256.0929$ ) which is in excellent agreement with the calculated exact mass for fC ( $m/z_{\text{calcd}} = 256.0928$ ; Figure 1B). To unequivocally prove that the signal is generated by the presence of fC we synthesized the fC compound independently, as reported previously by us,<sup>[10a]</sup> and co-injected a small amount of the synthetic material into the DNA digest obtained from the mES cell DNA. Indeed, the synthetic material eluted with the same retention time (see the Supporting Information). Finally MS<sup>n</sup>-fragmentation experiments were performed, which are highly compound specific. In these studies, the fC compound is fragmented directly in the mass spectrometer to give characteristic fragment ions. (Figure 1C,D) The MS/MS data obtained from the putative fC compound isolated from mES cell DNA were found to be identical with literature data<sup>[13]</sup> and with the MS/MS data obtained from the authentic synthetic fC material (data not shown). In addition, the obtained MS<sup>n</sup> data were identical with the unknown fragment ions detected in the mEB cell DNA. These data prove that the newly discovered compound



**Figure 1.** A) HPLC trace of digested mES cell DNA together with the MS signals from T and fC. The UV detection has a general delay of 0.2 min and was adjusted manually to the ion current. B) High-resolution mass data of fC. C) Fragment mass data from MS/MS and MS<sup>3</sup> studies proving the structure of fC. D) Fragmentation pattern of fC in the MS<sup>2</sup> and MS<sup>3</sup> experiments.

in mES cell DNA has the structure of fC. This base is present at significant levels in mES cells and in traces in mEB cell DNA.

We next quantified the amount of fC base in the mES sample (Figure 2A). To this end we co-injected a defined amount of synthetic fC together with digested mES cell DNA and integrated the ion currents of the combined fC signal. In



**Figure 2.** A) Correlation of the mass signal of fC and the UV signal of G in mES cell DNA, in DNA from nerve tissue, in DNA from mEBs after 2 and 3 days, in DNA from cultured myoblasts, and in DNA from mES lacking DNMT3a and -3b. The detection limit of fC was determined to be  $\leq 1$  pmol. B) Co-injection studies of fC with digested embryonic stem cell DNA indicated an amount of approx. 2 pmol, which corresponds to a level of 0.02%/G. C) Quantitative levels of mC in mES cell DNA, in DNA from nerve tissue, in DNA from mEBs after 2 and 3 days, in DNA from cultured myoblasts, and in DNA from mES lacking DNMT3a and -3b measured by quantitative mass spectrometry using an isotopically labeled mC standard.<sup>[10a, 12a]</sup> D) Quantitative levels of hmC in mES cell DNA, in DNA from nerve tissue, in DNA from mEBs after 2 and 3 days, in DNA from cultured myoblasts, and in DNA from mES lacking DNMT3a and -3b measured by quantitative mass spectrometry using an isotopically labeled hmC standard.<sup>[10a, 12a]</sup> Green: fC, yellow: mC, blue: hmC.

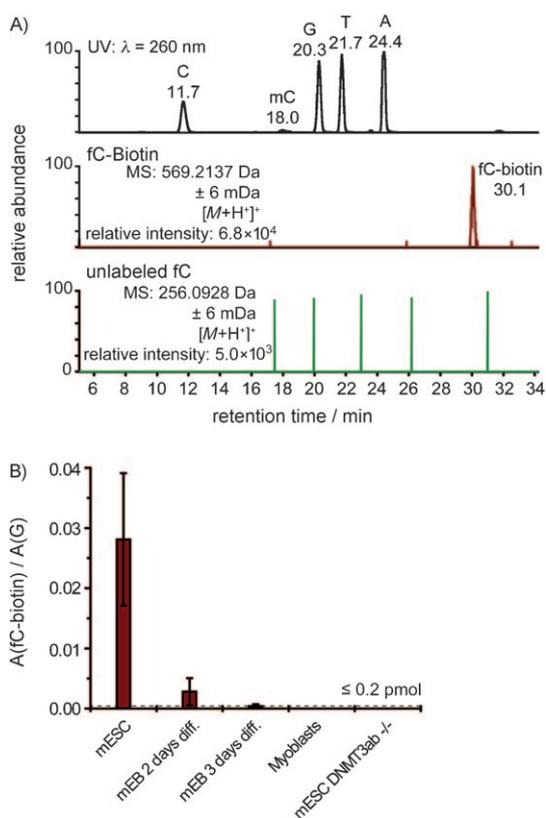
addition, we measured the fC ion current of the added amount of synthetic fC alone and compared the values to the fC ion current measured in the mES sample. The integrals allowed us to estimate the fC level to be around 0.02%/G (Figure 2B). This is a surprisingly high value compared with the mC levels (Figure 2C) as well as with the hmC levels, which we determined by using isotope spiking to be around 0.39% hmC/G (Figure 2D).<sup>[10a, 12a]</sup> Consequently, every 10th to 20th hmC base in the mES DNA is oxidized to fC, which shows that this oxidation is a significant process.

We next performed mass spectrometry experiments to study the presence of the hydrate form of fC (fC-hydrate)

since further oxidation of fC to caC or elimination of a formyl group would require the addition of water to the formyl group (Scheme 1A). We indeed detected the hydrate form in DNA containing synthetic fC at a level of about 0.5% (retention time of the hydrate = 21.7 and 26.5 min;  $m/z_{\text{found}} = 274.1037$ ,  $m/z_{\text{calcd}} = 274.1034$ ; see the Supporting Information) which is high enough to enable either direct elimination of a formyl group (Scheme 1) or further oxidation. To investigate the presence of the further oxidized compound caC in the DNA samples from mES and mEB cells, we performed additional MS studies that were also extended to a search for the hmC- and fC-derived deamination products hmU and fU. However, signals corresponding to caC, hmU, and fU were not found. In additional MS<sup>n</sup> experiments, fragmentation products characteristic for all these compounds were also not detected, which shows that if these compounds are present, their levels are below our detection limit.

We performed two more experiments to correlate the fC levels with the mC and hmC levels. Firstly we quantified the hmC levels in various DNA samples and secondly we studied the levels of mC, hmC, and fC in DNMT3a/3b double knock-out mES cells. Since the DNMT enzymes are needed to methylate C to mC, we hoped to learn if the newly discovered base fC is generated de novo from C (by a formylation reaction) or whether it is created from mC by stepwise oxidation via hmC. In agreement with earlier studies, we observed the highest hmC levels in nerve cell DNA (hmC/G = 0.65%). The levels of hmC in mES cell DNA are significantly lower and they decrease with differentiation (in accordance with the results of Szwagierczak et al.<sup>[14]</sup>), while the mC levels increase. A level of hmC/G = 0.39% was measured in mES cell DNA. Here also, the fC level is the highest with fC/G = 0.02%. In mEB cells, the hmC levels are hmC/G = 0.2% after 2 days of differentiation and hmC/G = 0.15% after 3 days of differentiation. The fC compound was only detected in trace amounts in both sets of mEB cells by MS<sup>n</sup> studies. A stronger difference can be seen in the DNMT3a/3b double knock-out cells. Here, the mC levels are greatly reduced to mC/G = 0.5%, compared to 5–6% determined in the mES and mEB cells. The mC level in cultured C2C12 myoblasts is slightly lower, in agreement with other cell lines,<sup>[12a]</sup> but still at around 4%, thus showing that C methylation is, as expected, strongly hindered in the DNMT double knock-out mES cells. The same trend can be observed when studying the hmC levels. The DNMT3a/3b -/- mES cells contain practically no hmC (0.02%) and also no fC. These experiments show that fC is likely produced from mC via hmC through further oxidation. We currently speculate that the TET enzymes may convert mC into fC by iterative oxidation, as it is found for related T7H enzymes that catalyze the stepwise oxidation of thymine to 5-formyluracil.<sup>[15]</sup>

Since the formyl group of the fC compound was shown to be reactive (hydrate formation with water), we next investigated the possibility of reacting the fC base in DNA with a reagent that would allow isolation<sup>[16]</sup> of fC-containing DNA fragments for sequencing<sup>[17]</sup> and more-sensitive detection of fC. Since formyl groups react selectively with hydroxylamines to give stable oxime derivatives,<sup>[18]</sup> we treated the mES DNA with the biotin-hydroxylamine reagent (Figure 3A).<sup>[19]</sup> After



**Figure 3.** A) HPLC and MS signals of biotin-labeled fC obtained after treatment of embryonic stem cell DNA with the biotin hydroxylamine followed by DNA digest. The lowest trace shows only the background noise, specific signals for residual fC were not observed. B) Relative amount (mass area of fC-biotin/dG) of fC in mES cell DNA, DNA from mEBs after 2 and 3 days, DNA from cultured myoblasts, and in DNA from mES containing a double knock-out in DNMT3a/3b. The detection limit of fC-biotin using quantitative mass spectrometry is  $\leq 0.2$  pmol, and thus five times lower than that of unlabeled fC.

24 h of incubation (pH 5.5, 25°C, *p*-methoxyaniline/NaOAc buffer),<sup>[20]</sup> the converted DNA was isolated and fully digested (see the Supporting Information). In a parallel experiment, we also added the biotin reagent to synthetic DNA in which the fC compound was synthetically incorporated by using a newly developed phosphoramidite building block (see the Supporting Information and for alternative synthetic strategies<sup>[21]</sup>). MS analysis of both digests showed the appearance of only a single new MS signal derived from the biotin-labeled fC nucleoside (fC-biotin). To our surprise, we no longer observed a signal for fC, thus showing that the reaction allows not only the highly selective but also complete (>90%) labeling of the fC nucleobase in genomic mES material (Figure 3A). Most importantly, the fC-biotin derivative produced a strongly increased MS signal, which allowed us to study the presence of fC in the mEB cell DNA in more detail. Indeed, after derivatization we could detect signals for the biotinylated fC base in mEB DNA (Figure 3B) directly in the mass spectrometer.

In summary, we provide here direct evidence for the presence of 5-formylcytosine (fC) in DNA isolated from mES und mEB cells. The fC levels were found to dramatically

decrease with ongoing differentiation. Interestingly, we do not detect the fC compound in DNA isolated from neuronal cells, which contain the highest amounts of hmC. We explain this result on the basis of data from a recent study by Song and co-workers,<sup>[9]</sup> who showed that active demethylation in adult brain cells proceeds likely through deamination of hmC to hmU followed by removal of the hmU base by the base excision repair pathway. Thus, fC is in this respect a clear marker nucleoside for the development of mES cells. It has not escaped our notice that the oxidative demethylation of methylcytosine via 5-formylcytosine we have postulated, immediately suggests a possible globally acting epigenetic control mechanism.

Received: June 8, 2011

Published online: June 30, 2011

**Keywords:** active demethylation · epigenetics · 5-formylcytosine · 5-hydroxymethylcytosine · mass spectrometry

- [1] a) R. Bonasio, S. J. Tu, D. Reinberg, *Science* **2010**, *330*, 612–616; b) P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068–1070; c) J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204–220.
- [2] a) K. Iqbal, S. G. Jin, G. P. Pfeifer, P. E. Szabo, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3642–3647; b) 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; c) P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter, M. A. Surani, *Mech. Dev.* **2002**, *117*, 15–23.
- [3] P. Hajkova, S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, M. A. Surani, *Science* **2010**, *329*, 78–82.
- [4] a) C. A. Miller, J. D. Sweatt, *Neuron* **2007**, *53*, 857–869; b) J. J. Day, J. D. Sweatt, *Nat. Neurosci.* **2010**, *13*, 1319–1323.
- [5] a) S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607–620; b) J.-K. Zhu, *Annu. Rev. Genet.* **2009**, *43*, 143–166.
- [6] a) S. Kiaucionis, 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; c) M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, DOI: 10.1002/anie.201101547.
- [7] a) G. Ficiz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398–402; b) S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129–1133; c) W. A. Pastor, et al., *Nature* **2011**, *473*, 394–397; d) K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343–348; e) 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; f) K. P. Koh, et al., *Cell Stem Cell* **2011**, *8*, 200–213; g) J. Walter, *Cell Stem Cell* **2011**, *8*, 121–122; h) Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, I. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Mol. Cell* **2011**, *42*, 451–461.
- [8] a) S. H. Feng, S. E. Jacobsen, W. Reik, *Science* **2010**, *330*, 622–627; b) W. Reik, W. Dean, J. Walter, *Science* **2001**, *293*, 1089–1093.
- [9] J. U. Guo, Y. Su, C. Zhong, G.-I. Ming, H. Song, *Cell* **2011**, *145*, 423–434.
- [10] a) D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367; b) C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580–583.

- [11] T. L. Arnyes, B. M. Wood, K. Chan, J. A. Gerlt, J. P. Richard, *J. Am. Chem. Soc.* **2008**, *130*, 1574–1575.
- [12] a) M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmler, S. Michalak, M. Müller, M. Biel, T. Carell, *Angew. Chem.* **2010**, *122*, 5503–5505; *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377; b) T. Le, K.-P. Kim, G. Fan, K. F. Faull, *Anal. Biochem.* **2011**, *412*, 203–209; c) M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671–5673.
- [13] H. Cao, Y. Wang, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1335–1341.
- [14] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res.* **2010**, *38*, e181.
- [15] a) C.-K. Liu, C.-A. Hsu, M. T. Abbott, *Arch. Biochem. Biophys.* **1973**, *159*, 180–187; for reviews about T7H enzymes, see b) J. M. Simmons, T. A. Muller, R. P. Hausinger, *Dalton Trans.* **2008**, 5132–5142; c) J. M. Simmons, D. J. Koslowsky, R. P. Hausinger, *Exp. Parasitol.* **2010**, *124*, 453–458.
- [16] C.-X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.-H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68–72.
- [17] S.-G. Jin, S. Kadam, G. P. Pfeifer, *Nucleic Acids Res.* **2010**, *38*, e125.
- [18] a) L. M. Hough, F. L. Oswald, *Annu. Rev. Psychol.* **2000**, *51*, 631–664; b) V. Raindlová, R. Pohl, M. Sanda, M. Hocek, *Angew. Chem.* **2010**, *122*, 1082–1084; *Angew. Chem. Int. Ed.* **2010**, *49*, 1064–1066.
- [19] a) J. Nakamura, V. E. Walker, P. B. Upton, S.-Y. Chiang, Y. W. Kow, J. A. Swenberg, *Cancer Res.* **1998**, *58*, 222–225; b) K. Kubo, H. Ide, S. S. Wallace, Y. W. Kow, *Biochemistry* **1992**, *31*, 3703–3708.
- [20] A. Dirksen, T. M. Hackeng, P. E. Dawson, *Angew. Chem.* **2006**, *118*, 7743–7746; *Angew. Chem. Int. Ed.* **2006**, *45*, 7581–7584.
- [21] a) N. Karino, Y. Ueno, A. Matsuda, *Nucleic Acids Res.* **2001**, *29*, 2456–2463; b) Q. Dai, C. He, *Org. Lett.* **2011**, DOI: 10.1021/ol201189n.

# VI. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing

S. Schiesser\*, **B. Hackner\***, T. Pfaffeneder, M. Muller, C. Hagemeyer, M. Truss, T. Carell,  
*Angewandte Chemie International Edition* **2012**, *51*, 6516-6520. (\* Equal contribution)

## **Introduction:**

As described in chapter I.2.3 several pathways for active demethylation have been proposed. A very attractive pathway is the decarboxylation of caC to C which avoids the formation of abasic sites that occurs in the BER pathway and is explained in detail in chapter I.2.3.3.5. In this manuscript hints for such a decarboxylation were found and chemical properties were elucidated using LC-MS.

## **Declaration of contribution:**

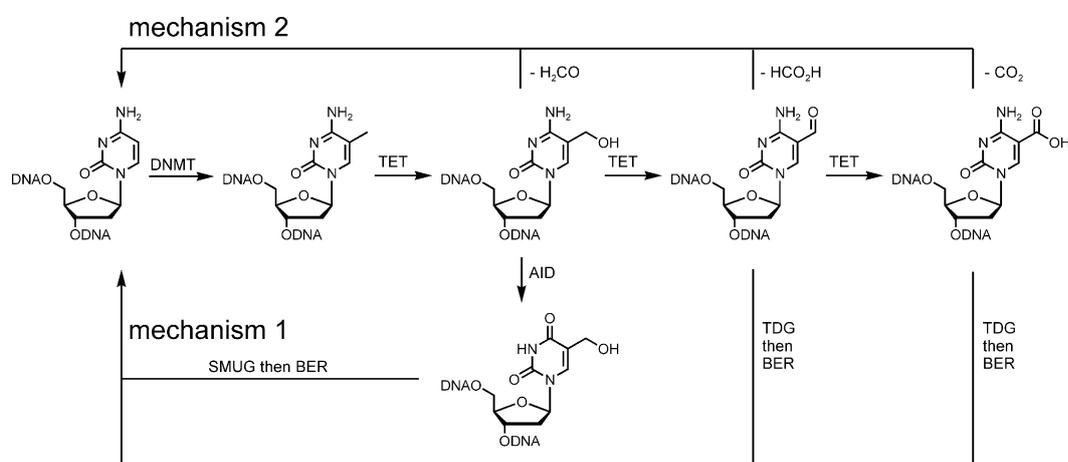
I “clicked” and purified the substrate oligonucleotide for the in vitro assay using stem cell nuclear extract. I established the *in vitro* assay (including the generation of nuclear extracts, isolation of oligonucleotides) and synthesized and purified  $^{15}\text{N}_2$ -labeled dC that was used as a standard for the LC-MS method established by *Toni Pfaffeneder*. The authentic material was also used for the chemical experiments performed by *Stefan Schiesser*.

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

Stefan Schiesser, Benjamin Hackner, Toni Pfaffeneder, Markus Müller, Christian Hagemeyer, 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.<sup>[1–3]</sup> 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)<sup>[4,5]</sup> and further to 5-formylcytosine (fC)<sup>[6]</sup> and 5-carboxycytosine (caC)<sup>[7,8]</sup> (Scheme 1).<sup>[9,10]</sup> 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.<sup>[11]</sup> Two possible active demethylation mechanisms are currently being intensively investigated (Scheme 1). Mechanism 1 is based on

special glycosylases,<sup>[12–15]</sup> 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).<sup>[15,16]</sup> The latter is formed after deamination of hmC.<sup>[15]</sup> 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,<sup>[†]</sup> B. Hackner,<sup>[†]</sup> 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. Hagemeyer, 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<sup>M</sup> 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>.

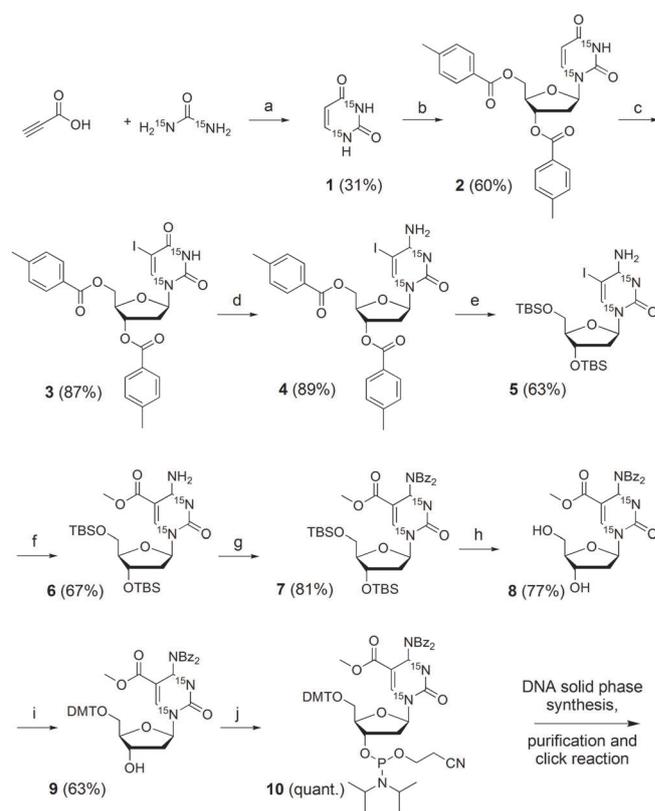
starting with hmC<sup>[17,18]</sup> or fC. Alternatively, decarboxylation of caC (Scheme 1) is a highly attractive alternative, because this mechanism allows exchange of mC by dC without formation of intermediate strand breaks, which are DNA lesions that are known to cause genome instability, and without 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.<sup>[19]</sup>

To gain deeper insight into the putative decarboxylation of caC in stem cells, we developed a sensitive isotope tracing experiment. A double <sup>15</sup>N-labeled caC phosphoramidite reagent [<sup>15</sup>N<sub>2</sub>]-caC was developed, which we incorporated into the 30 mer oligonucleotide **D1** with a sequence from the Ecat1 promoter<sup>[23]</sup> that is known to be the subject of active demethylation. The natural occurrence of the double <sup>15</sup>N isotopologue of dC ([<sup>15</sup>N<sub>2</sub>]-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  $[^{15}\text{N}_2]$ -caC phosphoramidite is depicted in Scheme 2 (see also the Supporting Information). The starting point is  $[^{15}\text{N}_2]$ -uracil **1**, which was prepared from  $[^{15}\text{N}_2]$ -urea and propiolic acid.<sup>[20]</sup> Vorbrüggen nucleosidation with the bis(toluoyl)-protected chlororibofuranoside<sup>[21]</sup> furnished the  $\beta$ -configured nucleoside **2** in 60% yield. Electrophilic iodination to **3**,<sup>[22]</sup> 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  $[^{15}\text{N}_2]$ -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 Ecad1 promoter containing two  $[^{15}\text{N}_2]$ -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.<sup>[23]</sup> 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 1 a, we added the DNA duplex **D1** (**ODN1** + **ODN2**) containing the two  $[^{15}\text{N}_2]$ -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 1 b–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 1 d, e) provided two distinct signals with  $m/z = 230.1014$  and  $230.0976$  in the mass range of the dC nucleoside (Figure 1 e). These  $m/z$  signals correspond to the naturally occurring isotopologues  $[^{13}\text{C}_2]$ - and  $[^{18}\text{O}]$ -dC, respectively, and  $[^{15}\text{N},^{13}\text{C}]$ -dC. The exact mass expected for double  $^{15}\text{N}$ -containing  $[^{15}\text{N}_2]$ -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 1 d, e). We then analyzed the digests obtained from **D1** after



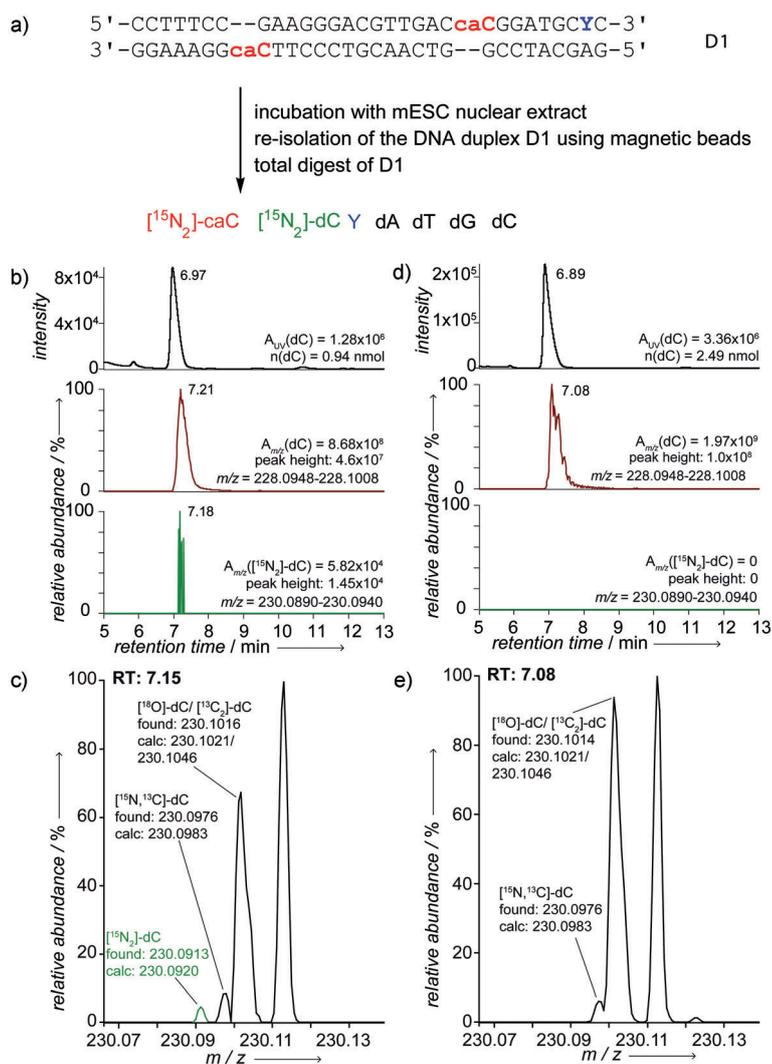
**ODN 1:** 5'-CCT TTC CGA AGG GAC GTT GAC XGG ATG CYC-3'  
calcd 9657 Da, found 9658 Da

**ODN 2:** 5'-GAG CAT CCG GTC AAC GTC CCT TXG GAA AGG-3'  
calcd 9259 Da, found 9259 Da, X =  $[^{15}\text{N}_2]$ -caC, Y = biotin-labeled dU



**Scheme 2.** Synthesis of the  $[^{15}\text{N}_2]$ -labeled caC phosphoramidite and of the DNA strands **ODN1** and **ODN2** containing one  $[^{15}\text{N}_2]$ -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,  $\text{CHCl}_3$ , RT, 2.5 h; c) CAN, Lil, MeCN, 80°C, 2 h; d) 1. 1,2,4-triazole,  $\text{POCl}_3$ , TEA, 30°C, overnight; 2.  $\text{NH}_2\text{OH}$ , 1,4-dioxane, RT, 10 min; e) 1.  $\text{K}_2\text{CO}_3$ , MeOH, RT, 19 h; 2. TBS-Cl, imidazole, DMF, RT, 4 days; f)  $[\text{Pd}(\text{MeCN})_2\text{Cl}_2]$ , 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,  $\text{CH}_2\text{Cl}_2$ , RT, 17 h. HMDS = Hexamethyldisilazane, TMS = trimethylsilyl, CAN = cerium ammonium nitrate, TEA = triethylamine, TBS = *tert*-butyldimethylsilyl, DMT = dimethoxytrityl.

incubation with mESC nuclear extract. Surprisingly we noted that after 1 h of incubation with mESC nuclear extracts, the  $[^{15}\text{N}_2]$ -labeled dC species was indeed detected in the total digest of the re-isolated duplex **D1**. The compound  $[^{15}\text{N}_2]$ -dC eluted with a retention time of 7.18 min. The obtained mass



**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 [<sup>15</sup>N<sub>2</sub>]-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).

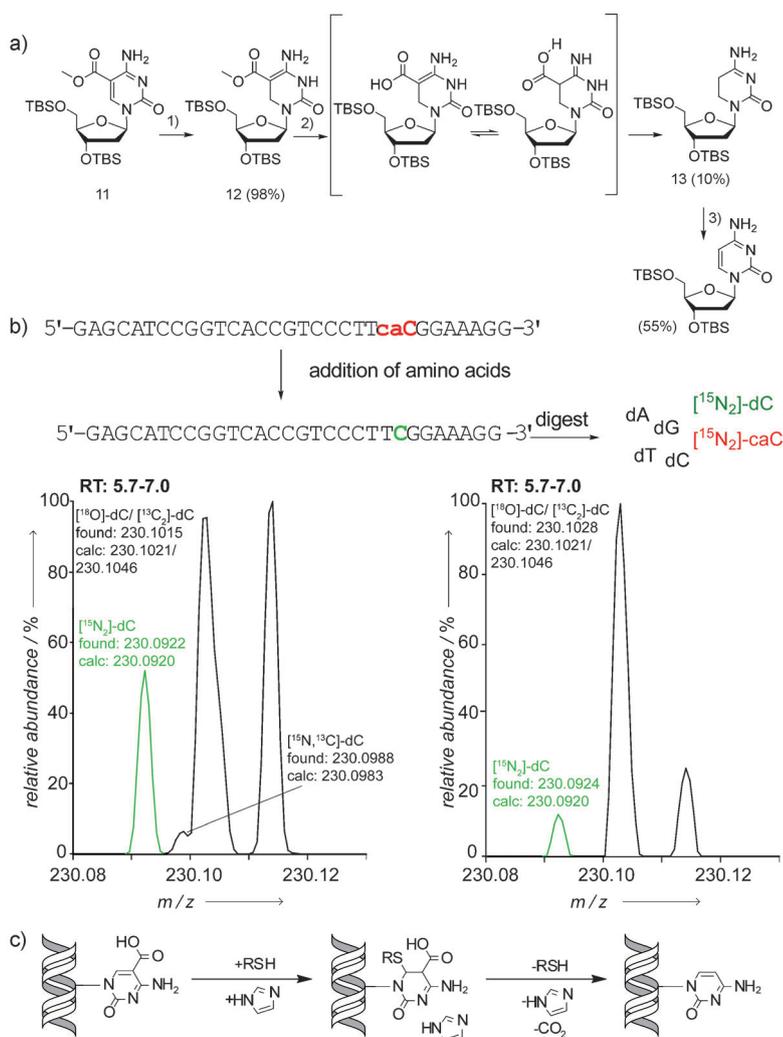
spectrometry signal has the exact mass of  $m/z_{\text{found}} = 230.0913$  (Figure 1 c). The signal is nicely separated from the signals of the other naturally occurring isotopologues [<sup>15</sup>N,<sup>13</sup>C]-dC ( $m/z=230.0976$ ) and [<sup>18</sup>O/<sup>13</sup>C<sub>2</sub>]-dC ( $m/z = 230.1016$ ) and thus is clearly detectable. Furthermore, the measured high-resolution data for the compound [<sup>15</sup>N<sub>2</sub>]-dC agrees very well with the calculated exact mass of [<sup>15</sup>N<sub>2</sub>]-dC ( $m/z_{\text{calcd}} = 230.0920$ ). 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  $\beta$ -iminocarboxylic acid, which would allow efficient decarboxylation (Figure 2).<sup>[10,24]</sup> To investigate this idea, we reduced the 5-carboxy-dC methyl ester **11** with NaBH<sub>4</sub>. 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.<sup>[10]</sup> As reaction with a H<sup>-</sup> 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 [<sup>15</sup>N<sub>2</sub>]-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 [<sup>15</sup>N<sub>2</sub>]-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 <sup>15</sup>N-labeled caC, we could gain first evidence that stem-cell nuclear extracts have the ability



**Figure 2.** a) Reaction of caC with NaBH<sub>4</sub> causes decarboxylation: 1) NaBH<sub>4</sub>, MeOH, RT, 20 min; 2) LiOH, MeCN/H<sub>2</sub>O, RT, 2 h; 3) DDQ, 1,4-dioxane, RT, 25 h. b) Incubation of caC containing DNA with cysteine and 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*-adenosylmethionine (SAM), which provides a CH<sub>3</sub><sup>+</sup>-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.<sup>[17,25]</sup> Another enzymatic reaction on pyrimidine nucleobases, in which the C5–C6 double bond is temporarily saturated is performed by isoorotate decarboxylase (IDCase), which activates water with the help of a Zn<sup>2+</sup> 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.<sup>[26,27]</sup> 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. Hagemeyer, 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. Matsu-moto, 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. Liutkevičiūtė, 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.
-

## **VII. Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and their Incorporation into Oligonucleotides by Polymerase Chain Reaction**

B. Steigenberger\*, S. Schiesser\*, **B. Hackner**, C. Brandmayr, S. K. Laube, J. Steinbacher, . Pfaffeneder, T. Carell, *Organic letters* **2013**, *15*, 366-369. (\* Equal contribution)

### **Introduction:**

Active demethylation as described in chapter I.2.3 is a process that is yet not fully understood. Further biochemical tools for its investigations are needed to shed light into it. In this manuscript the synthesis and incorporation of oxidized cytidine triphosphates is described.

### **Declaration of contribution:**

I performed the incorporation of fCTP and caCTP using polymerase chain reaction (PCR). I performed the analysis of the incorporation using LC-MS and LC-MS/MS methods. I contributed to the writing manuscript and prepared Figures 2 and 3.

# Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction

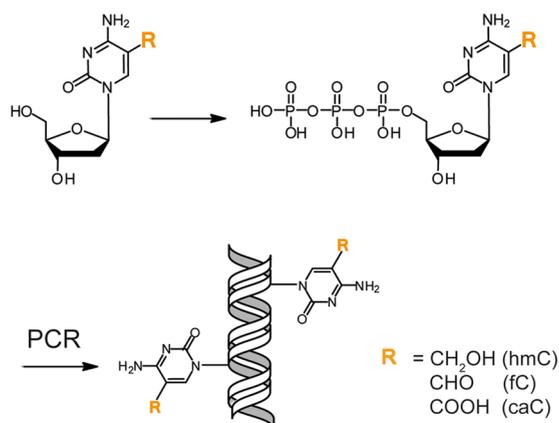
Barbara Steigenberger,<sup>†</sup> Stefan Schiesser,<sup>†</sup> Benjamin Hackner, Caterina Brandmayr, Silvia K. Laube, Jessica Steinbacher, Toni Pfaffeneder, and Thomas Carell\*

Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians Universität München, Butenandtstr. 5-13, 81377, Munich, Germany

Thomas.Carell@lmu.de

Received December 4, 2012

## ABSTRACT



The synthesis of the triphosphates of 5-hydroxymethyl-, 5-formyl-, and 5-carboxycytidine and the incorporation of these building blocks into long DNA fragments using the polymerase chain reaction (PCR) are reported. In this way DNA fragments containing multiple hmC, fC, and caC nucleobases are readily accessible.

Recently three new nucleobases were discovered in DNA isolated from mouse embryonic stem cells.<sup>1–4</sup> The three nucleobases are oxidation products of 5-methylcytosine (mC), which is a base that regulates transcriptional activity.<sup>5</sup> The oxidation reaction of 5-methylcytosine to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and finally 5-carboxycytosine (caC) is now understood to

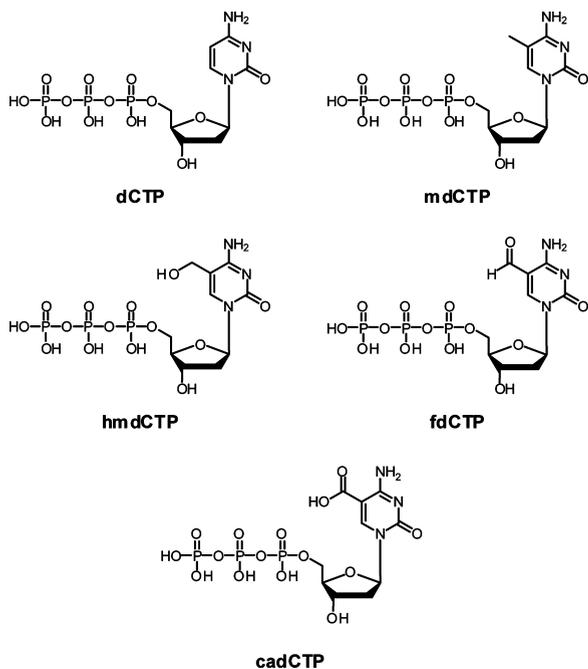
be performed by 10-11-translocase proteins (TET1–3).<sup>6</sup> These enzymes are  $\alpha$ -ketoglutarate dependent oxidases, which directly utilize molecular oxygen for the oxidation reaction.<sup>7</sup> The three new nucleobases are currently thought to be involved in epigenetic programming of cells, and they could be intermediates of a long searched for pathway of active demethylation.<sup>8–10</sup> Recently new sequencing methods that allow the genome wide localization of hmC<sup>11,12</sup> and fC<sup>13</sup> in genomic DNA were reported.

<sup>†</sup> These authors contributed equally.

- (1) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
- (2) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
- (3) Pfaffeneder, T.; Hackner, B.; Truss, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeyer, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008.
- (4) He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C.-X.; Zhang, K.; He, C.; Xu, G.-L. *Science* **2011**, *333*, 1303.
- (5) Law, J. A.; Jacobsen, S. E. *Nat. Rev. Genet.* **2010**, *11*, 204.

- (6) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.
- (7) Loenarz, C.; Schofield, C. J. *Chem. Biol.* **2009**, *16*, 580.
- (8) Jurkowski, T. P.; Jeltsch, A. *ChemBioChem* **2011**, *12*, 2543.
- (9) Gu, T.-P.; Guo, F.; Yang, H.; Wu, H.-P.; Xu, G.-F.; Liu, W.; Xie, Z.-G.; Shi, L.; He, X.; Jin, S.-g.; Iqbal, K.; Shi, Y. G.; Deng, Z.; Szabo, P. E.; Pfeifer, G. P.; Li, J.; Xu, G.-L. *Nature* **2011**, *477*, 606.
- (10) Ladwein, K. I.; Jung, M. *Angew. Chem., Int. Ed.* **2011**, *50*, 12143.

In the past couple of years phosphoramidite building blocks of hmC, fC, and caC, which allow the solid phase synthesis of oligonucleotides containing the new bases at defined sites, were developed.<sup>14–18</sup> However, to search for proteins that interact with these new epigenetic bases, and to decipher the biological/biochemical questions associated with the new nucleobases, longer oligonucleotides containing multiple hmC, fC, and caC bases are required. In this direction, we thought that using the corresponding triphosphates (Figure 1) in combination with the polymerase chain reaction (PCR) would solve this chemical problem.<sup>19–21</sup>



**Figure 1.** Depiction of the canonical DNA base dC and of mC as well as of the new epigenetic bases hmC, fC, and caC as triphosphates.

(11) Booth, M. J.; Branco, M. R.; Ficiz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S. *Science* **2012**, *336*, 934.

(12) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C.-X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J.-H.; Jin, P.; Ren, B.; He, C. *Cell* **2012**, *149*, 1368.

(13) Raiber, E.-A.; Beraldi, D.; Ficiz, G.; Burgess, H. E.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, *13*, R69.

(14) Tardy-Planechaud, S.; Fujimoto, J.; Lin, S. S.; Sowers, L. C. *Nucleic Acids Res.* **1997**, *25*, 553.

(15) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671.

(16) Dai, Q.; He, C. *Org. Lett.* **2011**, *13*, 3446.

(17) Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chem.—Eur. J.* **2011**, *17*, 13782.

(18) Dai, Q.; He, C. *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc.: 2001.

(19) Jäger, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M. *J. Am. Chem. Soc.* **2005**, *127*, 15071.

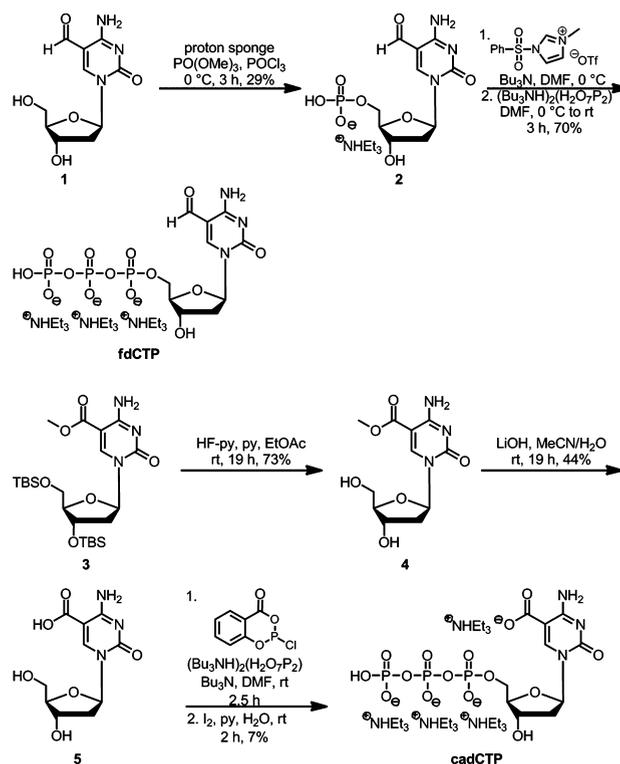
(20) Kuwahara, M.; Nagahima, J.; Hasegawa, M.; Tamura, T.; Kitagata, R.; Hanawa, K.; Hososhima, S.; Kasamatsu, T.; Ozaki, H.; Sawai, H. *Nucleic Acids Res.* **2006**, *34*, 5383.

(21) Shoji, A.; Hasegawa, M.; Hososhima, S.; Kuwahara, M.; Ozaki, H.; Sawai, H. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 776.

In such a PCR, one would exchange the dCTP either completely or partially by the corresponding xdCTP ( $x = \text{hm}, \text{f}$  or  $\text{ca}$ ) so that these DNA fragments contain the new bases hmC, fC, and caC at the corresponding dC positions.

While the triphosphate of hmC (hmdCTP) is already commercially available and its incorporation via PCR is established, the corresponding fdCTP and cadCTP compounds were unknown at the beginning of this study. The chemical synthesis of both building blocks was achieved as outlined in Scheme 1. For fdCTP we started the synthesis with the fC nucleoside **1**, which was prepared as described recently by us.<sup>22</sup> This compound was converted into the 5'-monophosphate **2**, which was obtained after HPLC purification in 29% yield. We subsequently employed the new triphosphate method recently described by S. D. Taylor et al. using sulfonyl imidazolium triflate as the activating reagent and pyrophosphate.<sup>23</sup> The procedure allowed us to access the triphosphate from the monophosphate in 70% yield. The triphosphate was best isolated by ion exchange chromatography at 4 °C using a DEAE-cellulose column<sup>24</sup> with a gradient from 100% water to 0.5 M TEAB (pH 7.5). The crude triphosphate product was further purified by FPLC (0.1 M TEAB, 1 M TEAB; 0–100% over 30 min) using a MonoQ 5/50 GL anion exchange column (GE). This two-step procedure allowed us to generate the reactive aldehyde-containing fC triphosphate in sufficient yield for all further studies.

### Scheme 1. Synthesis of fdCTP and cadCTP



(22) Globisch, D.; Münzel, M.; Müller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. *PLoS ONE* **2010**, *5*, e15367.

(23) Mohamady, S.; Desoky, A.; Taylor, S. D. *Org. Lett.* **2011**, *14*, 402.

The cadCTP compound was prepared from the TBS protected caC methyl ester **3** which was also prepared as described recently by us.<sup>17</sup> We first cleaved the TBS groups to obtain compound **4**. The methyl ester was subsequently saponified which provided the unprotected caC nucleoside **5**. **5** was next introduced into the one-step triphosphate synthesis reported originally by Eckstein et al.<sup>25</sup> Here the yield could be improved to 7% if the conditions developed by Huang et al. were employed.<sup>26</sup>

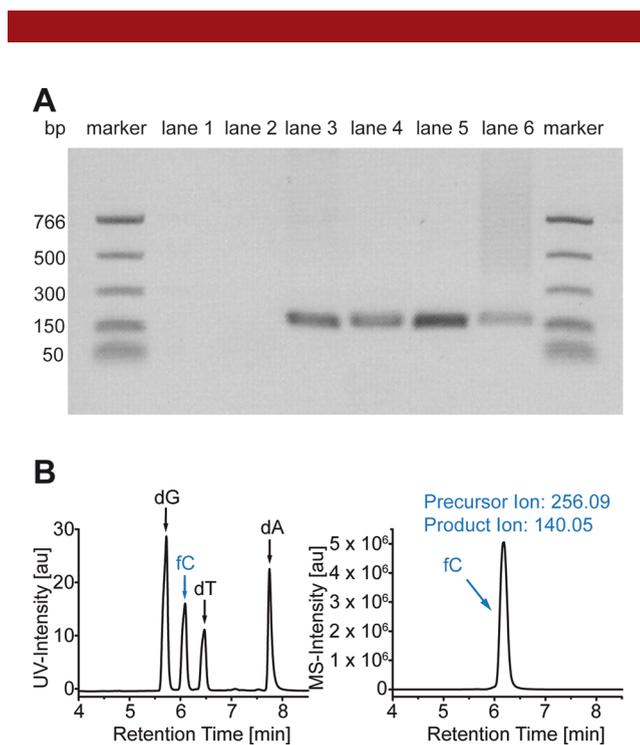
We recently reported the synthesis of hmC, fC, and caC phosphoramidites and the incorporation of these building blocks into DNA strands. Primer extension studies showed that none of the new bases are mutagenic.<sup>17</sup> Here we report the development of PCR conditions for the incorporation of fdCTP and cadCTP into long oligonucleotides. We chose the oct4 promoter sequence (see Supporting Information (SI)) as the DNA template. The primers for the PCR were designed to yield a 150 bp product containing 77 modified dCs (4 dC are present in the primer; these are not exchanged). For this purpose the forward and reverse primers were annealed to the template at 55 °C. The elongation of the primers was best performed at 75 °C (for fdCTP) and 72 °C (for cadCTP). Different polymerases were screened. We discovered that the DNA polymerase Vent (exo<sup>-</sup>) (for fdCTP) and KOD XL polymerase for (cadCTP) provided the best results (see SI). To ensure complete extension of the primer the elongation time was lengthened compared to the time used for incorporation of dCTP. The experimental results of the PCRs are described in Figure 2.

The PCR products obtained with cadCTP can only be visualized when a 1 × TBE buffer system is used for the analysis. When other buffers such as TAE were used we noted that the obtained oligonucleotide products did not give a distinct band in the gel electrophoresis potentially because of the additional carboxylic acid groups present on caC. Rather a broad smear is detected due to the lower buffer capacity. The TBE buffer system in contrast provides sharp bands for the caC containing DNA products. As depicted in lanes 5 and 6, both triphosphates fdCTP (lane 5) and cadCTP (lane 6) yielded PCR products with the correct length (negative control lanes 1 and 2) if our developed PCR methods are employed. In the shown experiments we replaced the dCTP completely by the corresponding xdCTP. Hence full length PCR product can only be formed when the triphosphate is accepted, as further shown by the negative control in lane 2. Further proof for the correct incorporation of fC and caC using PCR was obtained by LC-MS experiments. The PCR products were to this end fully digested. For this purpose the sugar phosphate backbone was first cleaved with nuclease S1 and snake venom phosphodiesterase, giving the 5'-monophosphates. These were further hydrolyzed to the nucleoside level by Antarctic phosphatase.<sup>3</sup> Using this

(24) Guan, L.; van der Heijden, G. W.; Bortvin, A.; Greenberg, M. M. *ChemBioChem* **2012**, *12*, 2184.

(25) Ludwig, J.; Eckstein, F. *J. Org. Chem.* **1989**, *54*, 631.

(26) Caton-Williams, J.; Smith, M.; Carrasco, N.; Huang, Z. *Org. Lett.* **2011**, *13*, 4156.



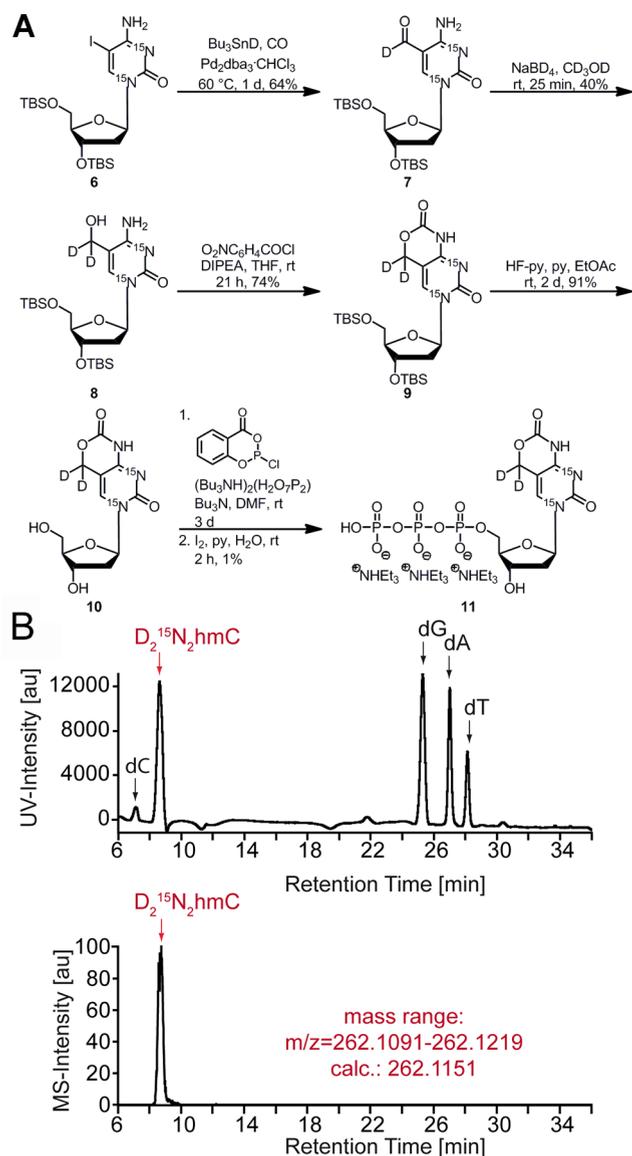
**Figure 2.** (A) Depiction of the results of the PCR analyzed by gel electrophoresis. Lane 1: without template. Lane 2: without any dC derivative. Lane 3: dCTP. Lane 4: protected D<sub>2</sub>, <sup>15</sup>N<sub>2</sub>-hmdCTP. Lane 5: fdCTP. Lane 6: cadCTP. (B) UV trace (left) and mass trace of fC (right) of a fully digested PCR mixture with fdCTP.

procedure even the highly modified DNA prepared here was fully digested. The resulting nucleoside mixture was subsequently analyzed by LC-HRMS or LC-MS/MS. The data obtained for incorporated fdCTP are shown in Figure 2B (for cadCTP, see SI). Clearly evident is the presence of dA, dT, and dG in addition to the fC-nucleoside. Our results show that both fC and caC can be inserted as triphosphates into long DNA fragments using PCR. The observation that formyl group containing nucleosides can be incorporated into PCR products despite their high reactivity is in line with a recent report by Hocek et al. This group reported the PCR based synthesis of aldehyde containing DNA products.<sup>27,28</sup> It is interesting that the polymerase tolerates also the negative charge associated with the carboxylic acid present in caC. For future quantification of fC and caC in natural material we also prepared isotope labeled fC and caC triphosphates and incorporated them into DNA strands using basically the same PCR conditions showing the broad applicability of the here reported technology (see SI).

We next turned our attention to the hmC base. Here, the corresponding triphosphate is generated from the monophosphate, which is directly isolated from natural sources. A chemical synthesis of the hmC triphosphate was not

(27) Raindlová, V.; Pohl, R.; Šanda, M.; Hocek, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 1064.

(28) Raindlová, V.; Pohl, R.; Hocek, M. *Chem.—Eur. J.* **2012**, *18*, 4080.



**Figure 3.** (A) Synthesis of the protected  $D_2, ^{15}N_2$ -hmdCTP. (B) UV trace (top) and  $D_2, ^{15}N_2$ -hmC mass trace (bottom) of a fully digested PCR mixture.

performed so far, which limits our ability to create DNA fragments with modified hmC, e.g. with isotopically labeled compounds as needed for mass spectrometry based quantification and proteomics studies. The synthetic challenge associated with the synthesis of hmdCTP is the benzylic hydroxyl group of the hmC heterocycle which is more reactive than the primary 5'-OH group. This makes it difficult to access the 5'-monophosphate directly from hmC. To circumvent this problem and to enable the PCR incorporation of modified hmC building blocks, we investigated the ability to insert a protected hmC derivative by PCR followed by deprotection. The synthesis of the

protected hmdCTP building block and the results of the PCR study are shown in Figure 3.

To exemplify the possibilities associated with the chemical method, we prepared for this study the unnatural  $2 \times [^{15}N]$  and  $2 \times D$  modified hmC (SI). The starting point for the synthesis is the TBS protected iodouracil **6**,<sup>29</sup> which was carbonylated to **7** and reduced to **8**. Compound **8** was protected as the carbamate **9**, and the TBS groups were cleaved to obtain **10**. We utilized the optimized one-pot triphosphate synthesis described above to obtain the labeled hmdCTP **11**. Subsequent PCR based incorporation studies showed that the reaction is best performed with the KOD XL polymerase with again slightly prolonged elongation times (30 s instead of 15 s). The PCR product was subsequently deprotected with 0.1 M NaOH in water/methanol 1:4 for 1 h at rt. The DNA was finally purified using a silica membrane (see SI). The agarose gel of the deprotected and purified DNA fragment is depicted in Figure 2A, lane 4. Again a clean PCR product is observed. The results of the total digest performed under the optimized conditions reported above are shown in Figure 3B. Again next to dA, dG, and dT as well as small amounts of residual dC from the primers, an additional signal is observed with the correct retention time and exact molecular weight for  $D_2, ^{15}N_2$ -hmC. Most importantly we do not observe a signal for un-deprotected hmC showing that full deprotection of the carbamate protecting group present on our hmdCTP building block was achieved. In summary we report here the first chemical synthesis of the three triphosphates hmdCTP, fdCTP, and cadCTP of the new epigenetic bases hmC, fC, and caC and describe PCR conditions which enable the incorporation of these building blocks into long DNA fragments. The chemical synthesis allows even the synthesis and incorporation of isotopologues of the new bases, which should facilitate mass spectrometry based quantification methods.

We believe that the reported synthetic methodologies disclosed here will strongly advance our ability to study the biology and biochemistry of the new epigenetic bases hmC, fC and caC.

**Acknowledgment.** We thank the excellence cluster EXC114 and the SFB749 as well as the Volkswagen foundation and the DFG for Grant CA275/8-4 for financial support. S.S. and T.P. thank the Fonds der Chemischen Industrie for predoctoral fellowships. C.B. thanks the Boehringer Ingelheim Fonds for predoctoral fellowship.

**Supporting Information Available.** Experimental procedures and spectroscopic data of all new compounds. This material is available free of charge via the Internet at <http://pubs.acs.org>.

(29) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeyer, C.; Truss, M.; Carell, T. *Angew. Chem., Int. Ed.* **2012**, *51*, 6516.

The authors declare no competing financial interest.

# **VIII. Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain**

P. Bashtrykov, A. Rajavelu, **B. Hackner**, S. Ragozin, T. Carell and A. Jeltsch,  
*ChemBioChem*, **2014**, accepted, DOI: 10.1002/cbic.201300740

## **Introduction:**

DNA methylation is a highly important and tightly regulated process for cellular development as described in chapter I.2.2. Maintenance methylation is predominantly regulated by the N-terminal part of DNMT1. In this manuscript an intrinsic regulation by ionic contacts of the RFTS-domain to the catalytic domain is described and its effects analyzed.

## **Declaration of contribution:**

I performed the genome wide mC level analysis in the mutants using LC-MS/MS. My main is summarized in figure 5.

DOI: 10.1002/cbic.201300740

# Targeted Mutagenesis Results in an Activation of DNA Methyltransferase 1 and Confirms an Autoinhibitory Role of its RFTS Domain

Pavel Bashtrykov,<sup>[a]</sup> Arumugam Rajavelu,<sup>[a]</sup> Benjamin Hackner,<sup>[b]</sup> Sergey Ragozin,<sup>[a]</sup> Thomas Carell,<sup>[b]</sup> and Albert Jeltsch<sup>\*[a]</sup>

The N-terminal regulatory part of DNA methyltransferase 1 (Dnmt1) contains a replication foci targeting sequence (RFTS) domain, which is involved in the recruitment of Dnmt1 to replication forks. The RFTS domain has been observed in a crystal structure to bind to the catalytic domain of the enzyme and block its catalytic centre. Removal of the RFTS domain led to activation of Dnmt1, thus suggesting an autoinhibitory role of this domain. Here, we destabilised the interaction of the RFTS domain with the catalytic domain by site-directed mutagenesis and purified the corresponding Dnmt1 variants. Our data show

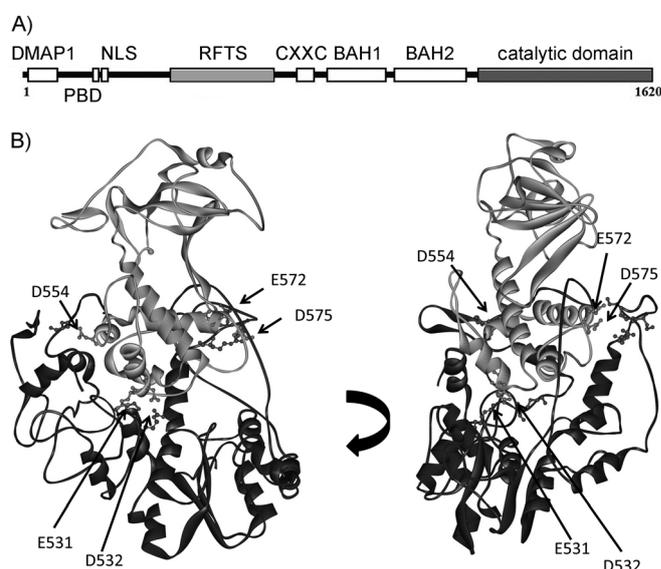
that these mutations resulted in an up to fourfold increase in Dnmt1 methylation activity in vitro. Activation of Dnmt1 was not accompanied by a change in its preference for methylation of hemimethylated CpG sites. We also show that the Dnmt1 E572R/D575R variant has a higher DNA methylation activity in human cells after transfection into HCT116 cells, which are hypomorphic for Dnmt1. Our findings strongly support the autoinhibitory role of the RFTS domain, and indicate that it contributes to the regulation of Dnmt1 activity in cells.

## Introduction

DNA methylation is a repressive epigenetic signal involved in the regulation of gene expression and chromatin structure. In mammals, DNA methylation takes place at position 5 of cytosines, primarily within CpG dinucleotides.<sup>[1–5]</sup> Approximately 60–70% of CpG sites are methylated, and form cell-type-specific DNA methylation patterns. Aberrant DNA methylation is implicated in the pathogenesis of many diseases, for example those of the immune system, the nervous system and cancer.<sup>[6–7]</sup> DNA methylation is durable and heritable: it can be transmitted through cell generations, because hemimethylated CpG sites in newly synthesised DNA are recognised and remethylated by Dnmt1. Hence, the preference of Dnmt1 for hemimethylated CpG sites over unmethylated sites plays a fundamental role in the maintenance of DNA methylation pattern.<sup>[5]</sup> Mouse Dnmt1 (1620 residues) consists of a C-terminal catalytic domain and an N-terminal regulatory multidomain, which includes the RFTS (replication foci targeting sequence, the subject of this study; Figure 1 A). In vitro experiments have shown that Dnmt1 has a preference of approximately 10- to 40-fold for hemimethylated CpG sites.<sup>[8–14]</sup> However, this preference is not enough to ensure accurate copying of methylation patterns that encompass millions of CpG sites, thus suggesting that additional mechanisms exist to increase the specificity for

hemimethylated CpG sites and to provide precise control of its activity in vivo.

Recently, based on structural and biochemical studies, two models of allosteric control of Dnmt1 activity by its N-terminal



**Figure 1.** Structure of Dnmt1. A) Domain architecture of Dnmt1 showing RFTS domain (light grey) and catalytic domain (dark grey). B) Binding of the RFTS domain to the DNA-binding pocket of Dnmt1 is stabilised by several hydrogen bonds and ionic contacts. Parts of the catalytic and CXXC domains are shown in dark and light grey, respectively; RFTS residues selected for mutagenesis are labelled; catalytic and CXXC domain residues involved in the interaction with the RFTS domain are shown as well. (Images prepared from crystal structure PDB ID: 3AV4.)<sup>[15]</sup>

[a] P. Bashtrykov, Dr. A. Rajavelu, Dr. S. Ragozin, Prof. Dr. A. Jeltsch  
Institute of Biochemistry, Faculty of Chemistry, University Stuttgart  
Pfaffenwaldring 55, 70569 Stuttgart (Germany)  
E-mail: albert.jeltsch@ibc.uni-stuttgart.de

[b] B. Hackner, Prof. Dr. T. Carell  
Center for Integrated Protein Science (CiPSM) at the  
Department of Chemistry, Ludwig Maximilians University  
Butenandtstrasse 5–13, 81377 Munich (Germany)

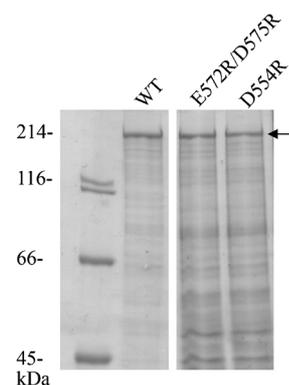
region were proposed. The structure of a truncated form of Dnmt1 (amino acids 650–1620: CXXC, BAH1, BAH2 and catalytic domains) bound to unmethylated DNA was solved, and biochemical experiments with this Dnmt1 fragment provided evidence that the CXXC domain has an autoinhibitory role.<sup>[11]</sup> However, in a separate investigation, the regulatory role of the CXXC domain in full-length Dnmt1 could not be shown.<sup>[13]</sup> In addition, the structure of a Dnmt1 variant (residues 291–1620) showed that the RFTS domain binds to the DNA-binding pocket in the catalytic domain of Dnmt1, and that the CXXC domain is pushed aside.<sup>[15]</sup> Based on this, it was proposed that Dnmt1 has to undergo a structural change to allow DNA binding and methylation; this includes movement of the RFTS domain. In a parallel study, Syeda et al. crystallised the isolated RFTS domain (351–600) and, based on modelling, proposed that the RFTS domain can bind to the catalytic domain and can occupy the catalytic site of the enzyme.<sup>[16]</sup> In support of their proposal, a human Dnmt1 variant (621–1616) without the RFTS domain showed stronger binding to DNA than a longer Dnmt1 fragment (351–1616) that included the domain. They also compared the DNA methylation activity of these two truncated variants and found that the turnover of Dnmt1 (621–1616) was approximately 40 times higher than that of Dnmt1 (351–1616), and that addition of the RFTS domain to the methylation reaction inhibited the activity of Dnmt1 (621–1616).<sup>[16]</sup> These data support the model that the RFTS domain is an endogenous inhibitor of DNA binding and methylation by Dnmt1. However, as Dnmt1 is a large multidomain protein, the removal of domains by truncation might have unpredictable consequences, as illustrated by the different results obtained with full-length Dnmt1 and with truncated versions in the studies related to the role of the CXXC domain in the regulation of the specificity of the enzyme.<sup>[11,13]</sup> In the current work we aimed to verify the autoinhibitory role of the RFTS domain in the context of the full-length Dnmt1, by targeted site-directed mutagenesis within the RFTS domain to reduce its interaction with the catalytic domain, and by measurement of the catalytic activity of the corresponding full-length Dnmt1 variants in vitro and in cells.

## Results and Discussion

### Design and expression of Dnmt1 mutants

According to the crystal structure of Dnmt1 (291–1620) published by Takeshita et al.,<sup>[15]</sup> the RFTS domain plays an autoinhibitory role in Dnmt1 by occupying the catalytic cleft of the enzyme and thereby preventing access of the DNA substrate to the catalytic site. To catalyse a methylation reaction, Dnmt1 has to adopt another conformation: one in which the RFTS domain is released from the catalytic domain. Binding of the RFTS domain to the DNA-binding cleft involves several hydrogen bonds and ionic contacts, including E531–K1537, D532–R1576, D554–S1495, A594–H1504 and L593–T1505 (first residue in the RFTS domain, second in the catalytic domain). In addition, the RFTS domain forms interactions with the CXXC domain, including E572–K691, E572–R601 and D575–R693 (first

residue in RFTS, second in CXXC), thus stabilising its binding to the catalytic domain. We selected residues in the RFTS domain for charge reversal mutagenesis to destabilise the interaction of the RFTS domain with the catalytic domain and thereby trigger its release from the catalytic domain. Amino acids E531, D532, D554, E572 and D575 were selected (Figure 1), and three Dnmt1 variants (E531R/D532R, D554R and E572R/D575R) were cloned in a baculovirus expression system. However, we were unable to express the E531R/D532R variant, so experiments were performed with just the other two (Figure 2). We did not



**Figure 2.** SDS PAGE of wild-type (WT) and mutant Dnmt1 to compare concentrations of WT and E572R/D575R and D554R mutants used in the kinetics experiments. Arrow indicates bands of full-length proteins.

mutate residues in the catalytic domain of Dnmt1, because this could cause direct effects on the catalytic activity; such effects would not be related to an autoinhibitory role of the RFTS domain and would confuse the interpretation of results.

### DNA methylation activity of Dnmt1 mutants

The DNA methylation activity of the mutants was investigated by in vitro DNA methylation assays. The methylation reactions were carried out with a 30-mer DNA substrate containing a single CpG site in a hemimethylated state and *S*-adenosyl-L-methionine (AdoMet) with a radioactively labelled methyl group. The activities of wild-type Dnmt1 and its variants were assessed by incorporation of radioactivity into the DNA. We observed a DNA methylation rate of  $1.3 \text{ h}^{-1}$  under our conditions (Figure 3A), which was too slow to conduct multiple-turnover experiments. Therefore, we measured DNA methylation during the initial reaction phase using excess DNA and AdoMet, to determine the single-turnover rate. Interestingly, the mutations resulted in an increase in DNA methylation activity in comparison to the wild type: approximately fourfold and 2.5-fold for E572R/D575R and D554R, respectively (Figure 3). Thus, removal of contacts involved in stabilising the autoinhibitory conformation of Dnmt1 resulted in its activation.

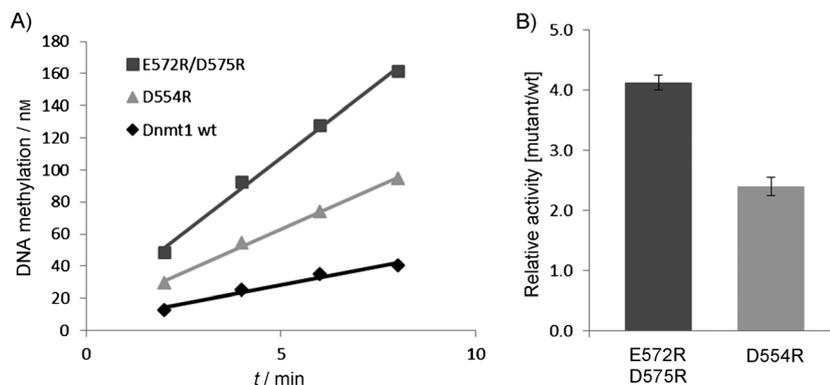
## Substrate specificity of Dnmt1 mutants

Next, we analysed the substrate specificity of the Dnmt1 mutants. For this we used an otherwise identical 30-mer DNA substrate, but with an unmethylated CpG site. Methylation reactions with hemimethylated and unmethylated substrates were carried out in parallel for wild-type Dnmt1 and the variants. Wild-type Dnmt1 showed a 16-fold preference for the hemimethylated substrate over unmethylated under the tested conditions (Figure 4), similarly to previous results.<sup>[13,14]</sup> The specificities of the Dnmt1 E572R/D575R and D554R variants were very similar and within the error range of the data obtained for wild-type Dnmt1. We conclude that autoinhibition of Dnmt1 by the RFTS domain plays a role in the control of the catalytic activity of Dnmt1, but that it is not involved in a regulation of the specificity of Dnmt1.

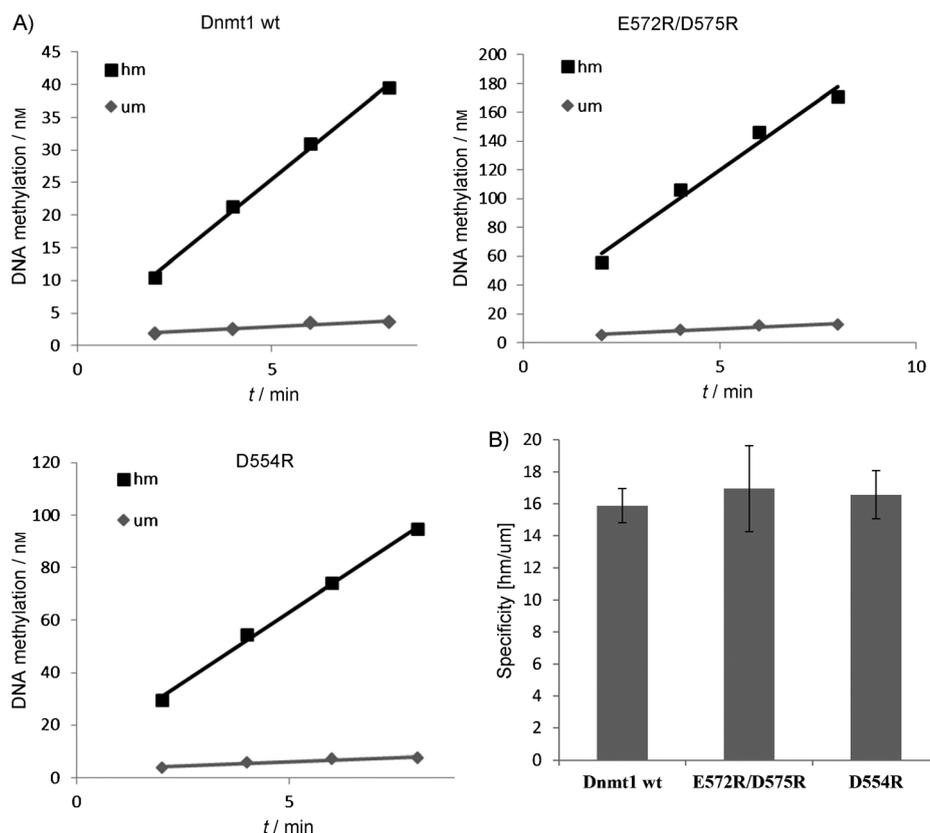
## Dnmt1 E572R/D575R has increased activity in cells

To study the DNA methylation activity of Dnmt1 in human cells we used HCT116 Dnmt1 hypomorphic cells, which contain a truncated Dnmt1 with reduced activity. Because of the impaired maintenance DNA methylation activity, these cells have an about 20% reduced DNA methylation and show increased levels of hemimethylation.<sup>[17,18]</sup> The cells were transfected with plasmids encoding YFP-fused Dnmt1 and Dnmt1 E572R/D575R. A YFP-expressing plasmid was used as a control. Transfection yields were assessed by FACS and found to be almost identical in all three experiments (Figure 5A). Genomic DNA was isolated, and 5-methylcytosine was quantified by LC-MS/MS (Figure 5B). Transfection with wild-type Dnmt1 led to a strong increase in DNA methylation of the hypomorphic cells (from 2.59 to 2.82%), thus indicating that the assay was working.

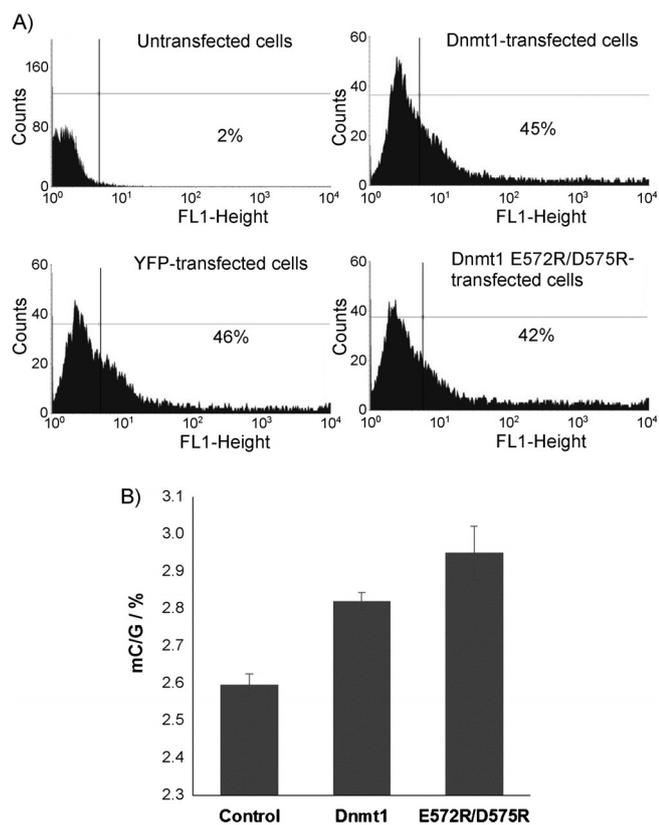
Transfection with the E572R/D575R variant led to an even higher increase in DNA methylation (to 2.95%) that was 60% larger than for wild-type Dnmt1. This shows that the E572R/D575R variant has higher activity in human cells, thus indicating that the binding of the RFTS domain to the catalytic pocket also reduces the activity of Dnmt1 in vivo.



**Figure 3.** Targeted mutagenesis in the Dnmt1 RFTS domain results in activation of DNA methylation activity. A) Examples of DNA methylation kinetics with Dnmt1 variants and hemimethylated substrates. B) Catalytic activities of Dnmt1 variants for methylation of a hemimethylated oligonucleotide substrate, relative to wild-type Dnmt1. Activities are averages of three to six independent experiments (error bars: SEM).



**Figure 4.** Substrate specificities of Dnmt1 variants. A) Example kinetics of wild-type Dnmt1 and mutants for methylation of unmethylated (um) and hemimethylated (hm) substrates. B) Dnmt1 specificity of variants is expressed as the ratio of the rate of methylation of hemimethylated substrate to that of unmethylated substrate. Mutagenesis in the RFTS domain did not change specificity of the mutants. Specificities are averages of three to six independent experiments (error bars: SEM).

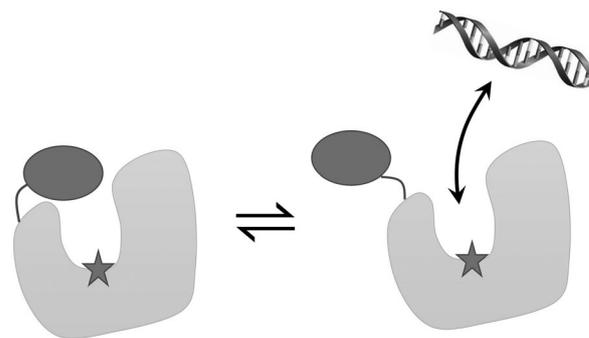


**Figure 5.** Mutant E572R/D575R is more active than wild-type Dnmt1 in HCT116 Dnmt1 hypomorphic cells. A) HCT116 Dnmt1 hypomorphic cells were transfected with YFP (control) or expression plasmids for YFP-fused Dnmt1 and Dnmt1 E572R/D575R. Transfection yields were determined by FACS, and found to be similar in all three experiments. The percentage of YFP-positive cells is indicated in the corresponding images. B) Genomic DNA was isolated from transfected cells, and the amount of methylated cytosine was quantified by LC-MS/MS (average of three measurements, SEM). Cytosine methylation is given here as the ratio of 5-methylcytosine and guanine as a measure of total cytosine (mC/G) in %.

## Conclusions

Autoinhibition is a common mechanism for regulation of enzyme activity. Protein kinase A is a well-investigated example. The enzyme consists of regulatory and catalytic subunits and forms a heterotetrameric complex in the absence of cAMP. This complex is catalytically inactive because the regulatory subunits contain pseudosubstrate sites, which mimic the substrate and tightly bind to the active site of the catalytic subunit, thereby blocking its activity. After binding of cAMP to the regulatory subunits, they undergo a conformational change and dissociate from the catalytic subunit; catalytic activity is switched on by this event.<sup>[19,20]</sup>

Here we provide evidence that the RFTS domain plays a similar autoinhibitory role in Dnmt1 by binding to its catalytic domain and blocking the DNA binding cleft of the enzyme (Figure 6). We were able to show that charge-reversal mutations of residues that mediate binding of the RFTS domain to the catalytic domain weaken this interaction and lead to activation of Dnmt1. As it is rare that such a drastic exchange of



**Figure 6.** Allosteric regulation of Dnmt1. The RFTS domain (dark grey) has an autoinhibition function, because it binds to the catalytic domain of the enzyme (light grey). This blocks access by the DNA to the DNA-binding cleft and catalytic centre (star). Mutations in the RFTS domain weakened its binding to the catalytic domain, thereby causing activation of the enzyme both in vitro and in cells.

residues of an enzyme leads to increased activity, our finding provides strong support for the autoinhibition model proposed previously by Syeda et al.<sup>[16]</sup> and Takeshita et al.<sup>[15]</sup> A similar increase in activity by the mutants was observed with substrates containing unmethylated or hemimethylated CpG sites, thus indicating that the RFTS domain represses activity of Dnmt1 but does not change its specificity. We interpret this finding as indicating that the weaker binding of the RFTS domain to the catalytic domain changes the ratio of Dnmt1 conformations with blocked and open DNA binding cleft, and that this results in higher overall activity. As specificity is determined mainly by specific interactions of the DNA substrate with the catalytic domain,<sup>[12–14]</sup> the change in the conformation of the RFTS domain did not affect the preference of the enzyme for hemimethylated DNA.

The elevated activity of Dnmt1 E572R/D575R in cells shows that the binding of the RFTS domain to the catalytic pocket also reduces the activity of Dnmt1 in vivo. This indicates that modulation of this interaction could be a regulatory mechanism in Dnmt1. However, the endogenous signal to trigger release of the RFTS domain from the DNA-binding pocket remains to be determined. One may speculate that interaction with other proteins plays a role in this. The protein Uhrf1 has been shown to recruit Dnmt1 to hemimethylated DNA and repressed DNA.<sup>[21–23]</sup> Syeda et al. proposed that Uhrf1 interacts with the RFTS domain of Dnmt1 and helps to release it from the catalytic site, and thereby activates DNA methylation by Dnmt1.<sup>[16]</sup> In agreement with this, we recently showed that binding of Uhrf1 stimulates Dnmt1, and that this effect is mediated by the RFTS domain.<sup>[28]</sup> Other factors involved in this regulation remain to be elucidated.

## Experimental Section

**Protein expression and purification:** Wild-type mouse Dnmt1 (accession: NP\_034196.5) and Dnmt1 mutants E531R/D535R, D554R and E575R/D575R were expressed by using the Bac-to-Bac baculovirus expression system (Invitrogen/Life Technologies) according to the manufacturer's instructions, basically as previously de-

scribed.<sup>[13,24]</sup> All genes were cloned into pFastBacHTa as N-terminal His<sub>6</sub> and EYFP fusions and purified with Ni-NTA Agarose (Genaxxon, Ulm, Germany) by following standard procedures. Dnmt1 mutants were generated by site-directed mutagenesis<sup>[25]</sup> using the following primers: E531R/D532R (forward primer 5'-CTG TAT ATC GGC GCT TGA TCA ATA-3', reverse primer 5'-TCA TCG TCG TCT GCC TCC TTCA-3'), D554R (forward primer 5'-TTC ACA GAG AGA TCT CTC TTA CGC CA-3', reverse primer 5'-TCA TCG TCG TCT GCC TCC TTCA-3'), E572R/D575R (forward primer 5'-TAC GAC AGG GCC AAG AGA GAT GAT GA-3', reverse primer 5'-TCA TCG TCG TCT GCC TCC TTCA-3').

**In vitro DNA methylation assay:** DNA methylation assays were performed as described previously.<sup>[24]</sup> Briefly, methylation reactions were carried out with DNA substrate (2 μM), [methyl-<sup>3</sup>H]AdoMet (1.125 μM; PerkinElmer) and recombinant Dnmt1 (0.2 μM) at 37 °C in methylation buffer (HEPES (20 mM, pH 7.5), EDTA (1 mM), KCl (50 mM)). DNA substrates for the assay were prepared by annealing a 30-mer oligonucleotide (TTG CAC TCT CCT CCGG GAA GTC CCA GCT TC) containing one unmethylated CpG site (bold) with a complementary oligonucleotide (either unmethylated or methylated), to form unmethylated and hemimethylated substrates. All methylation experiments were carried out at least in triplicate. Initial rates were determined by linear regression. Methylation assays were calibrated by complete methylation of the substrates with M.SssI (CpG methyltransferase; NEB).

**Cell culture and transfection:** To study the activity of Dnmt1 and the Dnmt1 E572R/D575R mutant in cells, we used HCT116 Dnmt1 hypomorph cells, kindly provided by Prof. Bert Vogelstein (Johns Hopkins University, Baltimore).<sup>[17]</sup> The cells were cultivated in McCoy's 5A medium (#16600; Gibco) supplemented with FBS (10%), penicillin (100 U mL<sup>-1</sup>) and streptomycin (0.1 mg mL<sup>-1</sup>) at 37 °C. EYFP fusions of Dnmt1 wild type and E572R/D575R mutant were cloned into pcDNA3 for expression in mammalian cells yielding the pEYFP-Dnmt1 and pEYFP-Dnmt1 E572R/D575R constructs. The plasmids were transfected into HCT116 cells, and DNA methylation was analysed by LC-MS/MS. The pEYFP plasmid (encoding only EYFP) was used as a control. The HCT116 cells were seeded in a T25 flask, grown to 60–70% confluence, then transfected with Lipofectamine (Invitrogen/Life Technologies) by following the manufacturer's instructions. Briefly, plasmid (12 μg) was incubated with Lipofectamine (12 μL) at room temperature for 20 min, then transferred to the T25 flask containing the cells and incubated at 37 °C. Fresh medium was supplemented the next day; cells were incubated over four to five days with daily supplementing of the medium. Transfected cells were quantified by FACS analysis (Calibur system; BD Bioscience) equipped with 488 nm laser and CellQuest Pro software, as previously described.<sup>[26]</sup> Fluorescence intensity was measured in the FL1 channel (log scale, 1 to 10<sup>4</sup>; 10000 cells per sample). Non-transfected HCT116 Dnmt1 hypomorph cells were used as a control. Cells with a fluorescence intensity above 7 were considered YFP-positive. Genomic DNA was prepared by using the Blood and Cell Culture DNA Midi Kit (Qiagen). DNA extraction was performed by following the manufacturer's instructions for genomic DNA isolation from cell culture. DNA amounts were quantified spectroscopically at OD<sub>260</sub>.

**Sample preparation of genomic DNA for LC-MS/MS analysis:** Genomic DNA (0.5–1 μg) in H<sub>2</sub>O (25 μL) was digested by addition of ZnSO<sub>4</sub> (480 μM, 7.5 μL) containing *Aspergillus oryzae* nuclease S1 (42 U; Sigma-Aldrich), antarctic phosphatase (5 U, NEB) and specific amounts of HPLC-purified, labelled, internal standards (VWR), as previously described.<sup>[27]</sup> The mixture was incubated at 37 °C for 3 h in a Thermomixer comfort (Eppendorf). After addition of EDTA-Na<sub>2</sub>

(7.5 μL, 520 μM) containing snake venom phosphodiesterase I (0.2 U; USB Corporation/Affymetrix, Santa Clara, CA), the sample was incubated for another 3 h at 37 °C (total volume 40 μL). The sample was stored at –20 °C until analysis. Samples were filtered through an AcroPrep Advance 96 filter plate (0.2 μm Supor; Pall Life Sciences, Port Washington, NY) and then analysed by LC-MS/MS as described below.

**LC-ESI-MS/MS:** Acetonitrile (LC-MS grade) was purchased from Carl Roth; formic acid (p.a. grade for MS) was purchased from Fluka; water was purified by a Milli-Q Plus system (Merck Millipore); general HPLC and source settings for LC-ESI-MS/MS settings were as previously described.<sup>[27]</sup> Compound-dependent LC-MS/MS parameters for analysis of genomic DNA were as follows: time segment 1.5–4.0 min; precursor ion (*m/z*) 245.13 for [D3]-mC and 242.11 for mC. MS1 resolution: enhanced, product ion (*m/z*) 129.09 for [D3]-mC and 126.07 for mC, MS2 resolution: enhanced, dwell time 30 ms; collision energy 60 V; collision cell accelerator voltage 1 V; positive polarity.

**Note added in revision:** Recently, an independent study showed that mutations of residues at the interface of the RFTS and catalytic domains stimulate a Dnmt1 (291–1620) fragment as well.<sup>[29]</sup>

## Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft DFG (JE 252/15-1).

**Keywords:** DNA methylation · DNA methyltransferases · enzyme catalysis · enzyme specificity

- [1] A. Jeltsch, *ChemBioChem* **2002**, *3*, 274–293.
- [2] R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89–97.
- [3] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204–220.
- [4] S. Feng, S. E. Jacobsen, W. Reik, *Science* **2010**, *330*, 622–627.
- [5] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *ChemBioChem* **2011**, *12*, 206–222.
- [6] P. C. Taberlay, P. A. Jones, *Prog. Drug Res.* **2011**, *67*, 1–23.
- [7] Y. Bergman, H. Cedar, *Nat. Struct. Mol. Biol.* **2013**, *20*, 274–281.
- [8] S. Pradhan, A. Bacolla, R. D. Wells, R. J. Roberts, *J. Biol. Chem.* **1999**, *274*, 33002–33010.
- [9] A. Hermann, R. Goyal, A. Jeltsch, *J. Biol. Chem.* **2004**, *279*, 48350–48359.
- [10] R. Goyal, R. Reinhardt, A. Jeltsch, *Nucleic Acids Res.* **2006**, *34*, 1182–1188.
- [11] J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, *Science* **2011**, *331*, 1036–1040.
- [12] J. Song, M. Teplova, S. Ishibe-Murakami, D. J. Patel, *Science* **2012**, *335*, 709–712.
- [13] P. Bashtrykov, G. Jankevicius, A. Smarandache, R. Z. Jurkowska, S. Ragozin, A. Jeltsch, *Chem. Biol.* **2012**, *19*, 572–578.
- [14] P. Bashtrykov, S. Ragozin, A. Jeltsch, *FEBS Lett.* **2012**, *586*, 1821–1823.
- [15] K. Takeshita, I. Suetake, E. Yamashita, M. Suga, H. Narita, A. Nakagawa, S. Tajima, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9055–9059.
- [16] F. Syeda, R. L. Fagan, M. Wean, G. V. Avvakumov, J. R. Walker, S. Xue, S. Dhe-Paganon, C. Brenner, *J. Biol. Chem.* **2011**, *286*, 15344–15351.
- [17] I. Rhee, K. E. Bachman, B. H. Park, K.-W. Jair, R.-W. Yen, K. E. Schuebel, H. Cui, A. P. Feinberg, C. Lengauer, K. W. Kinzler, S. B. Baylin, B. Vogelstein, *Nature* **2002**, *416*, 552–556.
- [18] G. Egger, S. Jeong, S. G. Escobar, C. C. Cortez, T. W. Li, Y. Saito, C. B. Yoo, P. A. Jones, G. Liang, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 14080–14085.
- [19] A. J. Boettcher, J. Wu, C. Kim, J. Yang, J. Bruystens, N. Cheung, J. K. Penypacker, D. A. Blumenthal, A. P. Kornev, S. S. Taylor, *Structure* **2011**, *19*, 265–276.
- [20] C. Kim, D. Vigil, G. Anand, S. S. Taylor, *Eur. J. Cell Biol.* **2006**, *85*, 651–654.

- [21] M. Bostick, J. K. Kim, P.-O. Estève, A. Clark, S. Pradhan, S. E. Jacobsen, *Science* **2007**, *317*, 1760–1764.
- [22] J. Sharif, M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, H. Koseki, *Nature* **2007**, *450*, 908–912.
- [23] X. Liu, Q. Gao, P. Li, Q. Zhao, J. Zhang, J. Li, H. Koseki, J. Wong, *Nat. Commun.* **2013**, *4*, 1563.
- [24] M. Fatemi, A. Hermann, S. Pradhan, A. Jeltsch, *J. Mol. Biol.* **2001**, *309*, 1189–1199.
- [25] A. Jeltsch, T. Lanio in *Methods in Molecular Biology, Vol. 182: In Vitro Mutagenesis Protocols* (Ed.: J. Braman), Humana, Totowa, **2002**, pp. 85–94.
- [26] A. N. Siddique, S. Nunna, A. Rajavelu, Y. Zhang, R. Z. Jurkowska, R. Reinhardt, M. G. Rots, S. Ragozin, T. P. Jurkowski, A. Jeltsch, *J. Mol. Biol.* **2013**, *425*, 479–491.
- [27] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
- [28] P. Bashtrykov, G. Jankevicius, R. Z. Jurkowska, S. Ragozin, A. Jeltsch, *J. Biol. Chem.* **2013**; DOI: 10.1074/jbc.M113.528893.
- [29] A. C. Berkyurek, I. Suetake, K. Arita, K. Takeshita, A. Nakagawa, M. Shirakawa, S. Tajima, *J. Biol. Chem.* **2014**, *289*, 379–386.

---

Received: November 25, 2013

Published online on ■ ■ ■ ■, 0000

# **IX. Deamination, Oxidation, and C–C Bond Cleavage**

## **Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine**

S. Schiesser\*, T. Pfaffeneder\*, K. Sadeghian, **B. Hackner**, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld, T. Carell, *Journal of the American Chemical Society* **2013**, *135*, 14593-14599. (\* Equal contribution)

### **Introduction:**

Many pathways for active demethylation have been suggested as stated in chapter I.2.3. The C-C bond cleavage of the oxidized cytidine derivatives has been previously described in chapter VI and supports the idea of a decarboxylation pathway. In this manuscript the chemical requirements for decarboxylation are clarified and the efficiency determined.

### **Declaration of contribution:**

I synthesized and purified the  $^{15}\text{N}_2$ -dC standard necessary for the quantification of dC resulting from C-C bond cleavage and contributed to assembling the SI of the manuscript.

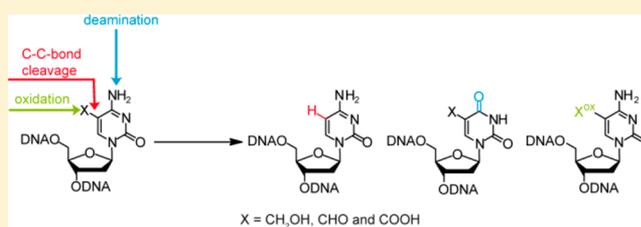
# Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine

Stefan Schiesser,<sup>†,‡</sup> Toni Pfaffeneder,<sup>†,‡</sup> Keyarash Sadeghian,<sup>†,‡</sup> Benjamin Hackner,<sup>†</sup> Barbara Steigenberger,<sup>†</sup> Arne S. Schröder,<sup>†</sup> Jessica Steinbacher,<sup>†</sup> Gengo Kashiwazaki,<sup>†</sup> Georg Höfner,<sup>§</sup> Klaus T. Wanner,<sup>§</sup> Christian Ochsenfeld,<sup>†,‡</sup> and Thomas Carell<sup>\*,†</sup>

<sup>†</sup>Center for Integrated Protein Science (CiPS<sup>M</sup>) at the Department of Chemistry, <sup>‡</sup>Chair for Theoretical Chemistry at the Department of Chemistry, <sup>§</sup>Center for Drug Research at the Department of Pharmacy, Ludwig-Maximilians-Universität München, Butenandtstrasse 5-13, 81377 Munich, Germany

## Supporting Information

**ABSTRACT:** Three new cytosine derived DNA modifications, 5-hydroxymethyl-2'-deoxycytidine (hmdC), 5-formyl-2'-deoxycytidine (fdC) and 5-carboxy-2'-deoxycytidine (cadC) were recently discovered in mammalian DNA, particularly in stem cell DNA. Their function is currently not clear, but it is assumed that in stem cells they might be intermediates of an active demethylation process. This process may involve base excision repair, C–C bond cleaving reactions or deamination of hmdC to 5-hydroxymethyl-2'-deoxyuridine (hmdU). Here we report chemical studies that enlighten the chemical reactivity of the new cytosine nucleobases. We investigated their sensitivity toward oxidation and deamination and we studied the C–C bond cleaving reactivity of hmdC, fdC, and cadC in the absence and presence of thiols as biologically relevant (organo)catalysts. We show that hmdC is in comparison to mdC rapidly oxidized to fdC already in the presence of air. In contrast, deamination reactions were found to occur only to a minor extent. The C–C bond cleavage reactions require the presence of high concentration of thiols and are acid catalyzed. While hmdC dehydroxymethylates very slowly, fdC and especially cadC react considerably faster to dC. Thiols are active site residues in many DNA modifying enzymes indicating that such enzymes could play a role in an alternative active DNA demethylation mechanism via deformylation of fdC or decarboxylation of cadC. Quantum-chemical calculations support the catalytic influence of a thiol on the C–C bond cleavage.

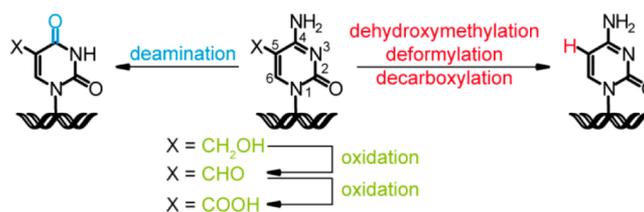


## INTRODUCTION

In addition to the canonical nucleosides dA, dC, dG, and dT, mammalian DNA contains 5-methyl-2'-deoxycytidine (mdC) and further dC derivatives that are generated from mdC by oxidation of the methyl group. The oxidizing enzymes are TET (ten eleven translocation) enzymes. These are  $\alpha$ -ketoglutarate dependent oxygenases, which oxidize 5-methyl-2'-deoxycytidine to 5-hydroxymethyl-2'-deoxycytidine (hmdC) and further to 5-formyl-2'-deoxycytidine (fdC) and 5-carboxy-2'-deoxycytidine (cadC).<sup>1–5</sup> The fate of these nucleosides and their function are currently controversially discussed. For hmdC for example it was postulated that the nucleoside may be deaminated *in vivo* to give 5-hydroxymethyl-2'-deoxyuridine (hmdU) by the action of special deaminases such as the AID (activation-induced deaminase)/APOBEC (apolipoprotein B mRNA-editing enzyme complex) protein family.<sup>6,7</sup> However, newer *in vitro* data suggest that deamination of hmdC is unlikely to occur enzymatically,<sup>8</sup> raising the question of whether the occurrence of hmdU may result from non-enzymatic spontaneous deamination of hmdC. Furthermore, since hmdC and fdC are known oxidative lesions of mdC<sup>9–11</sup> and because the levels of fdC and in particular cadC<sup>3–5,12</sup> are in the range of those reported for 8-oxo-dG,<sup>13</sup> which is a well-

known oxidative damage, one has to consider the possibility that they are formed by nonenzymatic oxidation processes

### Scheme 1. The dC Derivatives hmdC, fdC, and cadC Could Either Deaminate to dU Derivatives (Left), Undergo C–C bond Cleavage to dC (Right), or Oxidize



during DNA isolation and analysis.<sup>13,14</sup> Finally, while there is a possibility that enzymatic dehydroxymethylation of hmdC, deformylation of fdC, and decarboxylation of cadC might occur,<sup>15–19</sup> it cannot be ruled out that these processes occur

Received: April 11, 2013

Published: August 27, 2013

already to a significant extent without the help of an enzyme.<sup>18,20</sup>

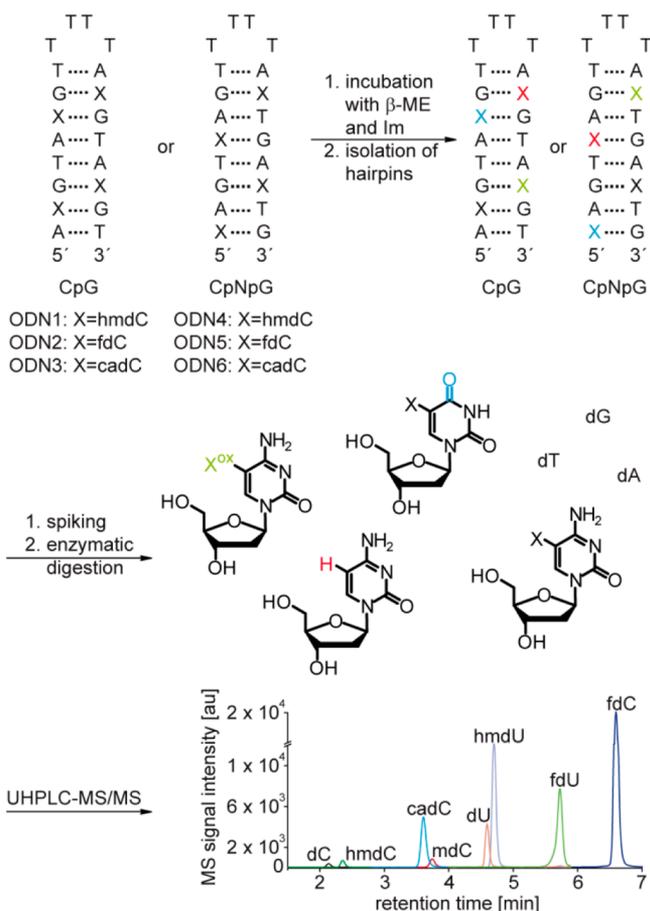
To differentiate between potential nonenzymatic background reactivities and enzyme-catalyzed processes, it is therefore essential to understand the intrinsic reactivity profile of the new nucleosides hmdC, fdC and cadC (Scheme 1).

## RESULTS AND DISCUSSION

Recently, others and us reported the synthesis of phosphoramidite and triphosphate building blocks for hmdC, fdC, and cadC and their incorporation into oligonucleotides using either solid phase phosphoramidite chemistry or by polymerase chain reaction (PCR).<sup>21–25</sup> For this study we synthesized hmdC, fdC, and cadC nucleosides and phosphoramidites and incorporated the latter into 20mer hairpin-oligodeoxynucleotides (ODN) with either hmdC (ODN1 and ODN4), fdC (ODN2 and ODN5), or cadC (ODN3 and ODN6) in a double-stranded xCpG or xCpNpG context using solid phase phosphoramidite chemistry (Scheme 2).

To investigate deamination-, oxidation-, and C–C bond cleavage reactions, the reactivities of the new epigenetic modifications were studied on the nucleoside level and in

**Scheme 2. Depiction of the Hairpins Used in This Study and of the Experimental Workflow<sup>a</sup>**

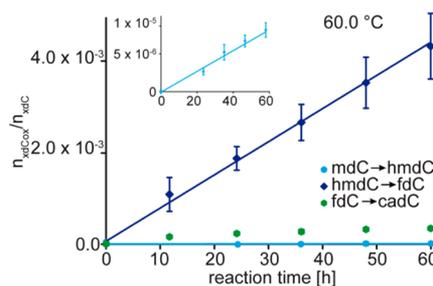


<sup>a</sup>After incubation of the xC-containing hairpins in  $\beta$ -mercaptoethanol ( $\beta$ -ME)/imidazole, the DNA was isolated. The corresponding isotopologues were added (omitted for clarity in the chromatogram), and the DNA was digested. Quantification of the reaction products was performed by UHPLC-MS/MS. Im = imidazole.

hairpin ODNs in buffer and in the presence of thiols. We investigated the reactivity with thiols in order to simulate potential enzymatic reactions, which often start by initial nucleophilic attack of the thiol at the C6 position of the pyrimidine. This then activates the corresponding C5 position (numbering see Scheme 1).<sup>18</sup> Examples are the DNA methyltransferases (DNMT), which convert dC to mdC.<sup>19</sup>

We first studied the reactivity of the different xC-nucleosides at pH 7.4 in buffer at different temperatures. We simultaneously quantified the various oxidation, deamination, and C–C bond cleavage reactions. Moreover, to elucidate how thiol-mediated C5 activation would influence the reactions, we performed studies with a systematic increase of  $\beta$ -mercaptoethanol ( $\beta$ -ME, 0 to 12 M). High thiol concentrations (12 M) were chosen to simulate the high effective molarity of, for example, the reactive cysteine moiety in the active sites of enzymes.<sup>17,18</sup> In all cases we performed product analysis using a UHPLC-MS/MS method with a triple quadrupole mass spectrometer (Scheme 2; for method development see Supporting Information [SI]). The developed method allows sensitive and accurate quantification of the whole product spectrum in one single analysis (12 min total run time). The exact quantification of the reaction products was conducted using the stable isotope dilution technique as described by others and us.<sup>12,13,26,27</sup> For the synthesis of the used isotopologues [<sup>15</sup>N<sub>2</sub>]-dC, [D<sub>2</sub>,<sup>15</sup>N<sub>2</sub>]-hmdC, [<sup>15</sup>N<sub>2</sub>]-fdC, [<sup>15</sup>N<sub>2</sub>]-cadC, [D<sub>3</sub>]-dT, [<sup>15</sup>N<sub>2</sub>]-dU, [D<sub>2</sub>]-hmdU, and [<sup>15</sup>N<sub>2</sub>]-fdU see SI.

**Oxidation Processes.** Our first question was how quickly air would oxidize mdC to hmdC, hmdC to fdC, and fdC to cadC. We dissolved the respective nucleosides in a phosphate buffer at pH = 7.4 and incubated the solution exposed to air for up to 60 h at 60.0, 67.5, 75.0, and 80.0 °C (Figure 1 and SI).



**Figure 1.** Oxidation kinetics of mdC to hmdC (cyan), hmdC to fdC (blue), and fdC to cadC (green) at 60.0 °C at pH 7.4. While hmdC is efficiently oxidized to fdC, fdC gives only a little cadC. Depicted are the means of triplicate experiments; error bars reflect the standard deviations. For details of the kinetic measurements at 67.5, 75.0, and 80.0 °C and linear regression analyses see SI.

For data analysis we assumed a pseudo-first order kinetic profile in which the oxygen concentration is a not rate limiting factor. Under these conditions we observed that mdC is oxidized only to a small extent to hmdC. In contrast, hmdC reacts efficiently to form fdC (Figure 1, blue line). Oxidation of fdC to cadC (Figure 1, green dots) is slow.

After 60 h at 60 °C, the yield of cadC was only 0.03%. Table 1 summarizes the determined pseudo-first-order rate constants. The data show that the formation of fdC by oxidation of hmdC is more than 2 orders of magnitude faster than the oxidation of mdC to hmdC. For the fdC to cadC oxidation, a rate constant

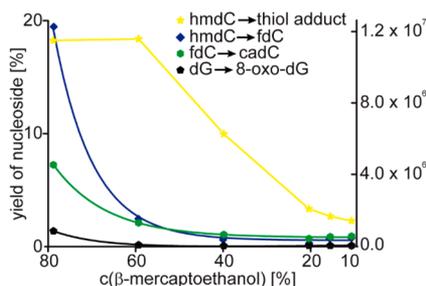
**Table 1.** Rate Constants  $k$  for the Oxidation of mdC to hmdC and hmdC to fdC at pH 7.4. For the Rate Constants at Higher Temperatures see SI<sup>a</sup>

	$k_{37\text{ }^\circ\text{C}}$ [ $\text{s}^{-1}$ ]	$k_{60\text{ }^\circ\text{C}}$ [ $\text{s}^{-1}$ ]	$E_a$ [ $\text{kJ mol}^{-1}$ ]
mdC→hmdC	n.d.	$4.7 \pm 0.6 \times 10^{-11}$	n.d.
hmdC→fdC	$1.3 \pm 0.1 \times 10^{-8}$	$2.2 \pm 0.2 \times 10^{-8}$	$20.8 \pm 2.2$

<sup>a</sup>n.d. = not determined.

could not be calculated, since the data deviated substantially from linearity (see Figure 1 and SI, Figure S2, Figure S11).

We next studied the oxidation reactions of hmdC and fdC in the hairpin-ODNs (Scheme 2). The reactions were performed without and with increasing amounts of  $\beta$ -ME (and imidazole at pH 5.0) in solution for 48 h at 50 °C (melting temperatures of hairpins:  $\geq 70$  °C; see SI, Table S1). The samples were desalted with a 0.025  $\mu\text{m}$  filter, spiked with the labeled internal standards, then fully digested with nuclease S1, snake venom phosphodiesterase, and antarctic phosphatase and analyzed by LC-MS/MS (Scheme 2). The amounts of the nucleoside products were normalized to the amount of dT [%]. Despite the shielding effect of the duplex environment, which could limit the reaction of the thiol at the C6 position, we measured increasing amounts of fdC with increasing concentrations of  $\beta$ -ME in ODN1 (hmdC). The fdC compound reached surprisingly high levels of up to 20% at 12 M  $\beta$ -ME (80% v/v  $\beta$ -ME; Figure 2, blue curve), showing that the C5–C6



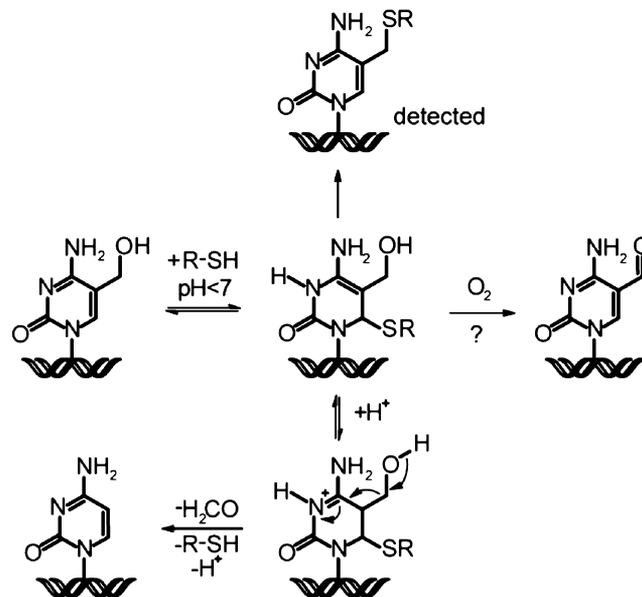
**Figure 2.** Investigation of the oxidation reactions of either hmdC or fdC in a hairpin-oligonucleotide (CpG-ODN 1, 2) with increasing concentration of  $\beta$ -mercaptoethanol/imidazole (pH 5.0, 50 °C, 48 h). Reaction yields (normalized to dT [%]) are plotted against the concentration of  $\beta$ -mercaptoethanol [% v/v]. The yellow data points show the intensity of the mass signal of the thiol adduct 5-((2''-hydroxyethyl)thio)methyl-dC (see top structure of Scheme 3), which was scaled to the right ordinate.

saturation by the thiol has a dramatic influence on the event of oxidation. In the absence of thiols we detected fdC at about 0.2% in ODN1 (48 h, 37 °C, pH 5.0). Other typical pyrimidine oxidation products were not detected regardless of the reaction conditions. We also monitored the levels of the well-established dG oxidation product 8-oxo-dG and noted here no significant level change (Figure 2, black curve), arguing for a thiol-catalyzed oxidation of hmdC to fdC.

Analyzing the mixture by mass spectrometry in more detail revealed the presence of 5-((2''-hydroxyethyl)thio)methyl-dC (Figure 2, yellow curve; top structure in Scheme 3), which shows that a 5-methylene intermediate may be formed during the reaction, which was first described by the Klimasauskas group.<sup>28</sup>

Treatment of ODN2 (fdC) at 80% v/v  $\beta$ -ME gave rise to the formation of the oxidized product cadC, but the reaction is

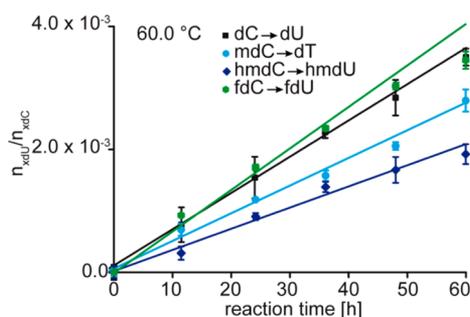
**Scheme 3.** Depiction of the Thiol-Catalyzed Oxidation of hmdC to fdC and the Proposed Mechanism of the C–C Bond Cleavage of hmdC to dC. R-SH =  $\beta$ -Mercaptoethanol R =  $\text{CH}_2\text{CH}_2\text{OH}$



comparatively slow (Figure 2, green curve). The formation of a dithioacetal of fdC was not detected (see SI). It should be noted that concomitant deformylation of fdC or decarboxylation of the oxidation product cadC to form dC in the presence of  $\beta$ -ME (see last section) may cause a slight underestimation of the oxidation rate.

In summary, experiments on the nucleoside and duplex level reveal that the oxidation of hmdC to fdC is a relatively fast process that is furthermore catalyzed by thiols. This result has to be taken into account when biological samples are investigated regarding the fdC levels.

**Deamination Reactions.** We next incubated the different nucleosides at pH 7.4 in water to investigate the deamination of dC, mdC, hmdC, and fdC to dU, dT, hmdU, and fdU (5-formyl-2'-deoxyuridine), respectively. The data are depicted in Figure 3. Clearly evident is that dC, mdC, hmdC, and fdC are deaminated under these conditions by about the same extent. Deamination of cadC was not detected. To obtain kinetic data at 37 °C, we determined the deamination rate constants at four



**Figure 3.** Deamination kinetics of dC to dU (black), mdC to dT (cyan), hmdC to hmdU (blue), and fdC to fdU (green) at 60.0 °C, pH 7.4. Depicted are the means of triplicate experiments; error bars reflect the standard deviations. For details of the kinetic measurements at 67.5, 75.0, and 82.5 °C and linear regression analyses see SI.

different temperatures (see SI) and extrapolated the pseudo-first-order deamination rates to 37 °C. The data are compiled in Table 2.

**Table 2. Rate Constants  $k$  and Activation Energies  $E_a$  for the Deamination of dC, mdC, hmdC, and fdC at pH 7.4**

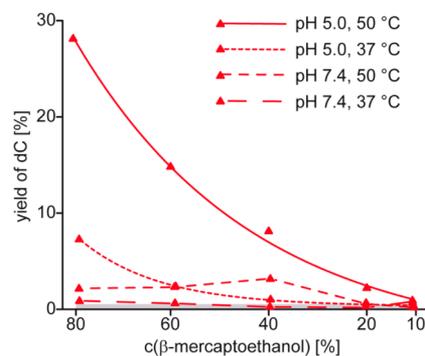
	$k_{37\text{ °C}}$ [ $s^{-1}$ ]	$E_a$ [ $kJ\ mol^{-1}$ ]
dC→dU	$9.4 \pm 0.5 \times 10^{-10}$	$108.7 \pm 1.9$
mdC→dT	$7.8 \pm 0.3 \times 10^{-10}$	$105.0 \pm 2.5$
hmdC→hmdU	$5.8 \pm 0.8 \times 10^{-10}$	$104.8 \pm 3.9$
fdC→fdU	$1.2 \pm 0.2 \times 10^{-9}$	$102.2 \pm 2.4$

At 37 °C, the deamination rates of dC, mdC, hmdC, and fdC to form their corresponding 2'-deoxyuridine derivatives are approximately the same with  $(6-12) \times 10^{-10}\ s^{-1}$  on the nucleoside level. The determined rate constant and activation energy for the deamination of dC are in good agreement with those reported for single-stranded DNA.<sup>29</sup> Bearing in mind that the rates are more than 2 orders of magnitude lower in double-stranded DNA,<sup>29,30</sup> our data argue that spontaneous deamination of hmdC, fdC, and cadC should be a negligible background reaction in comparison to the oxidation of hmdC to form fdC ( $k_{37\text{ °C}} = 1.3 \pm 0.1 \times 10^{-8}\ s^{-1}$ ;  $E_a = 20.8 \pm 2.2\ kJ\ mol^{-1}$ ). If, consequently, significant amounts of deaminated compounds are detected, we conclude that these are likely derived from an enzymatic process.

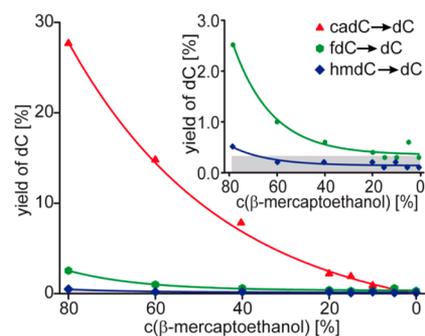
**C–C bond Cleaving Reactions.** To investigate the reaction of hmdC, fdC, and cadC to form dC we heated the nucleosides for 60 h in water at pH 7.4 at 60 and 80 °C. At 60 °C we detected only traces of dC and the obtained reaction rates at 80 °C were slow (SI, Figure S9 and Table 3), showing that uncatalyzed C–C bond cleavage reactions can be neglected.

This picture changes in the presence of thiols, which we recently reported to catalyze the decarboxylation of cadC.<sup>18</sup> To investigate the conditions of the thiol-mediated decarboxylation reaction in more detail, we treated the cadC-containing ODN3 with increasing concentrations of  $\beta$ -ME for 48 h at different pH-values and different temperatures in the presence of imidazole.<sup>18</sup> The oligonucleotides were isolated and analyzed by LC–MS/MS as outlined in Scheme 2. The results of the experiments are depicted in Figure 4. Clearly evident is that under the investigated conditions (pH 5.0 and 50 °C) decarboxylation of cadC is a relatively efficient reaction. The yield of dC increased along with the concentration of  $\beta$ -ME up to 28%. A lower temperature of 37 °C and a higher pH-value of 7.4 resulted in a strong reduction of the dC yield, which shows that the decarboxylation reaction is an activated proton catalyzed reaction. We next performed analogous experiments with hmdC and fdC (pH 5.0, 50 °C, 48 h). Figure 5 compares the C–C bond cleaving yields of all three nucleosides embedded in the hairpin duplex structures.

We noticed that hmdC dehydroxymethylates to give dC. However, the obtained yield of dC was very small (0.5%). Far higher yields were reported by the Klimasauskas group, who applied a mutated DNA methyltransferase and by the Sowers



**Figure 4.** Investigation of the C–C bond cleavage reaction of cadC to dC in a hairpin-oligonucleotide (CpG-ODN 3) in a  $\beta$ -mercaptoethanol/imidazole mixture. Depicted are the reaction yields (normalized to dT [%]) depending on the concentration of  $\beta$ -mercaptoethanol [% v/v] at pH = 5.0 or 7.4 and at 37 or 50 °C after 48 h. The gray area reflects the limit of quantification.



**Figure 5.** Investigation of the C–C bond cleavage reactions of either hmdC, fdC, or cadC in a hairpin-oligonucleotide (CpG-ODN 1, 2, 3) in a  $\beta$ -mercaptoethanol/imidazole mixture (pH 5.0, 50 °C, 48 h). Depicted are the reaction yields (normalized to dT [%]) of hmdC (blue), fdC (green), and cadC (red) to dC depending on the concentration of  $\beta$ -mercaptoethanol [% v/v]. The gray area in the inset shows the limit of quantification.

group, who used photohydration conditions.<sup>16,17,20</sup> For fdC, in contrast, we observed considerable deformylation and obtained yields of up to 2.5%. Decarboxylation of cadC is most efficient with an obtained yield of up to 28%.

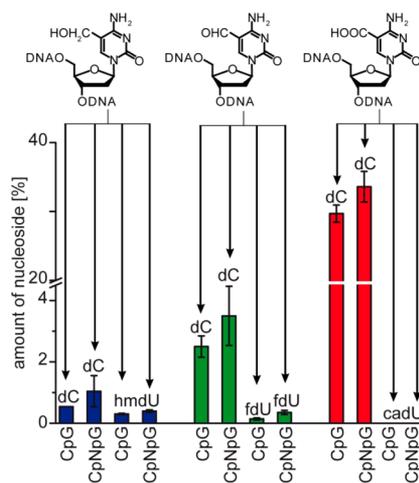
We next studied how different sequences would affect the C–C bond cleaving and deamination reactions and repeated the experiments with ODNs 4–6 which feature the xdC-derivatives in a non-CpG context (Figure 6). We observed only a small reactivity difference in these sequences compared to the CpG-ODNs. The yields of deamination products were found to be lower than 0.4% under these conditions.

The data show that both fdC and cadC can undergo C–C bond cleavage reactions mediated by thiols. Dehydroxymethylation of hmdC is in contrast a considerably slower process.

To gain deeper insights into the thiol catalysis of the C–C bond cleavage reactions, we finally computed the reaction energies using quantum-chemical methods. Computational details are described in the SI. Carboxylated nucleobases (caC and caU) were capped at the N1 position with a methyl

**Table 3. Rate Constants for the Non-Thiol-Mediated C–C Bond Cleavage of hmdC, fdC, and cadC to dC at pH 7.4 at 80 °C**

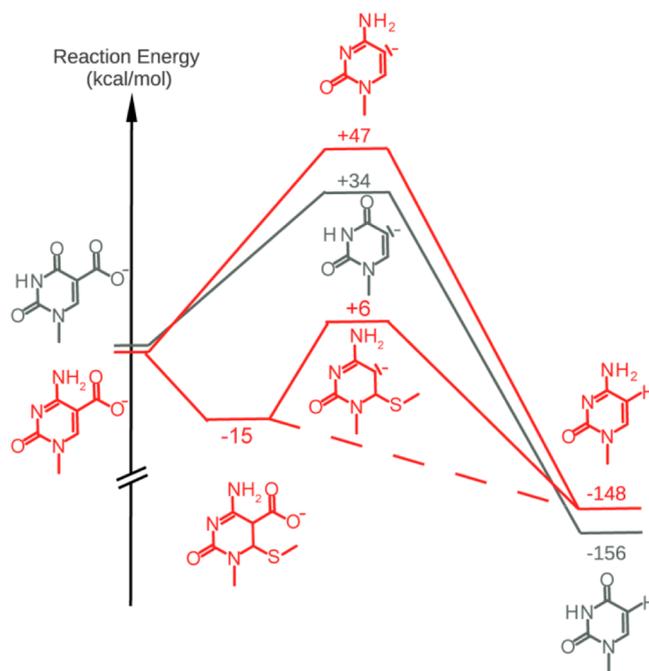
	hmdC→dC	fdC→dC	cadC→dC
$k_{80\text{ °C}}$ [ $s^{-1}$ ]	$1.3 \pm 0.2 \times 10^{-9}$	$7.3 \pm 1.1 \times 10^{-9}$	$6.9 \pm 0.7 \times 10^{-9}$



**Figure 6.** The C–C-bond cleavage of hmdC, fdC, and cadC to dC is almost independent of the sequence context. Investigation of the C–C bond cleavage and deamination reactions of either hmdC, fdC, or cadC in a hairpin-oligonucleotide (CpG-ODN 1, 2, 3; CpNpG-ODN 4, 5, 6) in a 80% (v/v)  $\beta$ -mercaptoethanol/imidazole mixture (pH 5.0, 50 °C, 48 h). Depicted are the reaction yields (normalized to dT [%]) of hmdC (blue), fdC (green), and cadC (red) to dC as well as the corresponding deamination products hmdU and fdU. Depicted are the means of triplicate experiments; error bars reflect the standard deviations.

group. To describe the explicit solvent–solute hydrogen bonds, five water molecules were included in the study. Ideally, one would like to describe more of the long-range electrostatic solute–solvent interactions by including more water molecules in the calculations. This would, however, mean a computational effort which is beyond the scope of this work. A crude way to approximate the influence of the continuum is to use an implicit solvent model. Here, we have performed calculations with an implicit solvent cavity using the COSMO-model<sup>31</sup> (data shown in SI). Although the energetics are clearly affected, the overall trend remains the same. Triple-zeta basis sets<sup>32</sup> were used throughout the calculations. RI-MP2<sup>32,33</sup> reaction energies, obtained using the DFT/B3LYP-D3<sup>34–36</sup> energy optimized structures, are depicted in Figure 7. No transition state search was carried out as the reaction rates are already obtained from the experimental data presented above.

We first computed the direct decarboxylation of isoorotate (caU), a reaction that is catalyzed by the enzyme isoorotate decarboxylase (IDCase).<sup>37,38</sup> We assumed in our study a direct decarboxylation via formation of a vinyl anion type intermediate. Similar mechanistic ideas were the basis of a recent detailed mechanistic and structural study of the IDCase.<sup>39</sup> We obtained by our calculations a rather high energy of +34 kcal mol<sup>-1</sup> for the vinyl anion of U (gray intermediate in Figure 7). In agreement with our results from the thiol-free reaction conditions, our data show that a direct decarboxylation mechanism for caC is unlikely. In comparison to the vinyl anion of U, the energy of the vinyl anion of C is with +47 kcal mol<sup>-1</sup> significantly higher. This may in part explain the observed weak activity<sup>39</sup> of IDCase to decarboxylate ScaC to C. In contrast, the energetics of the thiol addition at the C6 position of caC and the subsequent decarboxylation are much more favorable: Calculations predict an only slightly endothermic reaction energy of +6 kcal mol<sup>-1</sup> for the decarboxylation of the thiol-reacted anionic intermediate. Overall, the reaction of the C6 position with the thiol reduces



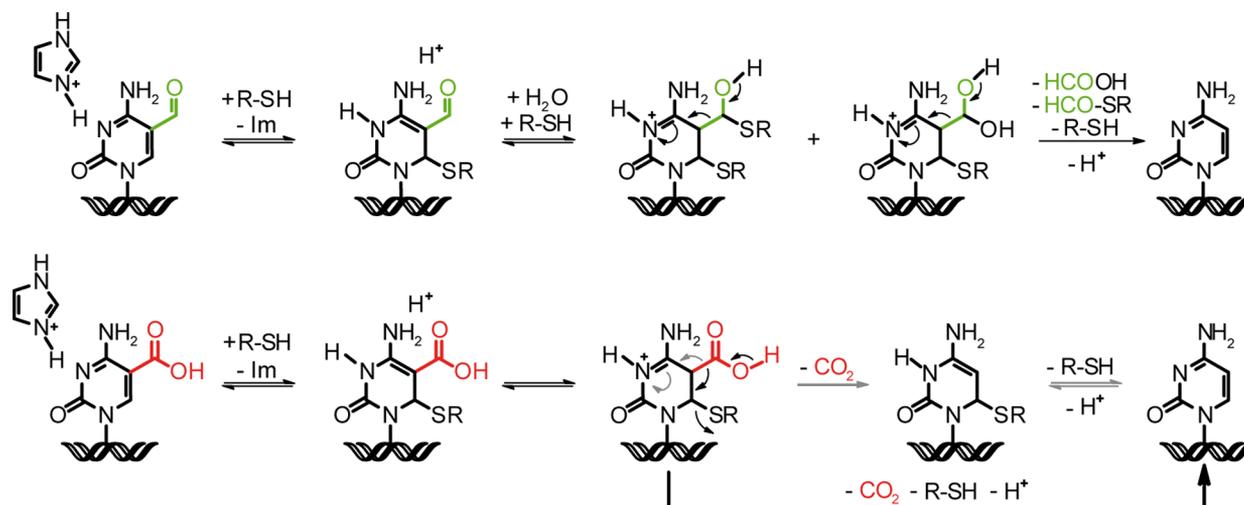
**Figure 7.** Schematic representation of different decarboxylation pathways. Depicted are the reaction energies obtained from quantum chemical calculations.

the energy of the corresponding anionic intermediate by more than 40 kcal mol<sup>-1</sup> (Figure 7, intermediates shown in red). This explains why the reaction of cadC with a thiol leads to fast decarboxylation. We also noted during the computational study that simultaneous decarboxylation and thiol elimination is an even more favorable process. This reaction pathway is symbolized with dotted lines in Figure 7 and illustrated in Scheme 4. In summary, the calculations support the influence of the thiol catalysis.

## CONCLUSIONS

We show that the new nucleosides hmdC, fdC, and cadC possess an increased reactivity compared to mdC and dC. First, the intrinsic deamination rates of hmdC and fdC are comparable to those of mdC and dC. Second, hmdC is more susceptible to oxidation. It reacts surprisingly quickly to fdC if exposed to atmospheric oxygen, which is a problem that needs to be considered when DNA is isolated from biological material for the determination of fdC. In addition, the reaction of hmdC to fdC is accelerated in the presence of thiols. If fdC is exposed to atmospheric oxygen further oxidation to cadC is a rather inefficient process.

Third and importantly, fdC and cadC can undergo thiol-mediated and acid-catalyzed C–C bond cleavage reactions to form dC under release of formic acid and CO<sub>2</sub>, respectively (see Scheme 4). Here decarboxylation is by a factor of 11 more efficient than the deformylation of fdC. If we consider that DNA demethylation requires stepwise oxidation of hmdC to fdC and cadC, both deformylation of fdC and decarboxylation of cadC could take place via alternative active demethylation mechanisms.

Scheme 4. Proposed Mechanisms of the Thiol-Mediated and Acid-Catalyzed C–C Bond Cleavage Reactions of fdC and cadC to dC<sup>a</sup>

<sup>a</sup>The reactions are thought to proceed via a covalent enamine intermediate. The deformylation reaction of fdC further requires the addition of a nucleophile (water or thiol) to the aldehyde before the release of formic acid (or the thiol ester) can proceed. In principal, the release of formic acid from fdC or carbon dioxide from cadC and thiol elimination could occur simultaneously or in two steps (shown by gray or black arrows, respectively). R-SH =  $\beta$ -mercaptoethanol; Im = imidazole.

## ■ ASSOCIATED CONTENT

### Supporting Information

Experimental procedures and spectroscopic data of standards for the UHPLC-MS/MS analysis, detailed procedures for mass spectrometric measurements, and method development. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

Thomas.Carell@cup.uni-muenchen.de

### Author Contributions

<sup>†</sup>S.S. and T.P. contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft via SFB749 (TPA4/C7) and the Excellence cluster EXC114 (CiPS<sup>M</sup>). S.S., T.P., and A.S.S. thank the Fonds der Chemischen Industrie for predoctoral fellowships. K.S. and C.O. thank Dr. Denis Flaig for helpful discussions. G.K. thanks the Japan Society for the Promotion of Science (JSPS) as a JSPS Postdoctoral Fellow for Research Abroad.

## ■ REFERENCES

- (1) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
- (2) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
- (3) Pfaffeneder, T.; Hackner, B.; Truß, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeyer, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008.
- (4) He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C.-X.; Zhang, K.; He, C.; Xu, G.-L. *Science* **2011**, *333*, 1303.
- (5) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.

(6) Guo, J. U.; Su, Y.; Zhong, C.; Ming, G.-L.; Song, H. *Cell* **2011**, *145*, 423.

(7) Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D.; Abramowitz, L. K.; Bartolomei, M. S.; Rambow, F.; Bassi, M. R.; Bruno, T.; Fanciulli, M.; Renner, C.; Klein-Szanto, A. J.; Matsumoto, Y.; Kobi, D.; Davidson, I.; Alberti, C.; Larue, L.; Bellacosa, A. *Cell* **2011**, *146*, 67.

(8) Nabel, S. N.; Jia, H.; Ye, Y.; Shen, L.; Goldschmidt, H. L.; Stivers, J. T.; Zhang, Y.; Kohli, R. M. *Nat. Chem. Biol.* **2012**, *8*, 751.

(9) Bienvenu, C.; Wagner, J. R.; Cadet, J. *J. Am. Chem. Soc.* **1996**, *118*, 11406.

(10) Cao, H.; Wang, Y. *Nucleic Acids Res.* **2007**, *35*, 4833.

(11) Wagner, J. R.; Cadet, J. *Acc. Chem. Res.* **2010**, *43*, 564.

(12) Liu, S.; Wang, J.; Su, Y.; Guerrero, C.; Zeng, Y.; Mitra, D.; Brooks, P. J.; Fisher, D. E.; Song, H.; Wang, Y. *Nucleic Acids Res.* **2013**, *41*, 6421.

(13) Taghizadeh, K.; McFaline, J. L.; Pang, B.; Sullivan, M.; Dong, M.; Plummer, E.; Dedon, P. C. *Nat. Protoc.* **2008**, *3*, 1287.

(14) Ravanat, J.-L.; Douki, T.; Duez, P.; Gremaud, E.; Herbert, K.; Hofer, T.; Lasserre, L.; Saint-Pierre, C.; Favier, A.; Cadet, J. *Carcinogenesis* **2002**, *23*, 1911.

(15) Globisch, D.; Münzel, M.; Müller, M.; Michalak, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. *PLoS One* **2010**, *5*, e15367 DOI: 10.1371/journal.pone.0015367.

(16) Hamm, S.; Just, G.; Lacoste, N.; Moitessier, N.; Szyf, M.; Mamer, O. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1046.

(17) Liutkeviciute, Z.; Lukinavicius, G.; Masevicius, V.; Daujotyte, D.; Klimasauskas, S. *Nat. Chem. Biol.* **2009**, *5*, 400.

(18) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeyer, C.; Truss, M.; Carell, T. *Angew. Chem., Int. Ed.* **2012**, *51*, 6516.

(19) Wu, S. C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607.

(20) Privat, E.; Sowers, L. C. *Chem. Res. Toxicol.* **1996**, *9*, 745.

(21) Tardy-Planechaud, S.; Fujimoto, J.; Lin, S. S.; Sowers, L. C. *Nucleic Acids Res.* **1997**, *25*, 553.

(22) Dai, Q.; He, C. *Org. Lett.* **2011**, *13*, 3446.

(23) Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chem.—Eur. J.* **2011**, *17*, 13782.

(24) Steigenberger, B.; Schiesser, S.; Hackner, B.; Brandmayr, C.; Laube, S. K.; Steinbacher, J.; Pfaffeneder, T.; Carell, T. *Org. Lett.* **2013**, *15*, 366.

- (25) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671.
- (26) Münzel, M.; Globisch, D.; Brückl, T.; Wagner, M.; Welzmler, V.; Michalakakis, S.; Müller, M.; Biel, M.; Carell, T. *Angew. Chem., Int. Ed.* **2010**, *49*, 5375.
- (27) Spruijt, Cornelia G.; Gnerlich, F.; Smits, Arne H.; Pfaffeneder, T.; Jansen, Pascal W. T. C.; Bauer, C.; Münzel, M.; Wagner, M.; Müller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, Arie B.; Lephikov, K.; Müller, U.; Walter, J.; Boelens, R.; van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M. *Cell* **2013**, *152*, 1146.
- (28) Liutkevičiūtė, Z.; Kriukienė, E.; Grigaitytė, I.; Masevičius, V.; Klimašauskas, S. *Angew. Chem., Int. Ed.* **2011**, *123*, 2138.
- (29) Frederico, L. A.; Kunkel, T. A.; Shaw, B. R. *Biochemistry* **1990**, *29*, 2532.
- (30) Shen, J.-C.; Rideout, W. M., III; Jones, P. A. *Nucleic Acids Res.* **1994**, *22*, 972.
- (31) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, *5*, 799.
- (32) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143.
- (33) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359.
- (34) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1998**, *37*, 785.
- (35) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (36) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456.
- (37) Smiley, J. A.; Angelot, J. M.; Cannon, R. C.; Marshall, E. M.; Asch, D. K. *Anal. Biochem.* **1999**, *266*, 85.
- (38) Palmatier, R. D.; McCroskey, R. P.; Abbott, M. T. *J. Biol. Chem.* **1970**, *245*, 6706.
- (39) Xu, S.; Li, W.; Zhu, J.; Wang, R.; Li, Z.; Xu, G.-L.; Ding, J. *Cell Res.* **2013**, DOI: 10.1038/cr.2013.107.

## **X. Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base", are associated with anaplasia in human brain tumors**

T. F. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Munzel, M. Muller, T. Pfaffeneder, **B. Hackner**, W. Feiden, U. Schuller, T. Carell, H. A. Kretzschmar, *International Journal of cancer* **2012**, 131, 1577-1590.

### **Introduction:**

Methylation of cytidine to mC plays a role in several biological processes including tumor biology. The TET enzymes oxidize mC to hmC utilizing Fe<sup>II</sup> and 2-oxoglutarate. This process is suggested to play a role in active DNA demethylation and other biological processes as described in chapter I.2.3.3.2. Aberrant levels of mC are associated with cancer and other severe diseases. In this manuscript the levels of hmC in tumor and healthy tissue are analyzed to determine whether or not they are associated with the disease.

### **Declaration of contribution:**

I performed the LC-MS analysis and contributed to the necessary DNA isolation of several tumor samples.

# Low values of 5-hydroxymethylcytosine (5hmC), the “sixth base,” are associated with anaplasia in human brain tumors

Theo F.J. Kraus<sup>1\*</sup>, Daniel Globisch<sup>2\*</sup>, Mirko Wagner<sup>2</sup>, Sabina Eigenbrod<sup>1</sup>, David Widmann<sup>1</sup>, Martin Münzel<sup>2</sup>, Markus Müller<sup>2</sup>, Toni Pfaffeneder<sup>2</sup>, Benjamin Hackner<sup>2</sup>, Wolfgang Feiden<sup>3</sup>, Ulrich Schüller<sup>1</sup>, Thomas Carell<sup>2</sup> and Hans A. Kretzschmar<sup>1</sup>

<sup>1</sup> Center for Neuropathology and Prion Research (ZNP), Ludwig-Maximilians-University, Munich, Germany

<sup>2</sup> Department of Chemistry, Center for Integrated Protein Science (CIPS<sup>M</sup>), Ludwig-Maximilians-University, Munich, Germany

<sup>3</sup> Institute for Neuropathology, University Hospital Homburg, Homburg, Germany

**5-Methylcytosine (5mC) in genomic DNA has important epigenetic functions in embryonic development and tumor biology. 5-Hydroxymethylcytosine (5hmC) is generated from 5mC by the action of the TET (Ten-Eleven-Translocation) enzymes and may be an intermediate to further oxidation and finally demethylation of 5mC. We have used immunohistochemistry (IHC) and isotope-based liquid chromatography mass spectrometry (LC-MS) to investigate the presence and distribution of 5hmC in human brain and brain tumors. In the normal adult brain, IHC identified 61.5% 5hmC positive cells in the cortex and 32.4% 5hmC in white matter (WM) areas. In tumors, positive staining of cells ranged from 1.1% in glioblastomas (GBMs) (WHO Grade IV) to 8.9% in Grade I gliomas (pilocytic astrocytomas). In the normal adult human brain, LC-MS also showed highest values in cortical areas (1.17% 5hmC/dG [deoxyguanosine]), in the cerebral WM we measured around 0.70% 5hmC/dG. 5hmC levels were related to tumor differentiation, ranging from lowest values of 0.078% 5hmC/dG in GBMs (WHO Grade IV) to 0.24% 5hmC/dG in WHO Grade II diffuse astrocytomas. 5hmC measurements were unrelated to 5mC values. We find that the number of 5hmC positive cells and the amount of 5hmC/dG in the genome that has been proposed to be related to pluripotency and lineage commitment in embryonic stem cells is also associated with brain tumor differentiation and anaplasia.**

Epigenetics, the study of mechanisms that control gene expression (in a potentially heritable way), may be the most rapidly expanding field in tumor biology. On a molecular level, (i) DNA methylation, *i.e.*, covalent modification of cytosine bases resulting in 5-methylcytosine (5mC), (ii) histone modifications and (iii) nucleosome positioning are regarded as the driving epigenetic mechanisms. They are fundamental to the regulation of many cellular processes, including gene and micro

RNA expression, DNA-protein interactions, suppression of transposable element mobility, cellular differentiation, embryogenesis, X-chromosome inactivation and genomic imprinting.

In tumor biology, DNA methylation is the best-studied epigenetic change. Epigenetic silencing of O<sup>6</sup>-methylguanine DNA-methyltransferase (MGMT)<sup>1</sup> has been described as a strong predictive factor of treatment response to chemotherapy with alkylating agents of glioblastoma (GBM) and anaplastic astrocytoma (AA).<sup>2</sup> Methylation of CpG islands in the MGMT promoter with ensuing repression of MGMT transcriptional activity is generally viewed as the cause for this correlation.<sup>3</sup>

Oxidation of 5mC leading to 5-hydroxymethylcytosine (5hmC) has been identified as a new epigenetic phenomenon in mouse Purkinje cells.<sup>4</sup> Three possible modes of action of 5hmC were discussed. (i) It might influence chromatin structure and local transcriptional activity by recruiting selective 5hmC-binding proteins or excluding 5mC-binding proteins. (ii) Conversion of 5mC to 5hmC might facilitate passive DNA demethylation by excluding the maintenance DNMT1, which recognizes 5hmC poorly. (iii) 5hmC may be an intermediate in a pathway of active DNA demethylation either by conversion to cytosine under certain conditions or by replacement of 5hmC by specific DNA repair mechanisms. Most recent findings support the hypothesis that 5hmC is an intermediate in a pathway for active DNA demethylation.<sup>5–7</sup> The TET proteins, identified as a new family of enzymes that

**Key words:** epigenetics, brain tumors, 5-hydroxymethylcytosine, 5hmC, glioma, ependymoma, meningioma, IDH, TET

**Grant sponsor:** DFG Normalverfahren; **Grant numbers:** CA275/8-4, SFB 749, SFB 646; **Grant sponsor:** Fonds der Chemischen Industrie

**DOI:** 10.1002/ijc.27429

**History:** Received 18 Oct 2011; Accepted 29 Dec 2011; Online 10 Jan 2012

\*T.F.J.K. and D.G. contributed equally to this work.

**Correspondence to:** Hans A. Kretzschmar, Center for Neuropathology and Prion Research (ZNP), Ludwig-Maximilians-University, 81377 Munich, Germany, E-mail: hans.kretzschmar@med.uni-muenchen.de or Thomas Carell, Center for Integrated Protein Science (CIPS<sup>M</sup>), Department of Chemistry, Ludwig-Maximilians-University, 81377 Munich, Germany, E-mail: thomas.carell@cup.uni-muenchen.de

**Table 1.** Twenty-two tissue samples were selected from nine control cases from the BrainBank Munich

Case	Sample	Age	Gender	Postmortem time (h)	Region	Cause of death	5hmC positive cells (%)	5mC/dG (%)	5hmC/dG (%)
1	a	85	F	20	Occipital cortex	Cardiac arrest	55.05	5.16	1.17
1	b	85	F	20	Occipital white matter	Cardiac arrest	19.07	4.26	0.76
1	c	85	F	20	Frontal cortex	Cardiac arrest	70.29	–	–
1	d	85	F	20	Frontal white matter	Cardiac arrest	23.27	–	–
2	a	61	M	24	Frontal cortex	Cardiac arrest	55.46	5.05	1.06
2	b	61	M	24	Frontal white matter	Cardiac arrest	16.45	–	–
3	a	87	M	48	Frontal cortex	Cardiac arrest	54.32	5.94	1.13
3	b	87	M	48	Frontal white matter	Cardiac arrest	22.85	–	–
3	c	87	M	48	Occipital cortex	Cardiac arrest	54.70	5.64	1.14
3	d	87	M	48	Occipital white matter	Cardiac arrest	23.08	4.23	0.64
4	a	46	M	n.a.	Occipital cortex	Cardiac arrest	64.29	5.74	1.12
4	b	46	M	n.a.	Occipital white matter	Cardiac arrest	36.34	4.36	0.54
5	a	84	M	19	Frontal white matter	Cardiac arrest	28.57	3.82	0.78
5	b	84	M	19	Frontal cortex	Cardiac arrest	59.92	4.55	1.19
6	a	61	M	n.a.	Frontal white matter	Cardiac arrest	61.61	4.14	0.76
6	b	61	M	n.a.	Frontal cortex	Cardiac arrest	70.33	5.35	1.35
7	a	26	F	n.a.	Frontal cortex	Hemorrhagic pericard effusion	80.17	–	–
7	b	26	F	n.a.	Frontal white matter	Hemorrhagic pericard effusion	58.06	–	–
8	a	41	M	n.a.	Frontal cortex	Multiple organ failure	65.02	–	–
8	b	41	M	n.a.	Frontal white matter	Multiple organ failure	38.70	–	–
9	a	39	M	n.a.	Frontal cortex	Haematothorax	46.75	–	–
9	b	39	M	n.a.	Frontal white matter	Haematothorax	30.00	–	–

alter the methylation status of DNA, are 2-oxoglutarate (2OG)- and Fe(II)-dependent enzymes that catalyze the conversion of 5mC to 5hmC,<sup>8</sup> 5-formylcytosine and 5-carboxylcytosine.<sup>5–7</sup> TET proteins and 5hmC have been reported in various tissues and both are tightly regulated during embryonic stem cells (ESC) differentiation.<sup>9</sup> 5hmC content has been described as significantly reduced in stem cell/progenitor cell compartments and in human cancers (prostate, breast and colon).<sup>10</sup>

To obtain a first insight into a possible role of 5hmC in the human brain, we used isotope-based liquid chromatography mass spectrometry (LC-MS) to measure 5hmC concentrations in various cortical and white matter regions of the adult human brain to elucidate whether hydroxymethylation may be present in human brain tumors, we measured 5hmC content in slowly growing frequent brain tumors, *i.e.*, astrocytomas WHO Grades I, II and rapidly growing AAs WHO Grade III and GBMs (WHO Grade IV)<sup>11</sup> and other brain tumors. In addition, we used immunohistochemistry (IHC) to investigate the distribution of 5hmC in the brain and in brain tumor cells.

## Material and Methods

### Sample selection normal brain regions

To evaluate the amount of 5hmC in human brain, we selected tissue of nine donors that had provided their brains

for research purposes after death. The samples were obtained from the Brain Bank Munich. As target regions we selected frontal cortex, occipital cortex, frontal white matter and occipital white matter. The tissue samples had either been formalin-fixed and paraffin-embedded (FFPE) or stored at  $-80^{\circ}\text{C}$ . For this study, we chose 22 human control samples, well characterized by age and gender (Table 1). The age of the patients varied from 26 to 87 with a mean age of 59 years.

### Sample selection tumors

To evaluate 5hmC in human brain tumors, we used FFPE and frozen material that had been stored at  $-80^{\circ}\text{C}$ . The samples were obtained from the Brain Tumor Bank, Center for Neuropathology, LMU Munich, as well as from the Brain Tumor Bank of the Institute for Neuropathology, University Hospital Homburg. For this study, we selected areas consisting of more than 75% of tumor cells as estimated from freshly made and H&E stained sections. In each case we separated clearly visible tumor areas. One hundred seventeen human tumor samples derived from 117 different tumors were selected, including 80 astrocytomas and GBMs, 23 ependymomas and 14 meningiomas (Table 2). The age of the patients ranged from 3 to 81 years with a mean of 46 years and the male to female ratio was 67:43. In the astrocytoma

**Table 2.** One hundred seventeen tumor samples were selected from the Brain Tumour Bank, Center for Neuropathology, LMU Munich and the Brain Tumour Bank of the Institute for Neuropathology, University Hospital Homburg

Case	Age	Gender	Tumor	Grade	Region	5hmC positive cells (%)	5mC/dG (%)	5hmC/dG (%)	IDH status
1	15	M	Pilocytic astrocytoma (WHO Grade I)	I	Temporal	13.66	4.75	0.18	No mutation
2	23	F	Pilocytic astrocytoma (WHO Grade I)	I	Thoracic spine	10.83	4.10	0.24	No mutation
3	23	M	Pilocytic astrocytoma (WHO Grade I)	I	n.a.	7.69	4.58	0.24	No mutation
4	12	F	Pilocytic astrocytoma (WHO Grade I)	I	Rhombencephalon	10.03	4.99	0.17	No mutation
5	9	M	Pilocytic astrocytoma (WHO Grade I)	I	Cerebellum	1.66	4.98	0.12	No mutation
6	21	M	Pilocytic astrocytoma (WHO Grade I)	I	Temporal	9.57	–	–	No mutation
7	34	M	Diffuse astrocytoma (WHO Grade II)	II	Fronto-temporal	3.75	4.66	0.17	Mutation
8	35	M	Diffuse astrocytoma (WHO Grade II)	II	Insula	1.47	–	–	Mutation
9	36	F	Diffuse astrocytoma (WHO Grade II)	II	Occipital	0.72	–	–	Mutation
10	15	M	Diffuse astrocytoma (WHO Grade II)	II	Thalamus	3.35	–	–	No mutation
11	28	F	Diffuse astrocytoma (WHO Grade II)	II	Precentral	1.55	–	–	Mutation
12	n.a.	n.a.	Diffuse astrocytoma (WHO Grade II)	II	n.a.	–	4.50	0.09	Mutation
13	n.a.	n.a.	Diffuse astrocytoma (WHO Grade II)	II	n.a.	–	4.57	0.24	Mutation
14	44	M	Diffuse astrocytoma (WHO Grade II)	II	Insula	7.04	–	–	No mutation
15	66	F	Diffuse astrocytoma (WHO Grade II)	II	Insula	5.19	–	–	No mutation
16	35	M	Diffuse astrocytoma (WHO Grade II)	II	Pons	9.06	–	–	No mutation
17	41	F	Diffuse astrocytoma (WHO Grade II)	II	Tectum	11.89	–	–	No mutation
18	30	M	Diffuse astrocytoma (WHO Grade II)	II	Temporal	0.00	–	–	Mutation
19	3	M	Diffuse astrocytoma (WHO Grade II)	II	Basal Ganglia	–	5.53	0.46	No mutation
20	74	M	Anaplastic astrocytoma (WHO Grade III)	III	Fronto-temporal	0.59	4.40	0.08	Mutation
21	43	M	Anaplastic astrocytoma (WHO Grade III)	III	Frontal	–	4.34	0.08	Mutation
22	41	F	Anaplastic astrocytoma (WHO Grade III)	III	Parietal	–	4.51	0.10	Mutation
23	59	M	Anaplastic astrocytoma (WHO Grade III)	III	n.a.	7.27	–	–	No mutation
24	62	M	Anaplastic astrocytoma (WHO Grade III)	III	Central	3.72	–	–	Mutation
25	41	M	Anaplastic astrocytoma (WHO Grade III)	III	Frontal	2.40	–	–	Mutation
26	37	F	Anaplastic astrocytoma (WHO Grade III)	III	Frontal	2.29	–	–	Mutation
27	37	F	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	2.68	–	–	No mutation
28	31	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	0.80	–	–	Mutation
29	32	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	2.69	–	–	Mutation
30	42	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	0.13	–	–	Mutation
31	44	F	Anaplastic astrocytoma (WHO Grade III)	III	n.a.	6.85	–	–	No mutation
32	44	F	Anaplastic astrocytoma (WHO Grade III)	III	Frontal	4.86	–	–	No mutation
33	81	F	Anaplastic astrocytoma (WHO Grade III)	III	Cerebellum	1.11	–	–	No mutation
34	70	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	5.35	–	–	No mutation
35	61	M	Anaplastic astrocytoma (WHO Grade III)	III	Thalamus	6.00	–	–	No mutation
36	53	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporo-mesial	–	4.76	0.30	No mutation
37	50	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	–	3.79	0.10	No mutation
38	72	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporoparietal	0.37	4.35	0.16	No mutation
39	47	M	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	–	4.33	0.13	No mutation
40	61	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	2.26	4.06	0.10	No mutation
41	69	F	Glioblastoma multiforme (WHO Grade IV)	IV	Occipital	0.25	3.77	0.09	No mutation

**Table 2.** One hundred seventeen tumor samples were selected from the Brain Tumour Bank, Center for Neuropathology, LMU Munich and the Brain Tumour Bank of the Institute for Neuropathology, University Hospital Homburg (Continued)

Case	Age	Gender	Tumor	Grade	Region	5hmC positive cells (%)	5mC/dG (%)	5hmC/dG (%)	IDH status
42	55	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	0.08	5.58	0.03	No mutation
43	44	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	0.28	4.70	0.09	No mutation
44	64	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	0.97	4.32	0.07	No mutation
45	25	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.37	4.70	0.03	Mutation
46	66	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	3.48	4.27	0.05	No mutation
47	64	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.49	4.06	0.11	No mutation
48	67	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.00	–	–	No mutation
49	67	M	Glioblastoma multiforme (WHO Grade IV)	IV	Occipital	0.55	–	–	No mutation
50	41	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.67	–	–	No mutation
51	66	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.82	–	–	No mutation
52	58	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	1.82	–	–	No mutation
53	30	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.22	–	–	Mutation
54	69	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.45	–	–	No mutation
55	53	f	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.60	–	–	No mutation
56	43	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporo-parieto-occipital	2.21	–	–	No mutation
57	59	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.14	–	–	No mutation
58	7	M	Glioblastoma multiforme (WHO Grade IV)	IV	Thalamus	2.28	–	–	No mutation
59	67	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.78	–	–	No mutation
60	64	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporo-mesial	1.61	–	–	No mutation
61	43	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.19	–	–	Mutation
62	72	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.49	–	–	No mutation
63	32	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.53	–	–	Mutation
64	33	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.20	–	–	Mutation
65	46	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.92	–	–	Mutation
66	40	F	Glioblastoma multiforme (WHO Grade IV)	IV	Parieto-occipital	0.96	–	–	Mutation
67	38	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.60	–	–	Mutation
68	61	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.31	–	–	No mutation
69	47	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.62	–	–	No mutation
70	78	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.38	–	–	No mutation
71	71	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.23	–	–	No mutation
72	62	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.43	–	–	No mutation
73	64	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.89	–	–	No mutation
74	70	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	0.87	–	–	No mutation
75	41	F	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	0.40	–	–	No mutation
76	74	M	Glioblastoma multiforme (WHO Grade IV)	IV	Postcentral	0.53	–	–	No mutation
77	55	M	Glioblastoma multiforme (WHO Grade IV)	IV	Thalamus	0.60	–	–	No mutation
78	37	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	–	4.36	0.08	Mutation
79	34	M	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	–	4.03	0.04	Mutation
80	34	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	–	3.68	0.04	Mutation
81	30	F	Myxopapillary ependymoma (WHO Grade I)	I	Lumbal spine	44.14	5.19	0.09	n.a.
82	68	M	Subependymoma (WHO Grade I)	I	Rhombencephalon	36.50	4.57	0.42	n.a.

**Table 2.** One hundred seventeen tumor samples were selected from the Brain Tumour Bank, Center for Neuropathology, LMU Munich and the Brain Tumour Bank of the Institute for Neuropathology, University Hospital Homburg (Continued)

Case	Age	Gender	Tumor	Grade	Region	5hmC positive cells (%)	5mC/dG (%)	5hmC/dG (%)	IDH status
83	20	M	Subependymoma (WHO Grade I)	I	Lateral ventricle	13.54	–	–	n.a.
84	59	M	Subependymoma (WHO Grade I)	I	Lateral ventricle	28.95	–	–	n.a.
85	54	M	Subependymoma (WHO Grade I)	I	Rhombencephalon	16.49	4.55	0.35	n.a.
86	46	F	Subependymoma (WHO Grade I)	I	Thoracic spine	1.28	–	–	n.a.
87	46	F	Subependymoma (WHO Grade I)	I	Lateral ventricle	16.81	–	–	n.a.
88	46	M	Subependymoma (WHO Grade I)	I	Rhombencephalon	15.22	–	–	n.a.
89	n.a.	n.a.	Subependymoma (WHO Grade I)	I	n.a.	–	4.63	0.20	n.a.
90	n.a.	n.a.	Subependymoma (WHO Grade I)	I	n.a.	–	4.58	0.37	n.a.
91	30	F	Ependymoma (WHO Grade II)	II	Filum	20.09	5.04	0.09	n.a.
92	41	M	Ependymoma (WHO Grade II)	II	Cervical spine	5.15	4.74	0.17	n.a.
93	39	M	Ependymoma (WHO Grade II)	II	Cervical spine	6.26	3.38	0.25	n.a.
94	70	F	Ependymoma (WHO Grade II)	II	Cervical spine	15.85	3.70	0.41	n.a.
95	23	M	Ependymoma (WHO Grade II)	II	Thoracic spine	14.43	6.16	0.21	n.a.
96	54	M	Ependymoma (WHO Grade II)	II	Cervical spine	10.86	5.01	0.12	n.a.
97	42	M	Ependymoma (WHO Grade II)	II	n.a.	12.02	–	–	n.a.
98	76	F	Ependymoma (WHO Grade II)	II	Lumbal spine	15.04	–	–	n.a.
99	21	M	Cellular ependymoma (WHO Grade II)	II	Lateral ventricle	12.49	5.05	0.13	n.a.
100	4	M	Anaplastic ependymoma (WHO Grade III)	III	Rhombencephalon	24.16	4.95	0.23	n.a.
101	5	F	Anaplastic ependymoma (WHO Grade III)	III	Posterior cranial fossa	10.47	4.22	0.03	n.a.
102	59	F	Anaplastic ependymoma (WHO Grade III)	III	Thoracic spine	16.59	4.78	0.32	n.a.
103	6	F	Anaplastic ependymoma (WHO Grade III)	III	Lateral ventricle	1.25	–	–	n.a.
104	72	F	Transitional meningioma (WHO Grade I)	I	Fronto-basal	32.12	4.48	0.12	n.a.
105	22	F	Meningotheliomatous meningiom (WHO Grade I)	I	Frontal	29.06	4.04	0.17	n.a.
106	60	F	Transitional meningioma (WHO Grade I)	I	Cerebellum	9.52	4.08	0.20	n.a.
107	59	M	Fibrous meningioma (WHO Grade I)	I	Occipital	6.82	–	–	n.a.
108	28	M	Meningothelial meningioma (WHO Grade I)	I	Precentral	18.30	–	–	n.a.
109	79	F	Meningothelial meningioma (WHO Grade I)	I	Temporal	12.46	–	–	n.a.
110	61	M	Atypical meningioma (WHO Grade II)	II	Frontal	1.11	4.65	0.08	n.a.
111	33	F	Chordoid meningioma (WHO Grade II)	II	Frontal	5.24	–	–	n.a.
112	72	F	Atypical meningioma (WHO Grade II)	II	Frontal	14.34	–	–	n.a.
113	79	F	Atypical meningioma (WHO Grade II)	II	Frontal	2.93	4.44	0.11	n.a.
114	26	M	Atypical meningioma (WHO Grade II)	II	Skullbase	8.94	–	–	n.a.
115	n.a.	n.a.	Atypical meningioma (WHO Grade II)	II	n.a.	–	4.24	0.27	n.a.
116	n.a.	n.a.	Atypical meningioma (WHO Grade II)	II	n.a.	–	4.29	0.08	n.a.
117	n.a.	n.a.	Atypical meningioma (WHO Grade II)	II	n.a.	–	3.98	0.16	n.a.

and GBM group, the mean age was 47 years, in ependymomas 40 years and in meningiomas 54 years.

#### Immunohistochemistry and cell counting

FFPE tissue samples were used for IHC. Standard protocols were followed, in brief 4- $\mu$ m FFPE sections were deparaffi-

nized by subsequent incubation in xylene (30 min), 100% ethanol (10 min), 96% ethanol (5 min) and 70% ethanol (5 min). The sections were then treated with 2N HCl (35°C, 20 min), pre-incubated with blocking solution (3% I-Block Protein-Based Blocking Reagent (Applied Biosystems, Darmstadt, Germany)) and incubated with the diluted primary

antibody for 18 hr at 4°C. As primary antibody, we used polyclonal rabbit anti-5hmC diluted 1:1,000 (Active Motif, Rixensart, Belgium). Subsequently, the sections were treated with Super Sensitive Polymer-Horseradish peroxidase (HRP) Detection System (BioGenix, San Ramon, CA), followed by incubation in 3,3'-diaminobenzidine (DAB) for 5 min and staining in hemalum. For immunofluorescence staining, we followed standard protocols. As primary antibody, we used polyclonal rabbit anti-5hmC diluted 1:1,000 (Active Motif, Rixensart, Belgium) and monoclonal mouse anti-Ki-67 diluted 1:500 (Dako, Hamburg, Germany). As secondary antibodies, we used Alexa Fluor 488 goat-anti-rabbit IgG diluted 1:300 (Invitrogen, Darmstadt, Germany) and Alexa Fluor 546 goat-anti-mouse IgG diluted 1:300 (Roth, Karlsruhe, Germany). For nuclear staining, we used 4',6-diamidin-2-phenylindol (Roth, Karlsruhe, Germany). Finally, we used Dako fluorescence mounting medium (Dako, Hamburg, Germany).

For cell counting, randomly selected sections were photographed at  $\times 10$  and  $\times 20$  magnification depending on tissue type and size. Following that at least 200 cells were counted in each case.

To identify IDH1 mutations in astrocytoma and GBM, FFPE samples were stained using the monoclonal mouse IDH1<sup>R132H</sup> antibody (Dianova, Hamburg, Germany), diluted 1:20, according to the manufacturer's protocol on a Ventana BenchMark system (Roche, Basel, Switzerland).

#### DNA extraction

For LC-MS, DNA isolation was performed on the basis of the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol with slight modifications. The RNA digestion was executed twice and instead of column purification phenol extraction was performed, *i.e.*, after addition of buffer AL, the sample was mixed and incubated. The sample was distributed equally to two 2 mL reaction tubes, if necessary. A 1/1 mixture of Roti®Phenol/chloroform (1 vol) was added and the tube was shaken vigorously at RT for 5 min. The tube was centrifuged (12,100g, 15 min) and the aqueous layer was collected. This procedure was repeated once. To the obtained aqueous layer chloroform (1 vol) was added and the tube was shaken at RT for 1 min. After centrifugation (12,100g, 5 min) the aqueous layer was collected. During collection of the aqueous layer, special care was taken to include the interphase. The sample was distributed equally to two 2 mL reaction tubes, if necessary. Ethanol (3 vol) was added. The sample was left to stand at RT for approximately 2 h. After precipitation of the DNA the tube was centrifuged (12,100g, 30 min). The supernatant was discarded and the pellet was dried. Subsequently, it was dissolved in water (100–400  $\mu$ L). The solution was centrifuged (12,100g, 30 min) and the supernatant was collected.

Additionally, we isolated DNA from FFPE samples for IDH mutation analysis using a Maxwell 16 system (Promega, Madison, WI) and Maxwell 16 FEV FFPE kit (Promega, Madison, WI) according to the manufacturer's protocols.

#### Enzymatic digestion

For the enzymatic digestion DNA mixtures (4–10  $\mu$ g in a final volume of 100  $\mu$ L H<sub>2</sub>O) were heated to 100°C for 5 min to denature the DNA and rapidly cooled on ice. After addition of standard solutions, buffer A (10  $\mu$ L, 300 mM ammonium acetate, 100 mM CaCl<sub>2</sub>, 1 mM ZnSO<sub>4</sub>, pH 5.7) and nuclease S1 (80 units, *Aspergillus oryzae*) the mixture was incubated for 3 hr at 37°C. Addition of buffer B (12  $\mu$ L, 500 mM Tris-HCl, 1 mM EDTA), antarctic phosphatase (10 units), snake venom phosphodiesterase I (0.2 units, *Crotalus adamanteus* venom) and incubation for further 3 hr at 37°C completed the digestion. Labeled nucleosides d<sub>2</sub>-5hmC and d<sub>3</sub>-5mC were added, followed by centrifugation of the sample (12,100g, 15 min). Each sample was performed at least in duplicate with independent concentrations of the two labeled nucleosides. The concentrations of standard solutions were chosen to be in the expected range of the sample nucleoside concentration.

#### LC-MS

The samples (100  $\mu$ L injection volume) were analyzed by LC-MS on a Thermo Finnigan LTQ Orbitrap XL and were chromatographed by a Dionex Ultimate 3000 HPLC system with a flow of 0.15 mL/min over an Uptisphere120-3HDO column from Interchim. The column temperature was maintained at 30°C. Eluting buffers were buffer C (2 mM HCOONH<sub>4</sub> in H<sub>2</sub>O (pH 5.5)) and buffer D (2 mM HCOONH<sub>4</sub> in H<sub>2</sub>O/MeCN 20/80 (pH 5.5)). The gradient was 0  $\rightarrow$  12 min; 0%  $\rightarrow$  3% buffer D; 12  $\rightarrow$  60 min; 3%  $\rightarrow$  60% buffer D; 60  $\rightarrow$  62 min; 60%  $\rightarrow$  100% buffer D; 62  $\rightarrow$  70 min; 100% buffer D; 70  $\rightarrow$  85 min; 100  $\rightarrow$  0% buffer D; 85  $\rightarrow$  95 min; 0 % buffer D. The elution was monitored at 260 nm (Dionex Ultimate 3000 Diode Array Detector). The chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by the use of a positive polarity mode over a full-scan range of *m/z* 200–1,000 with a resolution of 30,000. Parameters of the mass spectrometer were tuned with a freshly mixed solution of adenosine (5  $\mu$ M) in buffer C. The parameters used in this section were sheath gas flow rate, 16 arb; auxiliary gas flow rate, 11 arb; sweep gas flow rate, 4 arb; spray voltage, 5.0 kV; capillary temperature, 200°C; capillary voltage, 12 V, tube lens 86 V.

#### Pyrosequencing of DNA

To confirm IDH1 mutations detected by IHC staining, we sequenced a 88 bp long fragment of the IDH1 gene including codon 132. The IDH1 primers were designed using PSQ Assay design software (Qiagen, Hilden, Germany). As forward primer, we used 5'-biotin-AAAAATATCCCCCG GCTTG-3' and as reverse primer 5'-TGCCAACATGACTTA CTTGATCC-3'. PCR reactions were performed using the HotStarTaq DNA polymerase master mix (Qiagen, Hilden, Germany) according to the manufacturer's protocol and a

standard PCR protocol. The PCR product was then sequenced using a PyroMark Q24 System (Qiagen, Hilden, Germany) and the Pyro Gold reagent kit (Qiagen, Hilden, Germany). We used 5'-TGATCCCCATAAGCAT-3' as sequencing primer. The subsequent processing was performed according to the manufacturer's protocols. Data analysis was performed using PyroMark Q24 software (Qiagen, Hilden, Germany).

### Statistical analysis

Statistical calculations were performed using ANOVA single factor analysis and Tukey–Kramer *post hoc* test as well as unpaired *t*-test. Analysis was performed using GraphPad Prism 5.

## Results

### Not all cells express 5hmC in normal human brain and brain tumors at detectable levels

IHC with an antibody against 5hmC in general showed either strong staining or no staining with only a small number of faintly stained nuclei (Fig. 1). In normal human frontal brain, the majority of neurons of the cortex showed strong reactivity for 5hmC (Fig. 1a, upper right corner). In contrast, there was weaker immunohistochemical staining for 5hmC in the frontal subcortical white matter (Fig. 1a, lower left corner). About 70% of WM cells showed negative staining; these cells had mostly round, small nuclei and appeared to be oligodendrocytes. In the occipital cortex (Fig. 1b) and subcortical white matter (Fig. 1c), we found comparable results. In human brain tumors, we found significantly lower numbers of 5hmC positive cells. In astrocytomas and GBMs (Figs. 1d–1g), the number of 5hmC positive cells decreased from pilocytic astrocytomas (Fig. 1d) to diffuse astrocytomas (Fig. 1e), AAs (Fig. 1f) and GBMs with mean of 1.43% positive cells (Fig. 1g). Whilst there were about 10% of positive cells in pilocytic astrocytomas, GBMs showed only very few positive cells. In ependymomas (Figs. 1h–1j), we found that most of the cells were not stained. Grade I ependymomas (Fig. 1h) showed more 5hmC positive cells than Grade II (Fig. 1i) and Grade III tumors (Fig. 1j). Grades II and III ependymomas showed more or less equal staining but anaplastic ependymomas showed a weaker staining for 5hmC. Furthermore, there were strong regional variations. In meningiomas (Figs. 1k–1l), we found similar results to ependymomas. Grade I meningiomas (Fig. 1k) showed more 5hmC positive cells than Grade II tumors (Fig. 1l) and Grade II meningiomas showed a weaker staining for 5hmC with regional variations. We found that staining for 5hmC and the proliferation-associated Ki-67 was mutually exclusive; an example of this general finding is shown in Figures 1m–1o. We often observed a rhythmic variation of areas with higher proliferative activity (Ki-67) and sparse 5hmC staining and *vice versa*.

### There are significantly more 5hmC positive stained cells in the cortex compared to white matter areas. 5hmC/dG in human brain is significantly higher in cortex than white matter

When we counted the 5hmC positive and negative cells the values for various cortical areas did not show significant differences (Fig. 2a,  $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). Furthermore, we did not find any significant difference in the ration of 5hmC positive cells in various white matter areas (Fig. 2a,  $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). However, we found significant differences comparing frontal cortex and frontal white matter areas (Fig. 2a,  $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test) as well as occipital cortex and occipital white matter areas (Fig. 2a,  $p < 0.05$ , ANOVA and Tukey–Kramer *post hoc* test).

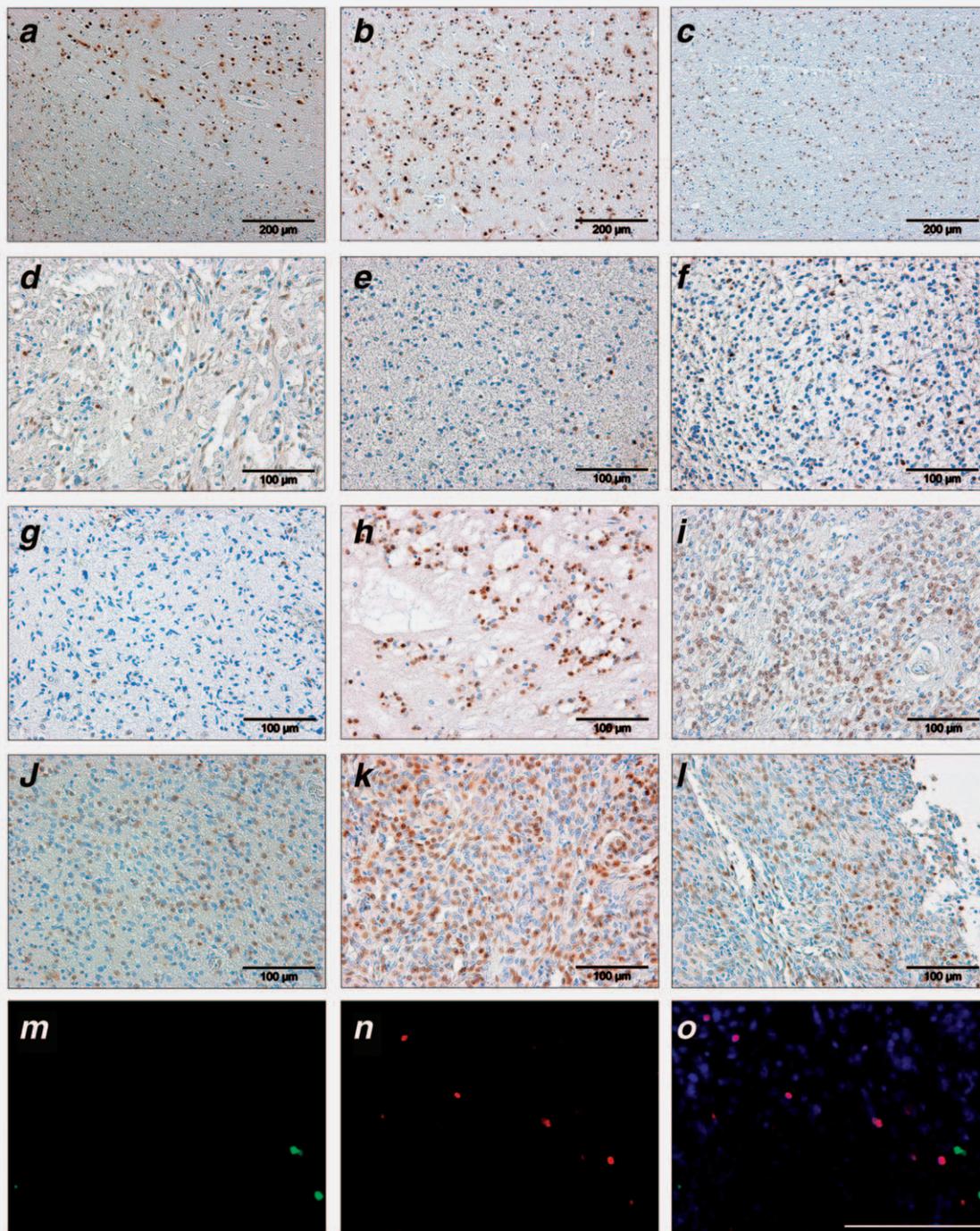
Additionally, we performed LC-MS measurements. 5hmC was found in measurable amounts in all brain regions but there was no significant difference between various cortical (Fig. 2b,  $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) and various white matter areas (Fig. 2b,  $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). But the values measured in the frontal cortex were significantly higher than in the frontal white matter (Fig. 2b,  $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test) and the values measured in the occipital cortex were significantly higher than in the occipital white matter (Fig. 2b,  $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test).

The 5mC values measured by LC-MS did not show significant differences between different cortical areas ( $p > 0.05$  using ANOVA and Tukey–Kramer *post hoc* test, Fig. 2c) nor between different white matter areas ( $p > 0.05$  using ANOVA and Tukey–Kramer *post hoc* test, Fig. 2c) but we found significant differences between cortical and white matter areas ( $p < 0.05$ , ANOVA and Tukey–Kramer *post hoc* test, Fig. 2c).

### The number of 5hmC positive cells and 5hmC/dG levels in human brain tumors are related to WHO grade

One hundred seventeen human tumors were selected for analysis including 6 pilocytic astrocytomas (Grade I), 13 diffuse astrocytomas (Grade II), 18 AAs (Grade III) and 43 GBMs (Grade IV), 10 ependymomas Grade I, 9 ependymomas (Grade II) and 4 anaplastic ependymomas (Grade III), 6 WHO Grade I meningiomas and 8 atypical meningiomas (Grade II).

In WHO Grade I tumors, the mean of 5hmC positive cells was 16.73%, in Grade II 7.87%, in Grade III 5.51% and in Grade IV 1.07% (Fig. 2d) with a highly significant difference between Grades I and II ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test), Grades I and III ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test), Grades I and IV ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test) as well as Grades II and IV ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test).



**Figure 1.** IHC with an antibody against 5hmC. (a) In the normal frontal lobe we see differences in staining of the cortex and subcortical white matter. In the cortex (upper right corner) there are numerous clearly visible positively stained neurons. In the subcortical white matter (lower left corner) there is diminished staining for 5hmC, but distinct positively stained cells are seen surrounded by negative ones. (b) In the normal occipital cortex, there is a clear staining of the majority (about 60%) of cells. (c) In normal occipital white matter, there are fewer 5hmC positive cells (about 30%) compared to the occipital cortex. (d) In human pilocytic astrocytomas (WHO Grade I) the majority of tumor cells are negative, only about 10% of tumor cells stain positive. (e) In diffuse astrocytomas less than 5% of cells stain positive for 5hmC. (f) AAs show lower values of 5hmC positive cells compared to low-grade gliomas. (g) GBMs show only very low positive staining for 5hmC. (h) In Grade I ependymomas most cells stain negative for 5hmC. (i) Ependymomas Grade II show lower numbers of 5hmC positive cells than Grade I tumors with a large intratumoral variation. (j) Anaplastic ependymomas show almost the same number of 5hmC positive cells compared to Grade II ependymomas but the cells stain much weaker. Furthermore there is a large intratumoral variation. (k) In Grade I meningiomas we see only a minority of positively stained cells with a large intratumoral variation. (l) Meningiomas Grade II show fewer 5hmC positive cells than Grade I tumors and the cells seem to stain more weakly than in Grade I tumors with large intratumoral variations. (m–o) Immunofluorescence staining of a GBM with antibodies against 5hmC (m), the proliferation marker Ki-67 (n) and DAPI. Scale bar is 100 µm. Cells stained positive for Ki-67 show no detectable signal for 5hmC in the merge picture (o).

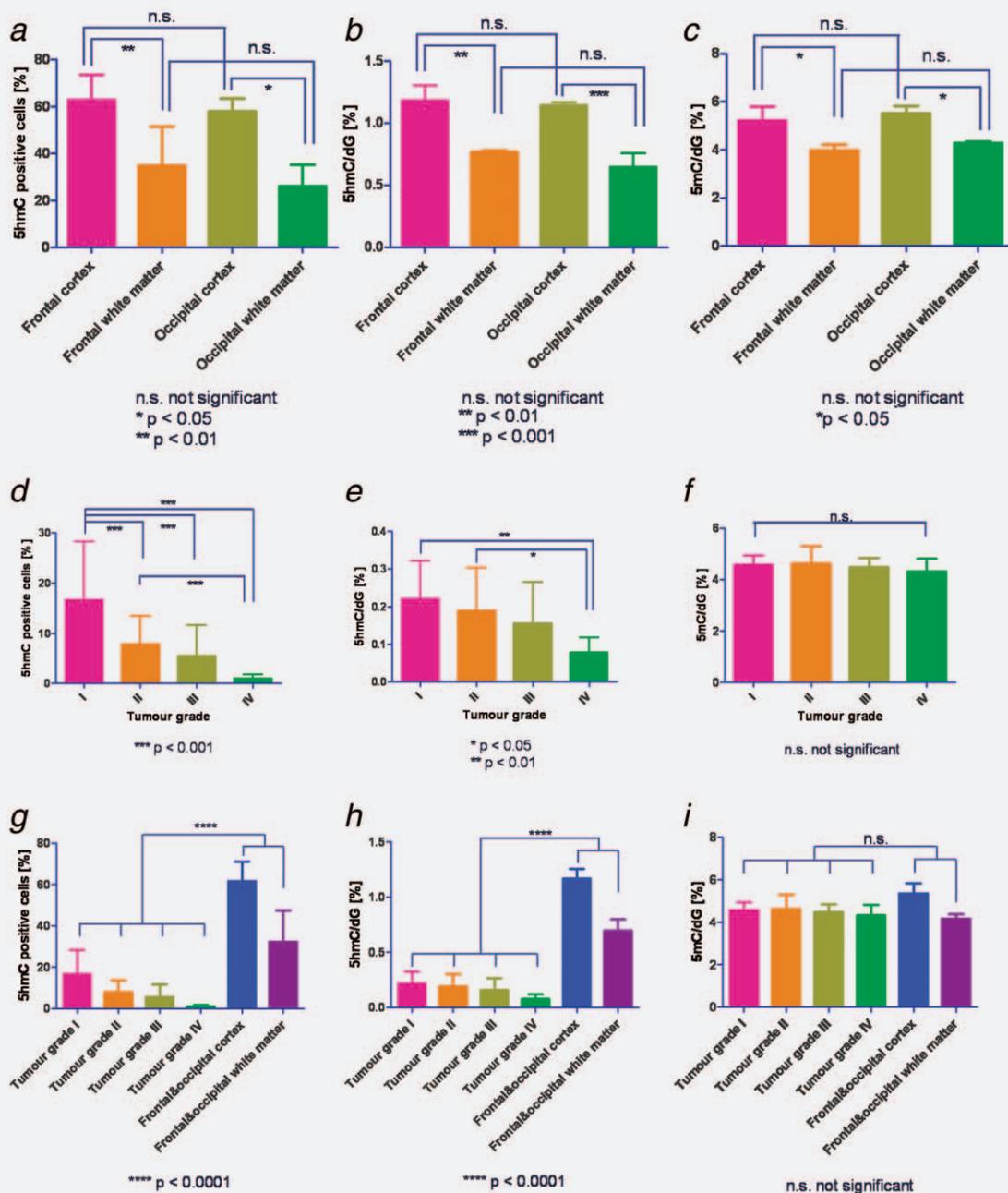


Figure 2. Percentage of 5hmC positive cells and LC-MS values of 5hmC and 5mC in normal human brain and in brain tumors. Indicated are the mean and SD. (a) There was no significant differences of 5hmC positive cells in various white matter areas as well as in various cortex regions of the human brain ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) but there were highly significant differences in the cortex compared to the white matter ( $p < 0.01$  in case of frontal cortex and frontal white matter and  $p < 0.05$  in case of occipital cortex and occipital white matter, ANOVA and Tukey–Kramer *post hoc* test). (b) LC-MS values showed no significant difference between various white matter regions and between various cortex regions ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) but highly significant differences between cortex and white matter regions ( $p < 0.01$  in case of frontal cortex and frontal white matter and  $p < 0.001$  in case of occipital cortex and occipital white matter, ANOVA and Tukey–Kramer *post hoc* test). (c) The distribution of 5mC in different brain regions showed no significant difference between different cortical regions as well as in different white matter regions ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) but there were significant difference between cortical and white matter regions ( $p < 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). (d) A comparison of 5hmC positive cells in human brain tumors showed highly significant differences between Grade I and Grade II tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test), Grade I and Grade III tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test), Grade I and Grade IV tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test) as well as Grade II and Grade IV tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test). (e) In LC-MS we found significant lower amounts of 5hmC in Grade I compared to Grade IV tumors ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test). (f) The 5mC values did not show a correlation with tumor grade ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). (g) A comparison of tumors with normal brain showed highly significantly lower numbers of 5hmC positive cells in tumors compared to normal brain ( $p < 0.0001$ , unpaired *t*-test). (h) In LC-MS we found that tumors show significant higher amount of 5hmC compared to normal brain ( $p < 0.0001$ , unpaired *t*-test). (i) There was no significant difference between tumor and normal brain in regard to 5mC ( $p > 0.05$ , unpaired *t*-test). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

By mass spectrometry we see that in Grade I tumors the 5hmC values were 0.22% 5hmC/dG, in Grade II 0.19% 5hmC/dG, in Grade III 0.16% 5hmC/dG and in Grade IV 0.078% 5hmC/dG with a significant difference between Grades I and IV tumors ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test) as well as Grades II and IV tumors ( $p < 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 2e).

No such relation was found in LC-MS measurements of 5mC in WHO tumor grades (Fig. 2f). Grade I tumors showed mean values of 4.58% 5mC/dG, Grade II 4.62% 5mC/dG, Grade III 4.47% 5mC/dG and Grade IV 4.32% 5mC/dG with no significant difference between different tumor grades ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test).

The number of 5hmC positive cells identified in brain tumors was significantly lower compared to normal cortex and white matter (Fig. 2g). While in normal human brain mean number of 5hmC positive cells was 61.48% in the cortex and 32.39% in the white matter, the mean in tumor tissue was 6.58% ( $p < 0.0001$ , unpaired *t*-test).

These data were paralleled by LC-MS (Fig. 2h). In normal cortex, we measured 1.17% 5hmC/dG, in the white matter 0.70% 5hmC/dG and in tumor tissue the average value was 0.16% 5hmC/dG ( $p < 0.0001$ , unpaired *t*-test).

As regards 5mC no significant difference was observed between normal brain tissue and tumor (Fig. 2i) with a mean of 4.85% 5mC/dG in the normal human brain and 4.50% 5mC/dG in tumors ( $p > 0.05$ , unpaired *t*-test).

When tumors were grouped according to their cellular lineages, we obtained similar differences of 5hmC related to the tumor grade.

In astrocytomas, the mean of 5hmC positive stained cells was 8.91% in Grade I pilocytic astrocytomas, 4.40% in Grade II diffuse astrocytomas, 3.34% in Grade III AAs and 1.07% in Grade IV GBMs (Fig. 3a) with highly significant differences between Grade I and Grade III tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test), Grade I and Grade IV tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test) and Grade II and Grade IV tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test) as well as significant differences between Grade I and Grade II tumors ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test) and Grade III and Grade IV tumors ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3a).

Using LC-MS, we found mean values of 0.19% 5hmC/dG in Grade I astrocytomas, 0.24% 5hmC/dG in Grade II, 0.13% 5hmC/dG in Grade III and 0.078% 5hmC/dG in Grade IV tumors (Fig. 3b) with significant differences between Grade II and Grade IV tumors ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3b).

In regard to 5mC, there were no significant differences related to WHO grades of astrocytic gliomas using LC-MS (Fig. 3c). The mean value of Grade I astrocytomas was 4.68% 5mC/dG, of Grade II 4.82% 5mC/dG, of Grade III 4.36% 5mC/dG and of Grade IV 4.32% 5mC/dG ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3c).

In ependymal tumors, we found a tendency of 5hmC positive cells between Grade I as well as Grade II and Grade III tumors by IHC staining but we did not find significant differences ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3d). Grade I ependymomas showed mean values of 21.62% positive stained cells, Grade II and Grade III tumors show almost equal values of 5hmC positively stained cells (12.47% and 13.12%).

LC-MS measurements showed that WHO Grade I ependymomas have mean values of 0.286% 5hmC/dG, Grade II ependymomas 0.197% 5hmC/dG and anaplastic ependymomas (WHO Grade III) 0.193% 5hmC/dG and thus paralleled the same tendency as the immunohistochemical stainings ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3e).

With regard to 5mC no significant differences were found using LC-MS; in Grade I ependymomas the mean was 4.70% 5mC/dG, in Grade II 4.73% 5mC/dG and in Grade III 4.65% 5mC/dG ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3f).

A comparison of WHO Grades I and II meningiomas in IHC showed lower numbers of 5hmC positive cells in Grade II meningiomas compared to Grade I meningiomas (mean of 18.05% positive cells in Grade I and 6.51% positive cells in Grade II) but the difference did not reach statistical significance ( $p > 0.05$ , unpaired *t*-test) (Fig. 3g).

By mass spectrometry, we found a mean value of 0.163% 5hmC/dG in Grade I and 0.140% 5hmC/dG in one Grade II atypical meningioma, which again paralleled the same tendency as in IHC but this did not reach statistical significance ( $p > 0.05$ , unpaired *t*-test) (Fig. 3h).

Again, 5mC/dG levels did not show significant differences between Grade I and Grade II meningiomas (4.20% 5mC/dG in Grade I meningiomas and 4.32% 5mC/dG in Grade II meningiomas,  $p > 0.05$ , unpaired *t*-test) (Fig. 3i).

#### The number of 5hmC positive cells is significantly related to IDH1 mutations in diffuse and anaplastic astrocytomas but not in glioblastomas

Additionally, we analyzed the IDH1 status of all astroglial tumors by IHC using an IDH1<sup>R132H</sup> antibody and pyrosequencing of the IDH1 gene. As expected we did not find IDH1 mutations in pilocytic astrocytomas (0% mutated IDH1), but we found 7 IDH1 mutations in diffuse astrocytomas (54% mutated IDH1), 9 IDH1 mutations in AAs (50% mutated IDH1) and 11 IDH1 mutations in GBMs (26% mutated IDH1).

Comparing 5hmC positive cells of diffuse and AAs as well as in GBMs (Fig. 4a) we found highly significant differences between diffuse astrocytomas with and without IDH1 mutation (1.50% 5hmC positive cells in tumors with IDH1 mutation compared to 7.31% 5hmC positive cells in tumors without IDH1 mutation,  $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test) and significant differences in AAs with and without IDH1 mutation (1.80% 5hmC positive cells in tumors with IDH1 mutation and 4.87% 5hmC positive cells in tumors without IDH1 mutation,  $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test). In GBM, we did not find significant differences

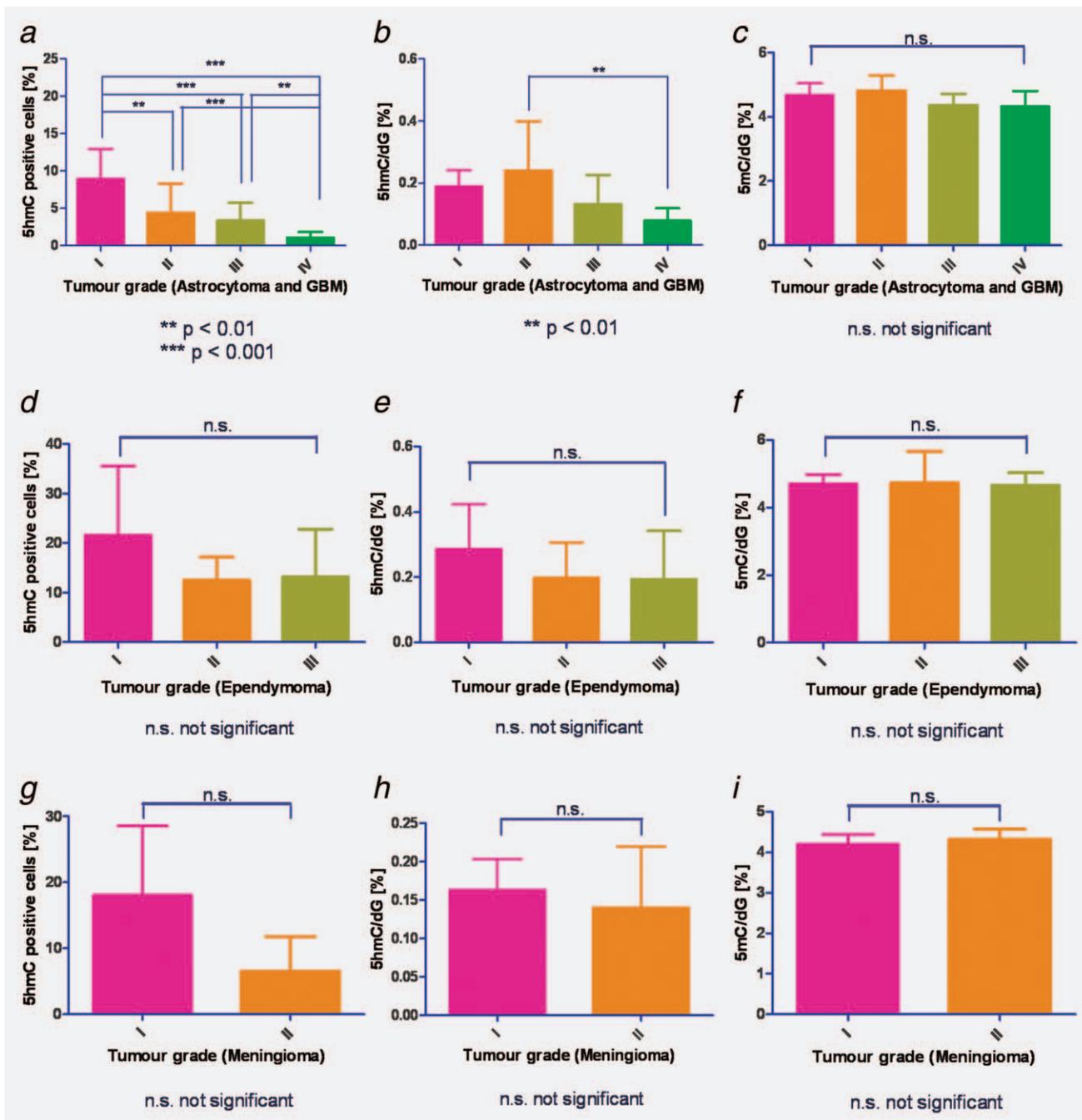
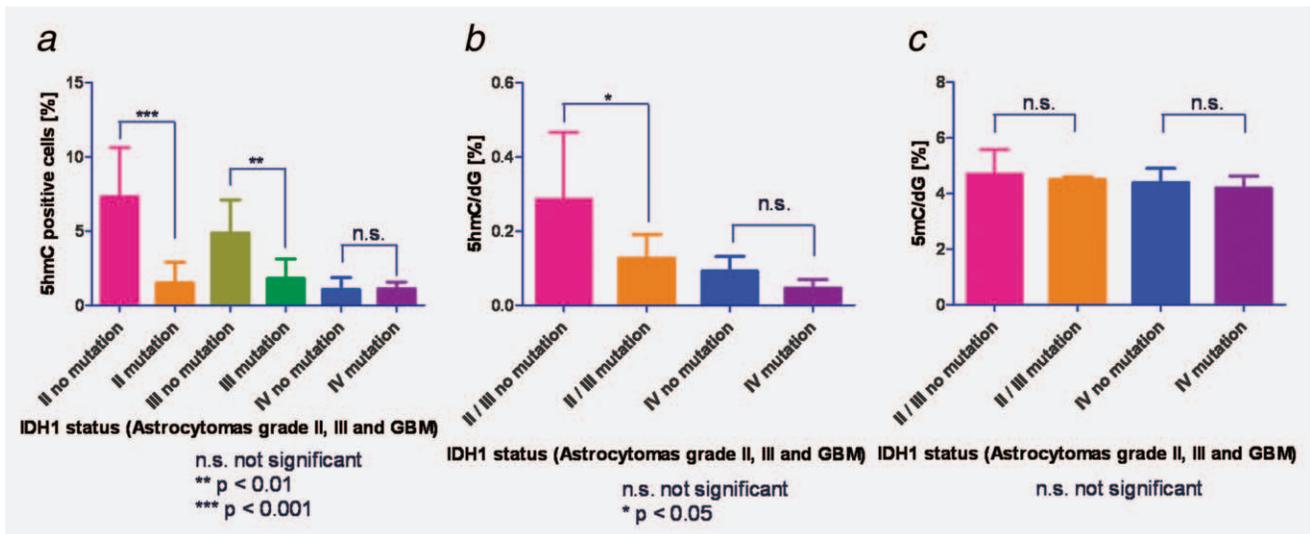


Figure 3. Percentage of 5hmC positive cells and LC-MS values of 5hmC and 5mC in various tumor entities. Indicated are the mean and SD. (a) In astrocytomas and GBMs a good correlation between the number of 5hmC positive cells and tumor grade was seen. Pilocytic Grade I astrocytomas showed highly significantly higher numbers of 5hmC positive cells compared to Grade III tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test) and Grade IV tumors ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test), and Grade II tumors showed highly significant higher numbers of 5hmC positive cells compared to Grade IV GBM ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test). Furthermore Grade I astrocytomas showed significantly lower numbers of 5hmC positive cells compared to Grade II tumors ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test) as well as Grade III tumors showed significantly lower numbers of 5hmC positive cells compared to Grade IV GBM ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test). (b) LC-MS showed similar results. Diffuse Grade II astrocytomas showed significantly higher amounts of 5hmC compared to undifferentiated Grade IV GBMs ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test). (c) 5mC showed no significant differences between different tumor grades ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). (d) Using IHC, in ependymomas Grade I there were more 5hmC positive cells than in ependymomas Grades II and III, but there was no significant decrease ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). (e) Using LC-MS we found higher amounts of 5hmC in Grade I ependymomas compared to Grades II and III tumors without statistical significance ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). (f) 5mC showed no significant differences in ependymomas of different WHO grades ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). (g) Meningiomas showed a decrease of 5hmC positive cells in Grade II tumors compared to Grade I tumors but the difference did not reach statistical significance ( $p > 0.05$ , unpaired *t*-test). (h) LC-MS showed higher amounts of 5hmC in Grade I tumors compared to Grade II tumors without statistical significance ( $p > 0.05$ , unpaired *t*-test). (i) The amount of 5mC did not show any significant differences in regard to tumor grade ( $p > 0.05$ , unpaired *t*-test) [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)].



**Figure 4.** Number of 5hmC positive cells and LC-MS values of 5hmC/dG and 5mC/dG in astrocytomas and GBMs with regard to IDH1 status. Indicated are the mean and SD. (a) Comparing the number of 5hmC positive cells in diffuse and AAs as well as in GBMs with and without IDH1 mutations we found that diffuse astrocytomas without IDH1 mutation had significantly more 5hmC positive cells ( $p < 0.001$ , ANOVA and Tukey–Kramer *post hoc* test). Similar and highly significant differences were noted between AAs with and without IDH1 mutation ( $p < 0.01$ , ANOVA and Tukey–Kramer *post hoc* test). GBMs with and without IDH1 mutation did not show significant differences in the number of 5hmC positive cells ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). Remarkably, Grade II diffuse and Grade III AAs that had a IDH1 mutation and GBMs with and without mutation had very low 5hmC counts, which were not significantly different between these groups ( $p > 0.05$ ). (b) Using MS-LC we could parallel these results with significant differences between Grades II and III astrocytomas with and without mutant IDH1 ( $p < 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). There was no significant difference in GBMs with and without IDH1 mutation ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). (c) The amount of 5mC did not show any significant differences in regard to IDH1 mutations ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test). [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

between tumors with or without IDH1 mutation ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 4a).

By mass spectrometry, we found a mean value of 0.287% 5hmC/dG in Grade II and Grade III tumors without IDH1 mutation and 0.127% 5hmC/dG in tumors with IDH1 mutation with a significant difference ( $p < 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 4b). In GBM, we did not find significant differences between tumors with or without IDH1 mutation ( $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 4b).

We did not find significant differences in the amount of 5mC/dG between Grades II and III astrocytomas with and without IDH1 mutation as well as GBMs (4.69% 5mC/dG in Grades II and III tumors without IDH1 mutation and 4.50% 5mC/dG in tumors with IDH1 mutation,  $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test, and 4.38% 5mC/dG in GBMs without IDH1 mutation and 4.19% 5mC/dG with IDH1 mutation,  $p > 0.05$ , ANOVA and Tukey–Kramer *post hoc* test) (Fig. 4c).

## Discussion

The profoundly distorted epigenetic landscape of tumor cells is only beginning to emerge.<sup>12–14</sup> DNA methylation, a major epigenetic mechanism of gene silencing, seems to play a major role in tumor pathogenesis. Transcriptional inactivation caused by promoter hypermethylation affects genes

involved in the main cellular pathways, DNA repair, vitamin response, Ras signaling, cell-cycle control, p53 network and apoptosis.<sup>15</sup> Hypermethylation patterns are tumor-type specific and it is still unclear why certain regions become hypermethylated, whereas others remain unmethylated. Hypomethylation at specific promoters can activate the aberrant expression of oncogenes and induce loss of imprinting.<sup>14</sup>

DNA methylation is mediated by the DNMT family of enzymes that catalyze the transfer of a methyl group from S-adenosyl methionine to DNA. However, it seems that the DNA methylome is reprogrammed during development and we hypothesized that this might also be the case in gliomas. The discovery that 5mC can be oxidized to 5hmC by the action of the TET enzymes has engendered ideas of a role of TET and 5hmC in genetic reprogramming, *i.e.*, that 5hmC could be an intermediate leading to demethylation of 5mC and consequently a reversal of gene silencing by methylation. Indeed, very recent research has shown that TET proteins can convert 5mC to 5-formylcytosine and 5-carboxylcytosine and that the latter can be excised by thymine-DNA glycosylase (TDG), which has been interpreted as a pathway for active DNA demethylation.<sup>5–7</sup>

The results presented here show that hydroxymethylation of genomic DNA cytosine residues is a feature found in the human central nervous system (CNS). The highest values were found in cortical areas (mean 1.17% 5hmC/dG) while

the white matter showed much lower values of 0.696% 5hmC/dG (ratio 1.68:1). Since neurons in the cortex are outnumbered by glia cells by a factor of 2–10 and IHC did not show a difference in the percentage of glial nuclei in the cortex and WM, we have to assume that the values of 5hmC/dG in neurons must be even higher. Percentage values of 5hmC/dG in the human adult cortex are the highest values published to date. The values of 5mC were around 5.3% of dG in the cerebral cortex and 4.2% in the cerebral white matter (ratio 1.26:1). Thus, overall it seems that a higher percentage of DNA is hydroxymethylated in neurons in absolute terms but also in relation to 5mC.

By comparison in the adult mouse, 5hmC/dG levels were highest in the CNS<sup>16</sup>; the adult murine cerebral cortex and hippocampus showed values around 0.6% 5hmC/dG while in other organs such as the kidney, heart and lung values were around 0.15%–0.2% and even lower in the liver, spleen and testis. Remarkably, 5mC/dG percentages were around 4.5% in most organs<sup>16</sup> and various areas of the brain.<sup>17</sup> Using an immunoprecipitation technique Jin *et al.* reported that 5hmC in DNA from human frontal lobe was more selectively targeted to genes than was 5mC and was particularly enriched at promoters and in intragenic regions. The presence of 5hmC was more positively correlated with gene expression than was the presence of 5mC.<sup>18</sup>

Surprisingly, our immunohistochemical results demonstrate that 5hmC is quite unevenly distributed in cells of neuronal and glial lineage. We would not conclude from our data that DNA in some cells is free of 5hmC, but our results unequivocally show that there is a wide variation in cellular 5hmC concentration.

In our investigation, 5hmC values in gliomas range from 0.086% to 0.24% 5hmC/dG and thus are surprisingly low when compared to measurements in the white matter. 5mC/dG values in tumors were around 4.5% and thus were slightly higher than in normal white matter. Since tumor infiltration areas could be a possible confounding factor that is difficult to exclude in investigating gliomas, we took great care to sample pure tumor areas excluding infiltration zones of the white matter or cortex as far as that is possible on light-microscopic inspection. However, tumor infiltration may still not be a totally negligible factor. When all tumor entities including astrocytomas, ependymomas and meningiomas were grouped by WHO grade, significantly lower 5hmC/dG values were noted in high-grade anaplastic tumors than in low-grade tumors while no significant differences were found in 5mC/dG values. When considering 5hmC/dG values of the individual tumor lineages, significant differences were found for low-grade astrocytomas *versus* GBMs, whereas only a tendency could be shown regarding ependymomas and meningiomas. Again, as in the normal brain, IHC demonstrated uneven cellular distribution in various tumor cell populations. We found that staining for 5hmC and the proliferation-associated Ki-67 was mutually exclusive and often observed a rhythmic variation of areas with higher proliferative activity (Ki-67) and sparse 5hmC staining and *vice*

*versa*. Again, 5mC/dG measurements showed no significant differences between WHO low- and high-grade tumors.

Thus by using LC-MS we have shown that 5hmC values related to the number of dG in the human brain are highest in gray matter areas, considerably lower in the white matter and even lower in brain tumors. Comparable values of 5hmC in the prefrontal cortex and a similar reduction in astrocytomas have recently been published by Jin *et al.*<sup>19</sup>; data on the white matter, GBMs and other brain tumors are not available at present. In the tumor group, lower 5hmC values were associated with anaplasia (WHO Grade IV), *i.e.*, a low degree of cellular differentiation and high mitotic activity. IHC with an antibody against 5hmC, more precisely the ratio of positively stained nuclei related to unstained nuclei, showed the same tendency as LC-MS; higher values were measured in the gray matter *vs.* white matter and highly differentiated tumors *vs.* anaplastic tumors. Additional factors such as the different cells of origin of these tumors may influence the 5hmC values. Pilocytic astrocytomas may be a particular case in point since they are known to be genetically different from higher-grade astrocytomas.

Acute myeloid leukemia (AML)-associated mutations of IDH1 and IDH2 that display a neomorphic enzyme activity resulting in the production of 2-hydroxyglutarate (2HG) have been shown to impair the catalytic function of TET2, an  $\alpha$ -ketoglutarate (2-oxoglutarate or 2OG)-dependent enzyme.<sup>20</sup> Xu *et al.* reported decreased 5hmC values in a mixture of Grade III astrocytomas and GBMs harboring mutant IDH1.<sup>21</sup> Similarly our data show in greater depth a significant difference in the percentage of 5hmC positive cells in diffuse and AAs with and without IDH1 mutation but not in GBMs (Fig. 4a). LC-MS measurements showed the same relationship between diffuse and AAs with and without IDH1 mutation and in GBMs.

Jin *et al.* found 5hmC levels were independent of IDH1 mutations in their sample of gliomas<sup>19</sup>; their data showed the same tendency as ours but did not reach statistical significance.

In a cohort of patients with AML, IDH1/2 mutations and TET2 loss-of-function mutations were mutually exclusive and elicited similar epigenetic defects.<sup>20</sup> Kim *et al.* found no TET2 mutation in 29 low-grade diffuse gliomas,<sup>22</sup> neither have we identified TET2 loss-of function mutations in seven investigated cases of astrocytomas and GBMs (data not shown). Thus, concluding from what we know from myeloid leukemia, in human gliomas the observed very low 5hmC values may be related to functional impairment of the TET enzymes by 2HG. Remarkably, Grade II diffuse and Grade III AAs and GBMs that had IDH1 mutations had almost identical low 5hmC values. GBMs with no IDH1 mutation, which by and large are primary GBMs, had equally low values; therefore other, unknown factors may impair 5mC oxidation in these tumors.

### Acknowledgements

The authors thank the Brain Bank Munich (Sigrun Roeber) for providing control tissues as well as Virginie Guibourt and Michael Schmidt for expert technical assistance. M. Mün. and T.P. thank the Fonds der Chemischen Industrie for pre-doctoral fellowships.

## References

- Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JE, Hau P, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med* 2005;352:997–1003.
- Wick W, Hartmann C, Engel C, Stoffels M, Felsberg J, Stockhammer F, Sabel MC, Koepfen S, Ketter R, Meyermann R, Rapp M, Meisner C, et al. NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with procarbazine, lomustine, and vincristine or temozolomide. *J Clin Oncol* 2009;27:5874–80.
- Kaina B, Christmann M, Naumann S, Roos WP. MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst)* 2007;6:1079–99.
- Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 2009;324:929–30.
- He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 2011;333:1303–7.
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 2011;333:1300–3.
- Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeyer C, Carell T. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* 2011;50:7008–12.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009;324:930–5.
- Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 2011;8:200–13.
- Haffner MC, Chau A, Meeker AK, Esopi DM, Gerber J, Pellakuru LG, Toubaji A, Argani P, Iacobuzio-Donahue C, Nelson WG, Netto GJ, De Marzo AM, et al. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* 2011;2:627–37.
- Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 2007;114:97–109.
- Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 2007;8:286–98.
- Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis* 2009;31:27–36.
- Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol* 2010;28:1057–68.
- Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* 2007;16 Spec No 1:R50–9.
- Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 2010;5:e15367.
- Munzel M, Globisch D, Bruckl T, Wagner M, Welzmler V, Michalakis S, Muller M, Biel M, Carell T. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl* 2010;49:5375–7.
- Jin SG, Wu X, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res* 2011;39:5015–24.
- Jin SG, Jiang Y, Qiu R, Rauch TA, Wang Y, Schackert G, Krex D, Lu Q, Pfeifer GP. 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* 2011;71:7360–5.
- Figuerola ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, Tallman MS, Sun Z, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 2010;18:553–67.
- Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, Ito S, Yang C, Xiao MT, Liu LX, Jiang WQ, Liu J, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* 2011;19:17–30.
- Kim YH, Pierscianek D, Mittelbronn M, Vital A, Mariani L, Hasselblatt M, Ohgaki H. TET2 promoter methylation in low-grade diffuse gliomas lacking IDH1/2 mutations. *J Clin Pathol* 2011;64:850–2.

## **XI. Tet enzymes oxidize T to hmU in mouse embryonic stem cells**

Toni Pfaffeneder<sup>1#</sup>, Fabio Spada<sup>1#</sup>, Mirko Wagner<sup>1#</sup>, Caterina Brandmayr<sup>1</sup>, Silvia Laube<sup>1</sup>, David Eisen<sup>1</sup>, Matthias Truss<sup>2</sup>, Jessica Steinbacher<sup>1</sup>, **Benjamin Hackner**<sup>1</sup>, Olga Kotljarova<sup>1</sup>, David Schuermann<sup>5</sup>, Stylianos Michalakis<sup>3</sup>, Olesya Kosmatchev<sup>1</sup>, Stefan Schiesser<sup>1</sup>, Barbara Steigenberger<sup>1</sup>, Nada Raddaoui<sup>1</sup>, Udo Müller<sup>4</sup>, Heinrich Leonhardt<sup>4</sup>, Primo Schär<sup>5</sup>, Markus Müller<sup>1\*</sup> and Thomas Carell<sup>1\*</sup> ; *Nat. Chem. Bio.* **2014**, *accepted*

### **Introduction:**

TET enzymes are involved in the oxidation of mC as described in chapter I.2.3.3.1. In this manuscript it is shown, that not only mC is oxidized but also dT. The product hmU was earlier only associated with deamination in mES cells as explained in I.2.3.3.3.

### **Declaration of contribution:**

I purified several standards for the LC-MS/MS level analysis.

# Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA

Toni Pfaffeneder<sup>1,8</sup>, Fabio Spada<sup>1,8</sup>, Mirko Wagner<sup>1,8</sup>, Caterina Brandmayr<sup>1</sup>, Silvia Laube<sup>1</sup>, David Eisen<sup>1</sup>, Matthias Truss<sup>2</sup>, Jessica Steinbacher<sup>1</sup>, Benjamin Hackner<sup>1</sup>, Olga Kotljarova<sup>1</sup>, David Schuermann<sup>5</sup>, Stylianos Michalakis<sup>4</sup>, Olesea Kosmatchev<sup>1</sup>, Stefan Schiesser<sup>1</sup>, Barbara Steigenberger<sup>1</sup>, Nada Raddaoui<sup>1</sup>, Gengo Kashiwazaki<sup>1</sup>, Udo Müller<sup>5</sup>, Cornelia G Spruijt<sup>6</sup>, Michiel Vermeulen<sup>6,7</sup>, Heinrich Leonhardt<sup>4</sup>, Primo Schär<sup>5</sup>, Markus Müller<sup>1\*</sup> & Thomas Carell<sup>1\*</sup>

**Ten eleven translocation (Tet) enzymes oxidize the epigenetically important DNA base 5-methylcytosine (mC) stepwise to 5-hydroxymethylcytosine (hmC), 5-formylcytosine and 5-carboxycytosine. It is currently unknown whether Tet-induced oxidation is limited to C-derived nucleobases or whether other nucleobases are oxidized as well. We synthesized isotopologs of all major oxidized pyrimidine and purine bases and performed quantitative MS to show that Tet-induced oxidation is not limited to mC but that T is also a substrate that gives 5-hydroxymethyluracil (hmU) in mouse embryonic stem cells (mESCs). Using MS-based isotope tracing, we show that deamination of hmC does not contribute to the steady-state levels of hmU in mESCs. Protein pull-down experiments in combination with peptide tracing identifies hmU as a base that influences binding of chromatin remodeling proteins and transcription factors, suggesting that hmU has a specific function in stem cells besides triggering DNA repair.**

mC is an epigenetically important nucleobase associated with the control of transcriptional activity, genomic imprinting, X-chromosome inactivation and suppression of transposable elements<sup>1</sup>. Controlled formation and removal of mC at specific genomic loci is critical for correct genome programming or reprogramming during cellular differentiation<sup>2</sup>. Recently, it was discovered that Tet proteins (Tet1–3) oxidize mC to give the oxidized C-derived nucleobases hmC<sup>3</sup>, 5-formylcytosine (fC)<sup>4,5</sup> and 5-carboxycytosine (caC)<sup>5,6</sup>, whose biological functions are still yet unclear (Fig. 1a)<sup>7</sup>. As fC and caC are both removed by thymine DNA glycosylase (Tdg)<sup>6,8</sup>, it is currently assumed that they serve as intermediates of an active DNA demethylation process involving base excision repair. In addition to these oxidized C derivatives, cells also contain oxidized T nucleobases such as hmU and fU. These compounds are currently known as oxidative lesions that are thought to form upon the reaction of T with reactive oxygen species (ROS)<sup>9,10</sup>. It was recently suggested that hmU might also be produced by deamination of hmC, a hypothesis that remains controversial<sup>11–14</sup>. Deamination of hmC, situated in a base pair with G (hmC:G), would give rise to hmU:G mismatches, which are known substrates for the DNA glycosylases Tdg, Smug1, Mbd4, Ung2 (ref. 15), Neil1 and Nth1 (ref. 16). Deamination of hmC:G to hmU:G followed by mismatch repair would therefore establish an alternative pathway to active demethylation (Fig. 1a).

To unravel the origin of oxidized nucleobases, and of hmU in particular, in DNA from mESCs, we performed isotope tracing and quantitative MS studies using the chemically synthesized

isotopologs of mC, hmC, fC, caC, hmU and fU as internal standards (Fig. 1b and Supplementary Results, Supplementary Fig. 1). For the assessment of oxidation products that are formed by the action of ROS, we additionally quantified 8-oxo-G because 8-oxo-G is a well-established ROS reaction product formed from G<sup>17,18</sup>. We show here that hmU is generated enzymatically from thymidine during stem cell differentiation by the action of the Tet enzymes. A proteomic analysis provides new insight into how genomic hmU can influence the binding of chromatin remodeling proteins and transcription factors.

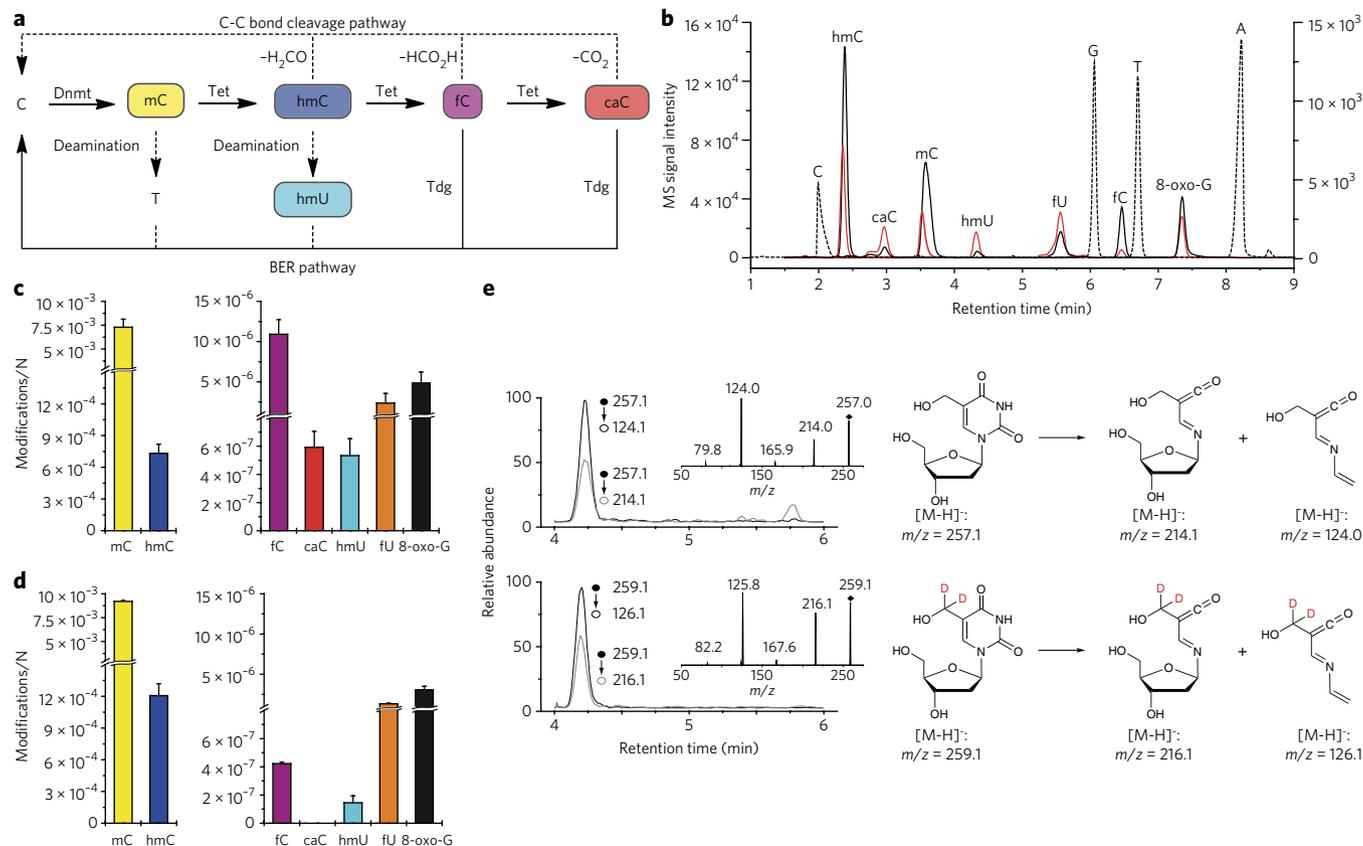
## RESULTS

### hmU is present at elevated levels in mESCs

We first created an inventory of the named nucleosides (Supplementary Fig. 1) in mESCs (Fig. 1c) and adult cortex tissue (Fig. 1d). In mESCs, we observed that hmC is, as expected, the most abundant oxidized pyrimidine (10% relative to mC), followed by fC (~1–2% of hmC). The ROS marker 8-oxo-G was detected at similar levels (~45% of fC), showing that nonenzymatic, ROS-induced oxidations of nucleobases are important processes, as expected. We also found relatively high levels of fU (22% of fC). Clearly detectable were also caC and hmU. hmU, which is at the center of this study, was unequivocally detected, as shown by its retention time and its specific fragmentation pattern, which were found to be identical with that of the internal standard [D<sub>2</sub>]hmU (Fig. 1e). Notably, both caC and hmU were present in comparable amounts (5% of fC). In adult mouse cortex DNA (Fig. 1d; for other tissues see

<sup>1</sup>Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, München, Germany. <sup>2</sup>Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Berlin, Germany. <sup>3</sup>Department of Biomedicine, University of Basel, Basel, Switzerland. <sup>4</sup>Center for Integrated Protein Science at the Department of Pharmacy—Center for Drug Research, Ludwig-Maximilians-Universität München, München, Germany. <sup>5</sup>Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. <sup>6</sup>Department of Molecular Cancer Research, Cancer Genomics Netherlands, Utrecht, The Netherlands. <sup>7</sup>Present address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands. <sup>8</sup>These authors contributed equally to this work.

\*e-mail: markus.mueller@cup.uni-muenchen.de or thomas.carell@cup.uni-muenchen.de



**Figure 1 | Metabolism of cytosine derivatives, their detection by LC-UV ESI-MS/MS and levels in mESCs and mouse cortex.** (a) Potential active demethylation pathways. (b) Overlaid LC-UV and LC/MS/MS chromatograms of a representative DNA sample from mESCs. The dotted LC-UV chromatogram of C, G, T and A is scaled arbitrarily; the overlaid LC/MS/MS chromatograms of hmC, mC, fC and 8-oxo-G are scaled to the left y axis; the LC/MS/MS chromatograms of caC, hmU and fU are scaled to the right y axis. Red chromatograms refer to the corresponding labeled internal standards depicted in **Supplementary Figure 1**. AU, arbitrary units. (c,d) DNA modification levels per nucleoside (N) in mESCs (c; WT01,  $n = 7$ ) and 3-month-old mouse cortex tissue (d;  $n = 3$ ). Depicted are mean values  $\pm$  s.d. (e) Representative LC/MS/MS chromatograms for identification and quantification of hmU. Shown are the overlaid chromatograms for the two characteristic fragment ions of hmU (top trace) and the  $[\text{D}_2]\text{hmU}$  internal standard (bottom trace) derived from a mESC DNA sample. The insets show the MS/MS full-scan spectra of synthetic hmU and  $[\text{D}_2]\text{hmU}$  matching the proposed fragmentation pathway.

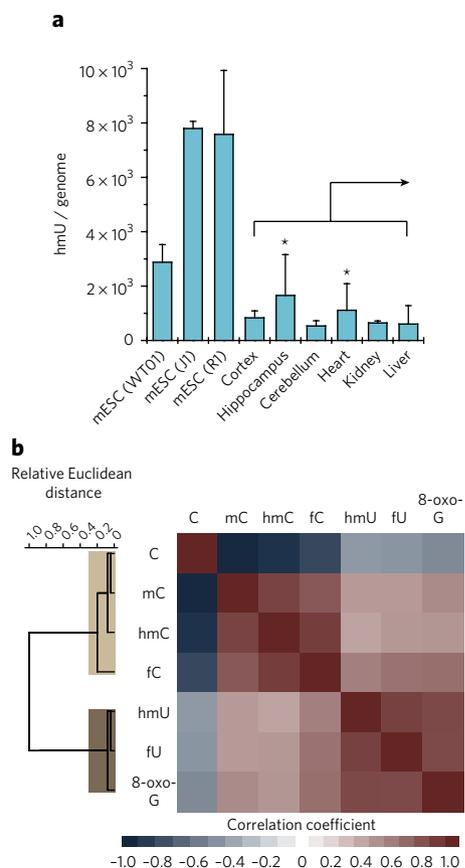
**Supplementary Fig. 2**), the hmC levels are very high (13% relative to mC), as previously reported<sup>19–21</sup>. In contrast, fC was detected only in very small amounts (0.03% relative to hmC), and caC was not observed at all<sup>22</sup>. The detected amounts of 8-oxo-G and fU are comparable. Notably, the detected levels of hmU in the cortex were significantly lower compared to that in mESCs (27%;  $P = 2.2 \times 10^{-4}$  by Student's *t*-test). Because the detected levels of 8-oxo-G were similar in the cortex (**Fig. 1d**) and in mESCs (**Fig. 1c**), the elevated hmU levels in mESCs cannot be explained by greater ROS-induced damage in mESCs and must have a different origin.

We next investigated this in more detail and quantified the levels of hmU in three different mESC lines (WT01, J1 and R1) and in a variety of tissues (**Fig. 2a**). Clearly, the hmU levels were higher in mESCs than in somatic tissues by factors of 2–15. The detected hmU levels correspond roughly to 500–1,700 hmU bases per genome in tissue and 2,900–7,800 in mESCs (**Fig. 2a**) depending on the cell type and growth conditions. We cannot explain these differences by elevated oxidative stress levels in mESCs (**Supplementary Fig. 3**). We next analyzed the levels of C, mC and the oxidized pyrimidines hmC, fC, hmU and fU as well as 8-oxo-G in somatic tissues (**Supplementary Fig. 2**) and performed a correlation and cluster analysis to reveal relationships of the modification levels (**Fig. 2b**). The data analysis confirmed that the low levels of hmU in somatic tissues correlate strongly (significant on a level  $< 0.001$ ) with the levels of fU and 8-oxo-G within a separated cluster.

This result showed that in somatic cells, hmU and fU are ROS-induced reaction products. If we assume that in somatic tissues hmU is exclusively formed by ROS, the data showed that in mESCs, 70–80% of the detected hmU is produced by ROS-independent processes (**Supplementary Fig. 3**).

### hmU formation by oxidation of T

To analyze the origin of the oxidized nucleobases in mESCs, we performed isotope tracing experiments (**Fig. 3a–c** and **Supplementary Fig. 4**). Substitution of L-methionine with  $[\text{methyl-}^{13}\text{CD}_3]\text{L-methionine}$  ( $[\text{methyl-}^{13}\text{CD}_3]\text{L-methionine}$ ) in the growth medium is known to give the labeled S-adenosylmethionine cofactor, which is needed for the conversion of C to mC. Replacement of L-Met by  $[\text{methyl-}^{13}\text{CD}_3]\text{L-Met}$  for 5 d (2 passages) in the medium at a concentration of 0.2 mM furnished 89% labeled  $[\text{methyl-}^{13}\text{CD}_3]\text{mC}$ , 88% labeled  $[\text{hydroxymethyl-}^{13}\text{CD}_3]\text{hmC}$  and 93% labeled  $[\text{formyl-}^{13}\text{CD}_3]\text{fC}$  (**Fig. 3b** and **Supplementary Fig. 4**). Within the detection limit (7 and 50 molecules per  $10^8$  nucleosides, respectively), no incorporation of the isotopes  $^{13}\text{C}$  and  $^{15}\text{N}$  into hmU and fU was observed, showing that hmC is not the precursor of hmU. We next added isotope-labeled  $[\text{thymidine-}^{13}\text{C},^{15}\text{N}_2]\text{thymidine}$  to the growth medium and observed  $\sim 76\%$  of label incorporation into T, hmU and fU (**Fig. 3a**, **Supplementary Fig. 4** and **Supplementary Table 1**). The combined data showed that hmU is not generated by deamination of hmC but by oxidation of T ( $\text{T} \rightarrow \text{hmU}$ ; **Fig. 3c**). Consequently, all of the detected hmU (and also fU) resides



**Figure 2 | hmU is present at elevated levels in mESCs compared to tissue.**

(a) hmU levels per genome in mESCs ( $n_{\text{WT01}} = 7$  replicates,  $n_{\text{I1}} = 2$ ,  $n_{\text{R1}} = 3$ ) and mouse tissue (3-month-old individuals,  $n = 3$ ). Levels per genome were obtained considering a mouse genome size of  $2.7 \times 10^9$  base pairs. Depicted are mean values  $\pm$  s.d. The differences between mESCs and mouse tissues are significant ( $*P = 2.9 \times 10^{-5}$  to  $3.6 \times 10^{-2}$ ; unpaired two-tailed  $t$ -test) except for WT01 and hippocampus ( $P = 0.292$ ) or heart ( $P = 0.069$ ). These exceptions are due to higher hmU levels in the hippocampus and heart caused by higher background oxidation (higher 8-oxo-G levels). hmU levels normalized to oxidative background (8-oxo-G) levels are in **Supplementary Figure 3**. (b) Unsupervised clustering analysis of Pearson correlation coefficients of 24 data sets of selected mouse organs at a defined time point (3-month-old individuals: cortex, hippocampus, cerebellum, heart, liver and kidney). Discussed correlations are strong to very strong (Pearson coefficient  $>0.7$ ) and significant on a level  $<0.001$ .

in an A base pair context (hmU:A). In this base pair, hmU is repaired by Smug1 but not Tdg<sup>12</sup>. This was confirmed by siRNA-mediated knockdown of Smug1 in mESC cells and HEK-293T cells. Indeed, the hmU level increased in these cells (**Supplementary Fig. 5**).

The fact that we were unable to detect hmU derived from hmC deamination, which would be situated in a base pair with G (hmU:G), could also be explained by very fast repair. If repair of the hmU:G base pair is extremely efficient, for example, because deamination and glycosylase-based repair occurs in a tight complex of the involved enzymes, we would be unable to detect this type of hmU because of low steady state levels. Indeed, it was proposed that hmC deamination and hmU excision requires a complex of the cytidine deaminase Aid and the glycosylase Tdg<sup>12</sup>. To assess this possibility, we performed isotope tracing experiments using [<sup>13</sup>CD<sub>3</sub>]<sub>L</sub>-Met in *Tdg*<sup>-/-</sup> mESCs stably complemented with either empty vector (control) or a minigene expressing a catalytically incompetent Tdg at near-endogenous levels. These cells are able to form the Aid-Tdg complex, but the Tdg is inactive, which should give elevated hmU

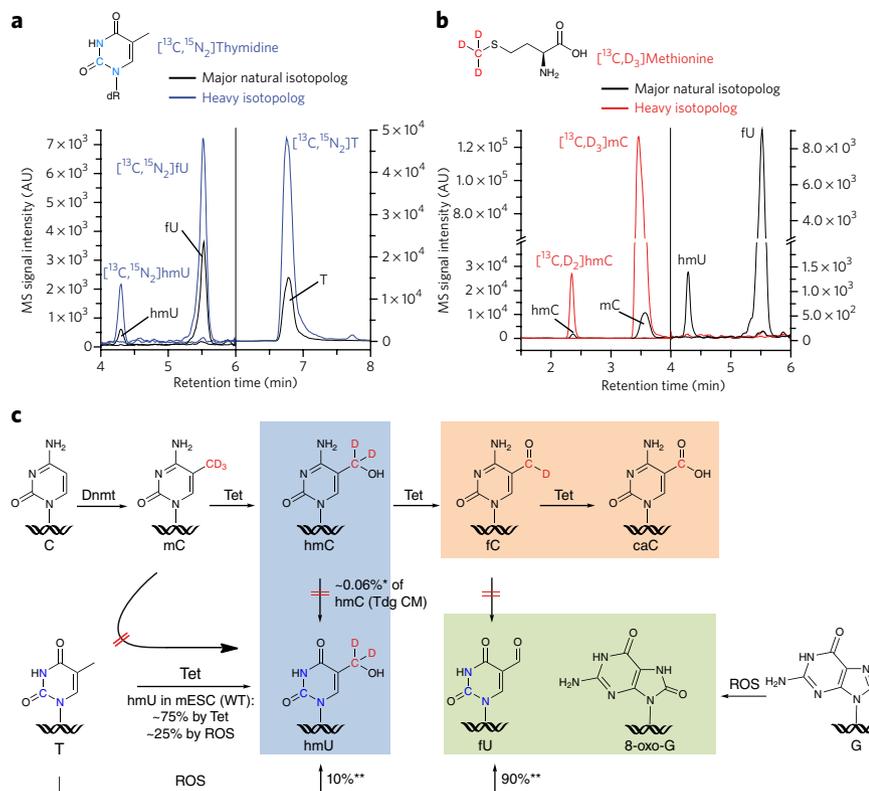
levels if deamination occurs. In both cell lines, we detected high levels of labeled hmC. In the control cells, [*hydroxymethyl*-<sup>13</sup>CD<sub>2</sub>] hmU was not detected. However, in the cell line complemented with inactive Tdg, some labeled [*hydroxymethyl*-<sup>13</sup>CD<sub>2</sub>]hmU was indeed observed, albeit only in small amounts ( $\sim 7\%$  of total hmU and  $\sim 0.06\%$  of total hmC; **Supplementary Fig. 6**). Thus, the Tdg protein is required for deamination of hmC to hmU. In wild-type (WT) mESCs, the hmU:G mismatches are obviously repaired so quickly that they do not contribute to steady state levels of hmU. All of the detected hmU was derived from T oxidation and resided in hmU:A base pairs.

### Tet enzymes form hmU in correlation with mC oxidation

We next investigated whether enzymatic oxidation or ROS-dependent processes are responsible for the observed oxidation of T to hmU. To this end, we studied mESCs with genetic or functional depletions of Tet or DNA methyltransferase (Dnmt) enzymes (**Fig. 4**). Although hmU levels were maintained at normal levels in the severely hypomethylated, Dnmt-depleted cell lines, they were substantially reduced upon knockdown of Tet1 and Tet2. As the levels of fU and 8-oxo-G remained unaffected, the data establish that, though both hmU and fU are generated by T oxidation, their formation must occur by two independent processes in mESCs. The data supported the idea that fU is a ROS-created lesion similar to 8-oxo-G<sup>9,10,17,18</sup>, whereas most hmU is generated by Tet-induced oxidation of T.

In support of enzymatic T oxidation by Tet proteins, ectopic expression of the catalytic domain of Tet1 (Tet1cd) in HEK-293T cells led to a 65-fold increase for both hmC and hmU levels. This was not observed when a catalytic mutant of Tet1 (Tet1cm) was expressed (**Supplementary Fig. 7a**). To determine whether under these conditions hmU is generated by deamination, as previously suggested<sup>11,12</sup>, we again replaced natural L-Met with [<sup>13</sup>CD<sub>3</sub>]<sub>L</sub>-Met in the medium, but we did not detect incorporation of heavy isotopes into hmU, providing evidence that the elevated hmU levels do not originate from hmC deamination (**Supplementary Table 1**). Finally, we confirmed enzymatic generation of hmU *in vitro* by incubating recombinant Tet1cd with a plasmid that was pre-methylated by bacterial methyltransferase M.SssI. In addition to oxidation of mC to hmC, fC and caC, we detected hmU at a prominent level of 9% relative to hmC (**Supplementary Fig. 7b**), showing that the catalytic center of the Tet enzymes clearly has the capacity to oxidize T to hmU.

To further confirm that hmU is formed in mESCs in the process of epigenetic reprogramming, we analyzed the dynamic changes of mC and all of the oxidized pyrimidines plus 8-oxo-G during differentiation. It was recently shown that mC and hmC levels sharply increase when mESCs maintained in the naive state are shifted to a primed state in serum-containing medium<sup>23-25</sup>. To investigate global kinetics of all of the Tet-generated oxidation products under more physiologically relevant priming conditions, we used established protocols based on serum-free N2B27 medium for differentiation of naive mESCs into states resembling that of post-implantation epiblasts<sup>26,27</sup>. mESCs were first grown for several passages in the presence of MEK and GSK3 inhibitors (dual inhibition or 2i conditions) and LIF to induce a hypomethylated state resembling that of the naive epiblast<sup>23-25</sup>. The data in **Figure 5a** are averaged from three independent differentiation experiments, each performed with two cell lines in the absence of growth factors. First, we observed that the fU and 8-oxo-G levels stayed constant (**Fig. 5a**), in line with the idea that both are ROS-derived products. In contrast, fC and caC levels peaked at about 8 h. Both hmU and hmC also peak between 8 h and 16 h. Isotope tracing with [<sup>13</sup>CD<sub>3</sub>]<sub>L</sub>-Met under these conditions provided no evidence for switched-on deamination of hmC to hmU (**Supplementary Table 1**). At these peak levels, we estimated that mESCs contain roughly 110,000 fC bases, 4,400 caC bases and



**Figure 3 | hmU and fU are thymine oxidation products in WT mESCs with no detectable contribution from hmC or fC deamination.** (a) Overlaid LC/MS/MS chromatograms of heavy ( $[^{13}\text{C},^{15}\text{N}_2]$ hmU,  $[^{13}\text{C},^{15}\text{N}_2]$ fU and  $[^{13}\text{C},^{15}\text{N}_2]$ T; blue) and major natural (black) isotopologs of hmU, fU and T. AU, arbitrary units. (b) Overlaid chromatograms of heavy ( $[^{13}\text{C},\text{D}_3]$ mC,  $[^{13}\text{C},\text{D}_3]$ hmC,  $[^{13}\text{C},\text{D}_3]$ hmU and  $[^{13}\text{C},\text{D}_3]$ fU; red) and the major natural (black) isotopologs of mC, hmC, hmU and fU. (c) Enzymatic and ROS-dependent pathways leading to the formation of hmC, fC, caC, hmU and fU. Single asterisks denote labeled hmU generated by deamination of labeled hmC, which was observed only in Tdg catalytic mutant (CM) cells (Supplementary Fig. 6), representing ~7% of the total hmU content and corresponding to deamination of ~0.06% hmC. Double asterisks denote basal rates of ROS-dependent T oxidation, which were determined in HEK-293T cells, where Tet activity is lowest (3.6 hmC per  $10^5$  nucleosides; Supplementary Table 2). Here, hmU and fU roughly represent 10% and 90% of T oxidation products relative to the sum of each other.

14,000 hmU bases per genome, showing that, at its peak level, hmU is three times more abundant than caC. The dynamic peaking data allowed us to estimate half-life times for fC (7 h), caC (5 h) and hmU (4 h) during the differentiation process (Supplementary Fig. 8). Using the dynamic quantitative data, we performed a correlation and cluster analysis of the DNA modification levels (Fig. 5b). Three independent clusters were obtained. One cluster involves members of cytosine methylation (and demethylation) dynamics, including C, mC, fC and caC, but, to our surprise, not hmC. Instead, hmC groups with hmU. We therefore concluded that the formation of these modifications is tightly coupled by the action of the Tet enzymes. The ROS-induced lesions fU and 8-oxo-G form the third, well-separated cluster. Notably, hmU does not correlate with the ROS lesions.

Recently, it was shown that exposure of naive mESCs to fibroblast growth factor 2 (FGF-2) and activin A (ActA) for 48 h under similar conditions as those described above gives a homogeneous cell population whose transcriptome closely resembles that of the post-implantation epiblast<sup>26</sup>. Under these conditions, mC levels increased more rapidly, approaching somatic levels within a time frame closely reflecting that observed during embryonic development (Supplementary Fig. 9a). Levels of hmC rose steadily throughout the 48-h time course. Whereas fC peaked at 36 h, hmU reached its maximum at 24 h (Supplementary Fig. 9b). qPCR data

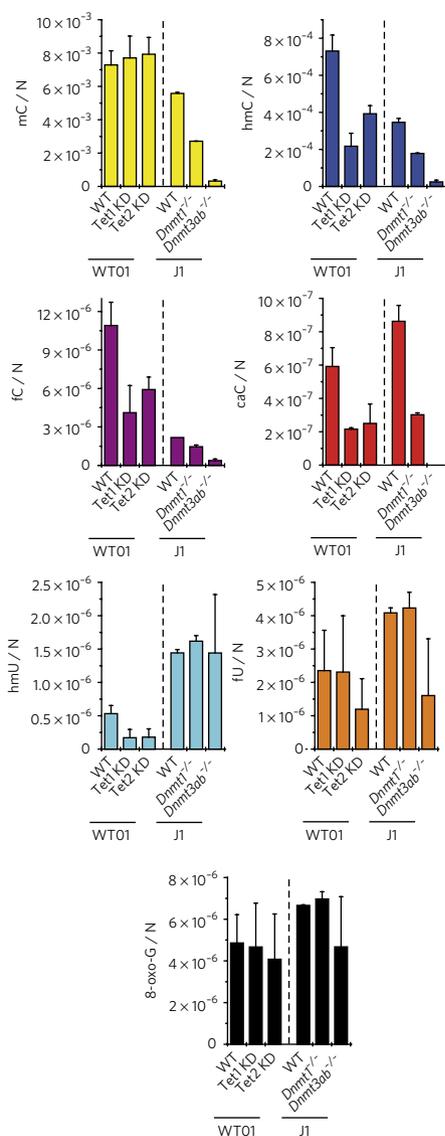
showed a robust induction of Dnmt3b that is most likely responsible for the increasing mC levels (Supplementary Fig. 9a). A transient peak of Tet1 expression at 24 h went in hand with rising hmU and fC levels, whereas Tet3 expression started rising slowly from the same time point and may drive further accumulation of hmC and fC (Supplementary Fig. 9b). Tdg was progressively upregulated together with the onset of methylation activity, whereas Smug1 showed little, if any, change (Supplementary Fig. 10). These data supported our view that hmU peaking is not caused by downregulation of the repair pathway.

Thus, the kinetic data showed a complex interplay between methylation and two oxidation reactions during differentiation, which depends on the exact conditions. Most important is the fact that hmU showed a time-dependent occurrence similar to the other oxidized bases hmC, fC and caC independently of the conditions investigated, confirming its formation during epigenetic reprogramming.

### hmU attracts specific readers

To obtain initial insight into a potential biological function of hmU in comparison to hmC, we screened for specific readers associating with hmU:A as well as with hmC:G and hmU:G using protein pull-down and relative quantification by LC/MS/MS. In previous stable isotope labeling by amino acids in cell culture (SILAC)-based proteomics studies with hmC, fC and caC containing oligonucleotides, we and others observed a high number of specific protein readers, arguing that the new bases influence a variety of different processes<sup>28,29</sup>. For this hmU study, we further developed this approach for the detection of only those proteins, which directly interact with the modified bases hmU (as well as hmC for comparison) to get a more direct insight into their function (Fig. 6a). For

the study, we not only included into the biotinylated DNA duplexes (24mers) an hmU (hmC) base but also equipped the counter strand with a polyethylene glycol-based linker carrying a reactive *N*-hydroxysuccinimide (NHS) ester moiety and a reductively cleavable disulfide bond in the middle. This reactive linker cannot be inserted into the DNA strand using solid phase synthesis. In addition, the NHS ester does not survive hybridization conditions. We therefore attached the linker as its azide derivative, using Cu(I)-catalyzed click chemistry, to an alkyne-bearing base present in the DNA duplex<sup>30,31</sup>. These DNA duplexes were subsequently incubated with nuclear extracts from mESCs. The DNA-bound protein complexes were isolated using streptavidin-coated magnetic beads (Fig. 6a)<sup>32</sup>. DNA duplexes with the canonical base pairs A:T and C:G at the respective positions served as reference strands. The NHS linker has two functions in the experiment: First, it will covalently trap the reader proteins specifically at the  $\epsilon$ -amino groups of lysines, which allows the identification of transiently binding proteins as well. Second, because the linker is cleavable, it will tag the trapped lysine residues, leaving a defined label on those proteins that bind in close proximity to the modified bases hmC and hmU. This allowed us to distinguish protein readers that bound close to hmC and hmU from proteins that are secondary members of the complexes. The covalently trapped proteins were next tryptically digested and labeled with tandem mass tagging (TMT) isobaric tags to allow protein

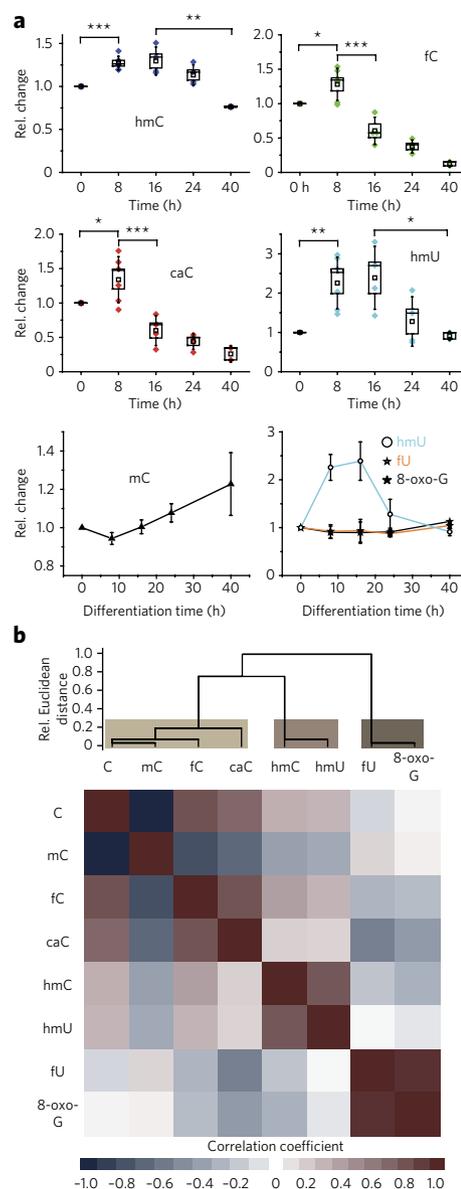


**Figure 4 | Tet1 and Tet2 generate hmU in mESCs.** Effect of Tet and Dnmt depletion on DNA modification levels in mESCs. J1 cell lines bearing homozygous *Dnmt*-null mutations and WT01 mESCs expressing Tet1 or Tet2 shRNAs (knockdown (KD)) were analyzed. Depicted are mean values per nucleoside (N)  $\pm$  s.d. of biological replicates as follows: wild type (WT01,  $n = 7$ ), Tet1 KD (WT01,  $n = 3$ ) and Tet2 KD (WT01,  $n = 3$ ); WT (J1,  $n = 2$ ), *Dnmt1*<sup>-/-</sup> (J1,  $n = 2$ ) and *Dnmt3a/b*<sup>-/-</sup> (J1,  $n = 6$ ).

identification and quantification by MS<sup>32,33</sup>. Only proteins that were enriched in both the forward and the reverse TMT experiment were considered to be specific hmU (hmC) readers (Fig. 6). All of the readers were subsequently divided into two groups. Proteins that were enriched relative to the control strands are termed specific readers (sRs). Of those, the proteins that were identified with a peptide containing the tag are termed direct-specific readers (dsRs).

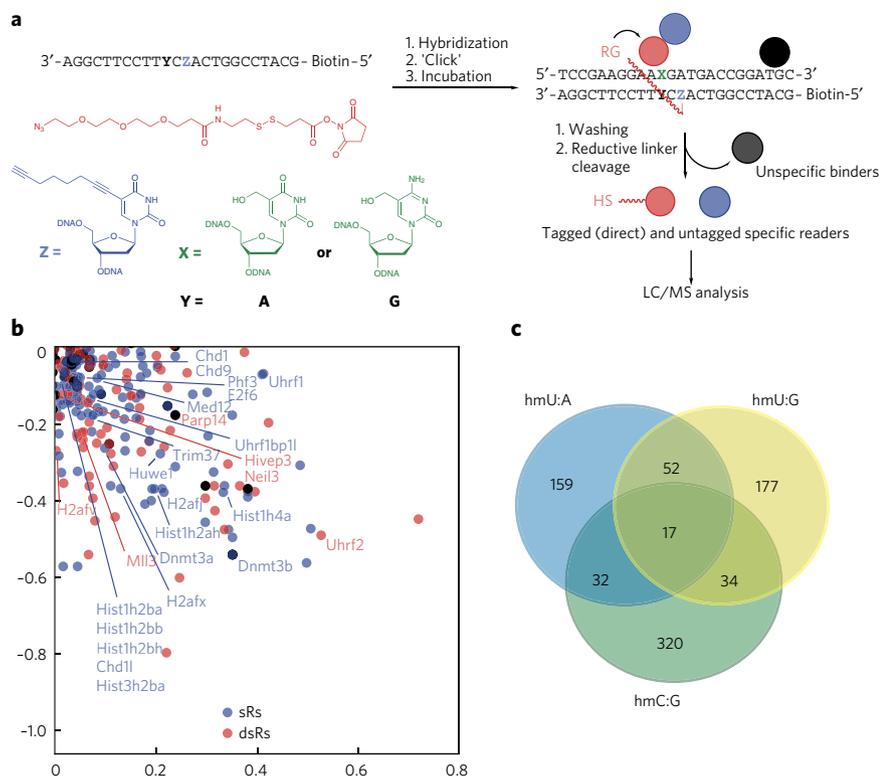
The scatter plot (Fig. 6b and Supplementary Fig. 11a) revealed over 250 sRs recognizing the hmU:A base pair (relative to T:A). Of those proteins, 99 were identified with tagged peptides. These proteins were therefore designated dsRs. A gene ontology analysis showed that 64 proteins of the 250 sRs are nucleotide binders, and 25 proteins are involved in chromatin organization (Supplementary Fig. 11b).

Among the dsRs we identified were the regulatory proteins Uhrf2, transcription factor HIVEP3 or the poly [ADP-ribose] polymerase



**Figure 5 | hmU is produced during mESC differentiation.** (a) Time course of DNA modification levels during early differentiation (0–40 h) of naive mESCs in the absence of growth factors. Box plot of relative modification changes for hmC, fC, caC and hmU averaged from six independent experiments using two different mESC lines. In the four upper plots, colored diamonds reflect mean values of technical triplicate measurements, open squares reflect mean values of biological replicates, boxes represent the s.e.m., and whiskers represent the s.d. of the biological replicates; \* $P < 0.05$ , \*\* $P < 0.01$  and \*\*\* $P < 0.005$  by unpaired two-tailed *t*-test. In the two lower plots, the relative modification changes of mC, hmU, fU and 8-oxo-G are depicted as biological mean values  $\pm$  s.e.m. (b) Unsupervised clustering analysis of Pearson correlation coefficients of 22 data sets obtained from the differentiation experiments depicted in a. Discussed correlations are strong to very strong (Pearson coefficient  $> 0.7$ ) and significant on a level  $< 0.001$ .

14 (Parp14), which showed that these proteins interacted directly with hmU. We also detected several H2A and H2B histone proteins among the dsRs and the sRs. Notably, in the group of the sRs, we observed also Dnmt3a and Dnmt3b, which are involved in gene expression regulation. Methyl CpG-binding protein 2 (MeCP2) was identified as well, but it showed only low enrichment. We also detected several transcriptional regulators, such as bromodomain



**Figure 6 | Identification of hmU:A readers.** (a) Workflow of pull-down experiments with hmU:A-containing oligonucleotides. DNA oligomers with the modified base Z containing an alkyne group for click reaction and biotin for the pull-down are hybridized with DNA oligomers containing hmU (X = hmU and Y = A). After the click reaction with the DNA-protein cross-linker (red), which has an NHS ester as a lysine-specific reactive group (RG), the dsDNA oligomers are incubated with a nuclear extract from mESCs, and specific binders are covalently linked to the DNA. After three washing steps, the linker is reductively cleaved, and the proteins are enzymatically digested and labeled with TMT2plex reagents. The same dsDNA oligomer without hmU modification (X = T and Y = A) was used as a control sample. Specifically enriched proteins are mixed with the control sample before LC/MS analysis. The same workflow was used for experiments with hmU:G or hmC:G and C:G as a control. (b) Magnified view of proteins enriched with hmU:A-containing oligomer. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross-linker and marked in red. Black dots are considered unspecific binders. A full-scale representation is shown in **Supplementary Figure 11**. (c) Venn diagram showing overlap of specific binders for hmU:A, hmU:G and hmC:G.

PHD finger transcription factor (Bptf), transcription factor E2F6 and mediator of RNA polymerase II transcription subunit 12 (Med12). Moreover, we identified hmU-binding E3 ubiquitin-protein ligases such as Uhrf1, Trim37 and Huwe1 and finally the chromodomain helicase DNA-binding proteins 1 and 9 (Chd1 and Chd9) among the hmU readers, which are chromatin remodeling factors and regulate polymerase I and II transcription. To exclude bias by sequence-specific binding, we repeated the experiment with hmU in a different sequence context. Indeed, 78% of the identified protein readers are also observed in this control experiment, and included among them were all of the proteins discussed above, with the notable exception of Uhrf1, which was not observed in this experiment (**Supplementary Data Set 1**). The data show that hmU:A recruits specific proteins that are involved in chromatin biochemistry in a broad sense.

As a first validation of the MS-identified readers, we focused on Uhrf1 and Uhrf2. Uhrf1 is a known mC and hmC binder<sup>34</sup>, whereas we recently identified the homolog Uhrf2 as a specific binder of hmC<sup>29</sup>. The MS data for hmU obtained in this study showed an enrichment of Uhrf1 and Uhrf2. In addition, Uhrf2 was identified as a direct binder. To validate the MS findings, we overexpressed

Uhrf1 or Uhrf2 together with Tet1cd in HEK-293T<sup>29</sup> and measured the levels of mC, hmC, fC, caC and also hmU. Indeed, we observed increased levels of hmU, proving a functional relation between hmU levels and Uhrf1 and Uhrf2 (**Supplementary Fig. 12**). As Uhrf2 is known to interact with Dnmt3a and Dnmt3b<sup>35</sup> and these two proteins are also identified as sRs of hmU, our initial data provide a consistent picture.

Analysis of the specific hmU readers showed that 49 of the identified proteins were also enriched with an hmC-modified strand. Among these proteins were, for instance, bromodomain PHD finger transcription factor (Bptf), MeCP2 and 19 other nucleotide-binding proteins. The obtained list of specifically binding proteins is in good agreement with our recently published data set<sup>29</sup>, which further supports the validity of the covalent trapping method. Finally, we compared the results of the hmU:A pull-down with a pull-down of hmU:G, which would be the product of hmC deamination. We found 69 of 280 sRs in both experiments. We identified transcription regulators as sRs as well in the hmU:G pull-down. However, we also found several DNA repair proteins, such as DNA repair protein complementing XP-C cells homolog (XPC), DNA repair protein Rad50 and five helicases, among others (**Supplementary Data Set 1**).

## DISCUSSION

This work aimed to analyze the origin of hmU in mESCs. Although hmU as well as fU are both well-characterized oxidation products of thymidine<sup>9,10</sup>, hmU in particular was recently discussed as an intermediate in active demethylation by deamination of hmC<sup>11,12</sup>. We used stable isotopologs of the main oxidation products hmC, fC, caC, hmU, fU and 8-oxo-G in combination with quantitative MS to show that the levels of hmU are strongly elevated in stem cells in comparison to somatic tissue. Notably, we did not observe equally elevated

levels of fU and 8-oxo-G, both of which are formed by the reaction of DNA with ROS. This analysis suggests that hmU in stem cells is not only formed by ROS. The quantitative data allowed us to estimate that in WT stem cells, about 70–80% of the found hmU is not formed by ROS but is derived from a different process. Isotope tracing studies with labeled [<sup>13</sup>CD<sub>3</sub>]-L-Met and [<sup>13</sup>C,<sup>15</sup>N<sub>2</sub>] thymidine showed that the detected hmU originates exclusively from T. Previous studies already proposed a connection between Tet activity and hmU levels, but these studies did not address the origin of hmU, which led to the hypothesis that hmU could form by deamination of hmC<sup>11,22</sup>. To investigate the involvement of deamination<sup>11–14</sup>, we studied TDG<sup>-/-</sup> stem cells reconstituted with a catalytically inactive TDG. In these cells, the putative deamination-repair complex<sup>12</sup> of TDG and Aid could form, and we expected to see elevated hmU. Indeed, in this experiment, we were able to detect labeled hmU, showing that deamination of hmC does occur, but the levels were low. Most notably, special conditions were required to detect this deamination. In WT cells, all of the detected hmU is clearly derived from T oxidation. As such, the detected hmU is situated in a base pair with A (hmU:A). In conclusion, the majority of hmU in mESCs is produced independently of both ROS and deamination.

Using a combination of functional knockdown in mESCs, ectopic expression of Tet1 catalytic domain in HEK-293T cells and *in vitro* studies with recombinant Tet protein, we showed that the Tet enzymes, known to oxidize mC to hmC, are also responsible for oxidation from T to hmU. A kinetic study in which the temporal changes of all of the oxidized pyrimidines were observed under two different differentiation kinetics showed finally that hmU is formed in processes that also generate hmC and fC. Together with the observed peak levels of hmU of about 14,000 bases per genome, this suggests that hmU could have an epigenetic function similar to hmC. It is conceivable that the Tet enzymes introduce a small amount of hmU during the hmC manufacturing process to trigger faster demethylation by induction of DNA repair. It is known that hmC is not a substrate for repair<sup>8,12</sup>, whereas hmU is efficiently recognized and repaired<sup>36</sup>, for example, by the repair glycosylase Smug1 (refs. 37,38). Sporadic introduction of hmU could therefore allow recruitment of repair factors, for example, for long patch repair, as recently suggested<sup>39</sup>. Alternatively, hmU might influence transcription factor binding, which is known to be a hallmark of epigenetic reprogramming<sup>40</sup>. To study the proteins that closely interact with hmU, a new pulldown experiment was devised in which we inserted a trapping linker into the oligonucleotide. This linker is able to react covalently with proteins that assemble on the DNA duplex close to the hmU base. Because the linker is reductively cleavable, it leaves a tag on the respective lysine residues, which is detected in subsequent MS-based proteomics measurements. Using this technology, we observed that hmU:A recruits transcription factors and proteins that are involved in chromatin biochemistry, including Uhrf1 and Uhrf2. This result is in line with the recent observation that oxidative lesions in DNA can change the binding affinity of the transcription factor CREB when they are present in the respective cognate sequence<sup>41</sup>. A further analysis in HEK-293T revealed that the hmU readers Uhrf1 and Uhrf2 are able to modulate the levels of all of the oxidized mC bases and of hmU. It is interesting to note that the observed level increases of hmC were small, whereas they were substantial for hmU and also for the further oxidized bases fC and caC. As Uhrf1 and Uhrf2 are proteins known to be involved in numerous epigenetic processes, our data supported the hypothesis that Tet-induced oxidation of T to hmU may have an epigenetic function. Even in the case that hmU is formed just as a side product of Tet-induced hmC formation, it is now clear that the oxidation chemistry performed by the Tet enzymes has to go in hand with effective DNA repair.

Received 15 December 2013; accepted 17 April 2014;  
published online XX XX 2014

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

## References

- Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
- Franchini, D.M., Schmitz, K.M. & Petersen-Mahrt, S.K. 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu. Rev. Genet.* **46**, 419–441 (2012).
- Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
- Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
- Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
- He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
- Tan, L. & Shi, Y.G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895–1902 (2012).
- Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
- Bjelland, S. *et al.* Oxidation of thymine to 5-formyluracil in DNA: mechanisms of formation, structural implications, and base excision by human cell free extracts. *Biochemistry* **34**, 14758–14764 (1995).
- Mouret, J.F., Polverelli, M., Sarradini, F. & Cadet, J. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chem. Biol. Interact.* **77**, 187–201 (1991).
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
- Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
- Nabel, C.S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
- Rangam, G., Schmitz, K.M., Cobb, A.J. & Petersen-Mahrt, S.K. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS ONE* **7**, e43279 (2012).
- Jacobs, A.L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
- Zhang, Q.M. *et al.* DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA Repair (Amst.)* **4**, 71–79 (2005).
- Burrows, C.J. Surviving an oxygen atmosphere: DNA damage and repair. *ACS Symp. Ser. Am. Chem. Soc.* **2009**, 147–156 (2009).
- Taghizadeh, K. *et al.* Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nat. Protoc.* **3**, 1287–1298 (2008).
- Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
- Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
- Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
- Liu, S. *et al.* Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res.* **41**, 6421–6429 (2013).
- Ficz, G. *et al.* FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
- Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
- Leitch, H.G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
- Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186 (2003).
- Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
- Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
- Burley, G.A. *et al.* Directed DNA metallization. *J. Am. Chem. Soc.* **128**, 1398–1399 (2006).
- Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. A stepwise Huisgen cycloaddition process: copper(i)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* **41**, 2596–2599 (2002).
- Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003); erratum **75**, 4942 (2003); erratum **78**, 4235 (2006).
- Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
- Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306 (2011).
- Pichler, G. *et al.* Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J. Cell. Biochem.* **112**, 2585–2593 (2011).
- Lewis, H.L., Muhleman, D.R. & Ward, J.F. Serologic assay of DNA base damage. I. 5-Hydroxymethyldeoxyuridine, a radiation product of thymidine. *Radiat. Res.* **75**, 305–316 (1978).
- Boorstein, R.J. *et al.* Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J. Biol. Chem.* **276**, 41991–41997 (2001).

38. Kavli, B., Otterlei, M., Slupphaug, G. & Krokan, H.E. Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst.)* **6**, 505–516 (2007).
39. Santos, F. *et al.* Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin* **6**, 39 (2013).
40. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132**, 532–536 (2008).
41. Moore, S.P.G., Toomire, K.J. & Strauss, P.R. DNA modifications repaired by base excision repair are epigenetic. *DNA Repair (Amst.)* **12**, 1152–1158 (2013).

## Acknowledgments

We thank the Excellence Cluster Center for Integrated Protein Science Munich (CiPS<sup>M</sup>) and the collaborative research centers SFB749, SFB646 and SFB1032 as well as German Research Foundation (DFG) grant CA275/8-4, the Volkswagen foundation, NGFNplus (01GS0870) and the Netherlands Organization for Scientific Research (NWO-VIDI) for financial support. T.P. and S.S. thank the Fonds der Chemischen Industrie for predoctoral fellowships. C.B. thanks the Boehringer Ingelheim Fonds for a predoctoral fellowship. G.K. thanks the Japan Society for the Promotion of Science (JSPS) for a postdoctoral fellowship for research abroad. We thank M. Moser (Max Plank Institute for Biochemistry) for providing R1- and C57Bl6/129-derived mESCs, G. Höfner and K.T. Wanner for their initial help with MS as well as M. Wirsing, L. Belzner and P. Laube for providing bioinformatic tools for data processing.

## Author contributions

T.P. synthesized MS standards, performed the sample preparation and ultra high-performance LC/MS/MS method development, did LC/MS analysis, interpreted data and performed statistical analysis. F.S. and N.R. performed mESC differentiation and isotope tracing experiments. M.W. and C.B. performed the HEK-293T experiments, did LC/MS analysis and interpreted data. C.B. performed qPCR and analyzed the data. S.L. and D.E. performed the protein pulldown studies and interpreted data. M.T. performed mESC knockdown and knockout experiments. J.S. and O. Kosmatchev did sample preparation and LC/MS analysis. B.H., S.S. and J.S. prepared MS standards. O. Kotljarova performed *in vitro* assays. B.S. synthesized oligonucleotides for protein capture. G.K. synthesized tandem mass tags, and S.M. provided mouse tissue samples. U.M. and H.L. constructed Tet expression plasmids. C.G.S. and M.V. performed Uhrf1/2 over-expression in HEK-293T cells. P.S. and D.S. provided plasmids and cell lines. M.M. and T.C. conceived and supervised the project, interpreted data and wrote the manuscript.

## Competing financial interests

The authors declare no competing financial interests.

## Additional information

Supplementary information is available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to M.M. and T.C.

## ONLINE METHODS

**General materials and methods.** Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR or Acros organics and used without further purification. Acetonitrile of LC/MS grade was purchased from Carl Roth GmbH + Co., KG. Formic acid, p.a. for MS, was purchased from Fluka, and water was purified with a Milli-Q Plus system from Merck Millipore.

The MS standards 5-methyl-2'-deoxycytidine (mC), 5-trideuteromethyl-2'-deoxycytidine ( $[D_3]mC$ ), 5-hydroxymethyl-2'-deoxycytidine (hmC), 5-dideuterohydroxymethyl-2'-deoxy- ( $N^1, N^3-^{15}N$ )-cytidine ( $[^{15}N_2, D_2]hmC$ ), 5-formyl-2'-deoxycytidine (fC), 5-formyl-2'-deoxy- ( $N^1, N^3-^{15}N$ )-cytidine ( $[^{15}N_2]fC$ ), 5-carboxy-2'-deoxycytidine (caC), 5-carboxy-2'-deoxy- ( $N^1, N^3-^{15}N$ )-cytidine ( $[^{15}N_2]caC$ ), 5-hydroxymethyl-2'-deoxyuridine (hmU), 5-(dideuterohydroxymethyl)-2'-deoxyuridine ( $[D_2]hmU$ ), 5-formyl-2'-deoxyuridine (fU) and 5-formyl-2'-deoxy- ( $N^1, N^3-^{15}N$ )-uridine ( $[^{15}N_2]fU$ ) were synthesized according to earlier published work<sup>19,21,42</sup>. All of the synthesized compounds were characterized and purity confirmed by <sup>1</sup>H-NMR, <sup>13</sup>C-NMR and ESI-MS, and some were additionally validated by <sup>15</sup>N-NMR. 8-hydroxy-2'-deoxy- ( $^{15}N_5$ )-guanosine ( $[^{15}N_5]8\text{-oxo-G}$ ) (99 atom% <sup>15</sup>N) was purchased from Cambridge Isotope Laboratories; 8-hydroxy-2'-deoxyguanosine (8-oxo-G) was from BIOLOG; 2'-deoxyguanosine (G) and 2'-deoxycytidine (C) were from ChemGenes. (methyl-<sup>13</sup>C, <sup>D</sup><sub>3</sub>)-L-Met (99 atom% D and <sup>13</sup>C) was purchased from Sigma-Aldrich, and 2'-deoxy- ( $C^2-^{13}C, N^1, N^3-^{15}N$ )-thymidine (99 atom% <sup>13</sup>C and <sup>15</sup>N) from Hartmann Analytic. Aqueous stock solutions of these compounds were stored at -20 °C and warmed up to RT before usage.

**Oligonucleotide synthesis.** Oligonucleotide synthesis was performed on an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions (DMT off) and acetyl-protected dC. Phosphoramidites (including the 5-hydroxymethyl-dU-CE phosphoramidite and the 5'-biotin phosphoramidite) and polystyrene carriers were obtained from Glen Research. 5-octadienyl-dU (Z) phosphoramidite was synthesized according to literature and characterized by <sup>1</sup>H-NMR, <sup>13</sup>C-NMR, <sup>31</sup>P-NMR and ESI-MS<sup>43</sup>. The crude oligonucleotide was cleaved from the resin and deprotected in 30% (v/v) ammonium hydroxide and 40% (v/v) methylamine (1:1) at 65 °C for 10 min and purified by preparative and analytical HPLC (Waters Breeze and Alliance, respectively). Separation was performed by applying a VP 250/10 Nucleosil 100-7 C18 column (flow: 5 ml/min) from Macherey-Nagel with a gradient of buffer A (0.1 M NH<sub>4</sub>OAc in water) and buffer B (0.1 M NH<sub>4</sub>OAc in 80% MeCN). DNA-containing fractions were characterized by MALDI-TOF (Bruker Autoflex II) and analytical HPLC, combined, and desalted by C18-Sep-Pak cartridges (Waters). For analytical HPLC, separation was performed by applying a CC 250/4 Nucleosil 120-3 C18 column from Macherey-Nagel (flow: 0.5 ml/min) with the aforementioned buffer system.

For the protein pull-down studies, an hmU-containing oligonucleotide was hybridized with a DNA strand modified with a 5'-biotin and a 5-octadienyl-dU. The sequences of the DNA strands are summarized in **Supplementary Table 3**. To 10 nmol of the dsDNA, 0.5 μl of a 200 mM solution of the cross-linking azide-PEG<sub>3</sub>-S-S-NHS ester (C<sub>18</sub>H<sub>29</sub>N<sub>5</sub>O<sub>8</sub>S<sub>2</sub>, Jena Bioscience, Jena, Germany) in DMSO was added. In a separate tube, CuBr was dissolved in a TBTA solution (DMSO/tBuOH 3:1, 100 mM) resulting in a 1:1 Cu(I):TBTA ratio. This solution was immediately added to the DNA/azide mixture. Furthermore, 45 μl of DMSO/tBuOH (3:1) were added, the mixture was shaken at 37 °C for 3 h, and the resulting product was purified by ethanol-precipitation.

**LC/MS analysis of DNA samples.** Quantitative LC-UV-ESI-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with an UV detector and an Agilent 6490 triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. An improved method, based on earlier published work<sup>29,42,44,45</sup>, was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature 50 °C, gas flow 15 l/min (N<sub>2</sub>), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 l/min (N<sub>2</sub>), capillary voltage 2,500 V in the positive ion mode, capillary voltage -2,250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 4** and **5**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 μm, 2.1 mm × 150 mm) at 30 °C using a gradient of water

and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 ml/min: 0 → 5 min; 0 → 3.5% (v/v) MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 7.2 min; 5 → 80% MeCN; 7.2 → 10.5 min; 80% MeCN; 10.5 → 11.3 min; 80 → 0% MeCN; 11.3 → 13 min; 0% MeCN. The effluent up to 1.5 min and after 9 min was diverted to waste by a Valco valve. The autosampler was cooled to 10 °C. The injection volume was amounted to 29 μl. Calibration curves, method validation and data processing are in **Supplementary Note 2**. A complete compilation of LC/MS quantifications results see **Supplementary Note 3**.

**DNA digestion.** 5–25 μg of genomic DNA in 25 μl H<sub>2</sub>O were digested as follows: An aqueous solution (7.5 μl) of 480 μM ZnSO<sub>4</sub>, containing 42 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards (**Supplementary Note 2**) were added, and the mixture was incubated at 37 °C for 3 h. After addition of 7.5 μl of a 520 μM [Na]<sub>2</sub>-EDTA solution, containing 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for another 3 h at 37 °C and then stored at -20 °C. Prior to LC/MS/MS analysis, samples with up to 15 μg DNA, for which the quantification of low amounts of caC was aspired, were filtered by using an AcroPrep Advance 96 filter plate 0.2 μm Supor (Pall Life Sciences). In contrast, samples with 15–25 μg DNA (isotope-tracing experiments) were filtered by using an AcroPrep Advance 96 filter plate 10K Omega (Pall Life Sciences).

**Genomic DNA isolation.** Tissues of female WT mice (C57-Bl/6N) were dissected at postnatal day 90 and prepared as earlier described<sup>19,21</sup>. Genomic DNA was extracted using the Qiagen Blood and Cell Culture DNA Midi Kit except for mESC samples differentiated in the presence of growth factors (see below). Extraction was performed following the manufacturer's instructions for genomic DNA isolation from cell culture samples or tissue samples, respectively. All buffers until loading of the sample on Genomic-tip 100/G were additionally supplemented with antioxidants 2,6-di-*tert*-butyl-4-methylhydroxytoluene (BHT, 200 μM) and deferoxamine mesylate salt (desferal, 200 μM) as well as the deaminase inhibitor tetrahydrouridine (THU, 200 μM), according to published standards, to reduce background oxidation or deamination<sup>18</sup>. Elution buffer QF was supplemented with 200 μM BHT. Following elution, all steps were performed on ice. DNA was then precipitated with NaOAc (0.3 M final) and 0.7 volumes *i*PrOH. DNA pellets from cultured cells were washed twice with ice-cold 70% EtOH and resuspended in H<sub>2</sub>O containing 20 μM BHT using a Qiagen TissueLyser (30 Hz, 2 min). DNA pellets from mouse tissues were resuspended in PBS buffer and additionally extracted with phenol/CHCl<sub>3</sub>, precipitated, washed and resuspended as described above.

R1 mESC samples differentiated in the presence of growth factors or transfected with Smug1 esiRNAs were lysed directly in the plates with RLT buffer (Qiagen) supplemented with BHT and desferal as described above. DNA was isolated using the Zymo Quick gDNA Midi Kit according to the manufacturer's instruction, except that elution was repeated four times with 100 μl of elution buffer supplemented with BHT (200 μM). Eluted DNA was precipitated with 2 M ammonium acetate and two volumes of absolute ethanol and finally resuspended in H<sub>2</sub>O containing 20 μM BHT. The flow-through from the spin columns was used to isolate RNA (see real-time PCR analysis).

**mESC cell culture.** Feeder independent WT01 mESCs (C57BL/6 strain)<sup>46</sup> were cultured in the presence of serum and LIF as previously described<sup>4</sup>. *Tdg*<sup>-/-</sup> and *Tdg*<sup>-/-</sup> mESCs were described previously<sup>47</sup>. *Tdg*<sup>-/-</sup> mESCs were complemented by random integration of either empty vector (hereafter referred to as *Tdg*<sup>-/-</sup> mESCs) or a minigene expressing catalytically inactive Tdg (N151A)<sup>47</sup>. Clonal mESC lines with targeted Tdg alleles, R1 cells (strain 129/Sv)<sup>48</sup>, J1 cell lines (strain 129S4/SvJae)<sup>49</sup> and a mESC line derived from C57Bl/6/129 mixed background<sup>50</sup> were routinely maintained on gelatinized plates in DMEM (PAA or Sigma) supplemented with 10% FBS, 1 × MEM-nonessential amino acids (NEAA), 0.2 mM L-alanyl-L-glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin (all from PAA), 0.1 mM β-mercaptoethanol, 20 ng/ml ( $\geq 1 \times 10^3$  U/ml) mouse recombinant LIF (ORF Genetics), 1 μM PD 0325901 and 3 μM CHIR 99021 (2i; both from Axon Medchem). In these conditions, the global levels of genomic mC were very low (and, as a consequence, the levels of its oxidized derivatives were even lower; data not shown). Before DNA isolation, 2i cultures were passaged twice (over 5 d) in DMEM supplemented with FBS and LIF as above but lacking 2i. With this strategy, primed mESC cultures were obtained with no sign of overt differentiation and modified genomic cytosines reached reproducibly higher and stable levels. For isotope tracing with heavy

thymidine in serum-primed mESCs 2i cultures of R1, cells were passaged twice (5 d) in the same serum-containing medium lacking 2i and simultaneously supplemented with 100  $\mu\text{M}$  [ $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ T. For isotope tracing with heavy methionine in serum-primed mESCs, 2i cultures of R1 cells and mESC lines with targeted Tdg alleles were passaged twice (over 5 d) without 2i in L-Met-free DMEM (Life Technologies) supplemented as above and with 0.2 mM of either [methyl- $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ -L-Met or natural L-Met. For mESC differentiation without growth factors, R1 cells and the C57Bl/6/129 mixed background cell line were first plated at  $1 \times 10^5$  cells/cm $^2$  on gelatin-coated plates in N2B27 medium containing 1,000 U/ml LIF to favor attachment and initial survival<sup>51</sup>. After 12 h, the medium was replaced without addition of LIF (defined as time point 0 h). The medium was replaced once more at 24 h. For isotope tracing with [methyl- $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ -L-Met during mESC differentiation, R1 cells were cultured for two passages in L-Met-free N2B27 medium supplemented with LIF, 2i and 0.2 mM of either [methyl- $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ -L-Met or natural L-Met. Differentiation of R1 cells in the presence of FGF-2 and ActA was as described<sup>26</sup>, with minor modifications. Briefly, mESCs were cultured in N2B27 medium containing 2i and 1,000 U/ml LIF for several passages and then seeded at  $2.2 \times 10^5$  cells/cm $^2$  in N2B27 medium containing 1% KnockOut Serum Replacement (Life Technologies), 12 ng/ml FGF-2 (PeproTech) and 20 ng/ml ActA (ORF Genetics) on plates coated with a thin layer of Geltrex extracellular matrix preparation (Life Technologies). The medium was exchanged after 24 h.

**Knockdown experiments in mESCs.** shRNA expression vectors targeting Tet1 and Tet2 were generated by cloning synthetic oligonucleotides in pLKO.1 (ref. 52). Recombinant lentiviruses were produced by cotransfecting pLKO.1 shRNA expression vectors and packaging plasmids in HEK-293 cells. 48 h after transduction in the presence of 8  $\mu\text{g}/\text{ml}$  polybrene, shRNA-expressing mESCs were selected with 4  $\mu\text{g}/\text{ml}$  puromycin. Cell pools were continuously cultured in the presence of puromycin. shRNA target sequences were as follows: SCR (control), 5'-CCT AAG GTT AAG TCG CCC TCG-3' (ref. 52); Tet1, 5'-TGT AGA CCA TCA CTG TTC GAC-3' (see ref. 52), Tet2: 5'-TTC GGA GGA GAA GGG TCA TAA-3'. esiRNAs for Smug1 knockdown were generated as described<sup>53</sup>. The cDNA template for *in vitro* transcription was generated by PCR using following primers: forward, 5'- CGT AAT ACG ACT CAC TAT AGG GAG CCC GTG GGT G-3', and reverse, 5'-CGT AAT ACG ACT CAC TAT AGG GGT TTC GTC CAC TGG G-3'. R1 mESCs were weaned from 2i for two passages in FBS- and LIF-containing medium as described above. Upon plating the second passage, the cells were transfected in a p60 plate with 6  $\mu\text{g}$  of Smug1 esiRNAs (34.5 nM) and 20  $\mu\text{l}$  of Lipofectamine RNAi MAX (Life Technologies) according to the manufacturer's instructions and were lysed 72 h after transfection.

**Culture and transfection procedures for HEK-293T cells.** All transfections were performed using jetPRIME transfection reagent (PEQLAB Biotechnologie GmbH) according to the manufacturer's instructions. HEK-293T cells were seeded 24 h before transfection at a density of  $2.5 \times 10^6$  cells per 75 cm $^2$  flask and incubated in 10 ml of medium. The transfection solution (500  $\mu\text{l}$  of jetPRIME buffer, a specific amount of plasmid DNA (Supplementary Note 4) and 20  $\mu\text{l}$  of jetPRIME reagent) was added to the medium, and the cells were incubated for 48 h, with an additional medium exchange 24 h after transfection. When cotransfection of esiRNA was performed, a second transfection step (500  $\mu\text{l}$  of jetPRIME buffer, 5  $\mu\text{g}$  of esiRNA and 20  $\mu\text{l}$  of jetPRIME reagent) was carried out 4 h after transfection of plasmid DNA. esiRNAs were purchased from Sigma (human TDG esiRNA EHU038971; human SMUG1 esiRNA EHU098861; human CDK5RAP1 esiRNA EHU079221). **Supplementary Note 4** summarizes the overexpression and knockdown procedures.

**Isotope tracing with [ $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ thymidine or [ $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ -L-Met in HEK-293T cells transfected with Tet1cd.** 24 h before transfection,  $2.5 \times 10^6$  cells were seeded in a 75-cm $^2$  flask containing 10 ml either of (for [ $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ thymidine) DMEM medium supplemented with 50  $\mu\text{M}$  [ $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ T or (for [ $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ -L-Met) DMEM medium lacking L-Met, L-cystine and pyruvate, which was supplemented with 10% dialyzed FBS, 2 mM [ $^{13}\text{C}$ , $^{15}\text{N}$ ] $_{2}$ -L-Met and 0.2 mM L-cystine. Transfection was performed as described above using labeled medium.

**Real-time PCR analysis of mRNA expression.** For analysis of Tet1 and Tet2 knockdown in mESC total RNA was prepared with Trizol (Invitrogen), cDNA synthesis was performed with Quantitect reverse transcription kit from Qiagen, and real-time PCR was performed with the Power Sybr Green PCR master mix from Applied Biosystems on an Applied Biosystems 7500 Fast sys-

tem. Knockdown efficiencies relative to control samples transfected with SCR esiRNAs were 79% and 70% for Tet1 and Tet2, respectively. The primers used to estimate them are listed in **Supplementary Note 4**. For analysis of Smug1 knockdown in mESCs and EpiLC differentiation samples, total RNA was prepared with RNeasy spin columns (Qiagen), followed by DNase treatment using TURBO DNA-free (Ambion, Life Technologies); cDNA synthesis was carried out using iScript cDNA Synthesis kit (Biorad); real-time PCR was performed with SsoFast EvaGreen Supermix (Biorad). Smug1 knockdown efficiency relative to control samples transfected with esiRNAs targeting GFP was estimated to be 60%. Quantification of Tet, Dnmt, Tdg and Smug1 transcripts during EpiLC differentiation and Smug1 knockdown samples was performed using the primers listed in **Supplementary Note 4**. Expression levels were quantified with respect to the housekeeping gene *Gapdh* and normalized to time point 0 h.

**Tet *in vitro* assay.** A plasmid was prepared from *dam* $^-$ /*dcm* $^-$  competent *E. coli* strain (New England BioLabs) and methylated with M.SssI (New England BioLabs). 1.5  $\mu\text{g}$  of plasmid DNA were then treated with recombinant Tet1 from the 5hmC TAB-Seq Kit (Wisegene) corresponding to ref. 54. After 3 h incubation at 37  $^{\circ}\text{C}$  and proteinase K treatment, the oxidized plasmids were purified with GeneJET PCR Purification Kit from Thermo Scientific and eluted in 25  $\mu\text{l}$  water. Samples were then subjected to LC/MS/MS analysis as described<sup>42</sup>. The results are compiled in **Supplementary Note 3**.

**Correlation and cluster analyses.** Statistical data analysis was performed using IBM SPSS Statistics 19. Results of bivariate correlation analyses are summarized in **Supplementary Note 5**. Unsupervised clustering of species with respect to its correlation coefficients was applied by average linkage hierarchical clustering using a squared Euclidean distance measure.

**Pulldown assay.** For the pulldown assay, 250  $\mu\text{g}$  (50  $\mu\text{l}$ ) of the crude nuclear protein extracts were filled up to 500  $\mu\text{l}$  with 50 mM TEAB and 1 mM MgCl $_2$ . The binding conditions were 45 mM TEAB, 1.1 mM MgCl $_2$ , 2 mM HEPES, 42 mM NaCl and 20  $\mu\text{M}$  EDTA. Complete Protease Inhibitor Cocktail Tablets were used from Roche Diagnostics (Indianapolis, IN, USA). The DNA oligomers with DNA-protein cross-linker were dissolved in neat DMSO, and 1 nmol was added to the protein lysate and incubated for 20 min at room temperature. Streptavidin-coated magnetic particles (Roche Diagnostics, Indianapolis, IN, USA) were washed three times with binding buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 7.4) before 200  $\mu\text{l}$  of the bead slurry (equal to 2 mg beads) were added to the sample. Following 2 h incubation at room temperature under constant rotating, the beads were washed three times with 50 mM TEAB and 1 mM MgCl $_2$ . The beads were reconstituted in 50 mM TEAB and 1 mM MgCl $_2$ . Disulfide bonds of the cross linker were cleaved and alkylated in the process of enzymatic digestion, and the magnetic particles were removed before adding trypsin (described below).

**Protein sample preparation.** Cell lysis of mouse embryonic stem cells was performed as described in ref. 29. For each lysis, approximately  $7.5 \times 10^7$  cells were used. Protein concentration was determined by Bradford assay. For each experiment 250  $\mu\text{g}$  (50  $\mu\text{l}$ ) of the crude nuclear protein extract were used. Protein samples for MS analysis were reduced by adding 100 mM TCEP and by incubating on a shaker at 650 r.p.m. for 1 h at 60  $^{\circ}\text{C}$  and subsequently alkylated with 200 mM iodoacetamide in the dark for 30 min at 25  $^{\circ}\text{C}$ . Protein samples were digested with 0.5  $\mu\text{g}$  trypsin (Promega, Madison, MA, USA) for 16 h at 37  $^{\circ}\text{C}$ . The reaction was stopped using 1 mM phenylmethylsulfonylfluoride. After tryptic digestion, peptide labeling with the TMT2plex reagents (Thermo Fisher Scientific, Waltham, MA, USA) was performed according to the manufacturer's instructions. TMT2plex reagents 126 and 127 were used to label the samples. When the sample (proteins enriched with hmU- or hmC-containing DNA strands) was labeled with TMT126, the control sample (proteins enriched with no modified DNA-bases) was labeled with the TMT127 reagent and vice versa. Subsequent to the labeling, both sample and control, were combined. This way, each experiment was performed twice as a so-called label swap experiment. Organic solvent was removed by vacuum centrifugation, and the sample was finally reconstituted in 1% (v/v) formic acid for MS analysis.

**LC/MS analysis of protein samples.** The samples were analyzed using an UltiMate 3000 nano liquid chromatography system (Dionex, Fisher Scientific, Waltham, MA, USA) coupled to an LTQ-Orbitrap XL (Thermo Fisher

Scientific, Waltham, MA, USA). Of each eluate, 15  $\mu$ l were injected for the analysis. The samples were desalted and concentrated on a  $\mu$ -precolumn cartridge (PepMap100, C18, 5  $\mu$ M, 100  $\text{\AA}$ , size 300  $\mu$ m i.d. x 5 mm) and further processed on a custom-made analytical column (ReproSil-Pur, C18, 3  $\mu$ M, 120  $\text{\AA}$ , packed into a 75  $\mu$ m i.d. x 150 mm and 8  $\mu$ m picotip emitter). A 57-min multistep analytical separation was performed at a flow rate of 300 nl/min. In the first 50 min, a linear gradient was ramped up from 5% (v/v) solvent B (acetonitrile containing 0.1% formic acid and 5% DMSO) and 95% solvent A (water containing 0.1% formic acid and 5% DMSO) to 95% solvent B. This level was held for 5 min and then ramped down again to 5% solvent B within 2 min. Mass spectrometric analyses were performed starting with a full mass scan in the mass range between  $m/z$  300 and  $m/z$  1,650. This survey scan was followed by three MS/MS scans using the FTMS mass analyzer and high normalized collision energy of 70 in the HCD cell and three additional scans using the ion trap mass analyzer and a normalized collision energy of 35.

**Protein identification and relative quantification method.** The Thermo Proteome Discoverer 1.1 software (Thermo Fisher Scientific, Waltham, MA, USA) was used for protein identification and for relative quantification. The Sequest (Thermo Fisher Scientific, Waltham, MA, USA) search engine was used in combination with a Uniprot database (*Mus musculus*; date of download, 04/2013). As a limit of detection, a ratio of threefold signal over the noise filter was applied. A maximum of two missed cleavage sites was allowed. The mass tolerances were 10 p.p.m. for the precursor mass and 0.5 Da for the fragment ion mass. Carbamidocysteine was set as static modification. Dynamic modifications were: cation, Na (D, E); the residue of the DNA-protein crosslinker (+146.028 Da; K, Y); Oxidation (M) as well as TMT2plex (N-term. and K). Identified, nonredundant peptides, which were labeled with the TMT2 reagent, were used for relative quantification. The integration window tolerance was 20 p.p.m., and the integration method was set to 'most confident centroid'. The signals of the TMT2 reporter ions 126 and 127 were used to calculate ratios and monitor either preferred or nonpreferred binding of the identified proteins to the modified DNA bases in comparison to the control strand. From the

identified proteins, the only proteins considered as 'specific readers' that were enriched in both, the forward and the reverse experiment.

42. Schiesser, S. *et al.* Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
43. Gierlich, J. *et al.* Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **8**, 3639–3642 (2006).
44. Cao, H. & Wang, Y. Collisionally activated dissociation of protonated 2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives. *J. Am. Soc. Mass Spectrom.* **17**, 1335–1341 (2006).
45. Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).
46. Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
47. Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
48. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90**, 8424–8428 (1993).
49. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
50. Montanez, E. *et al.* Kindlin-2 controls bidirectional signaling of integrins. *Genes Dev.* **22**, 1325–1330 (2008).
51. Ying, Q.L. & Smith, A.G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
52. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
53. Kittler, R., Heninger, A.K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779–784 (2005).
54. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).

## XII. DNA methylation in the context of oxidized cytidine derivatives

### Introduction:

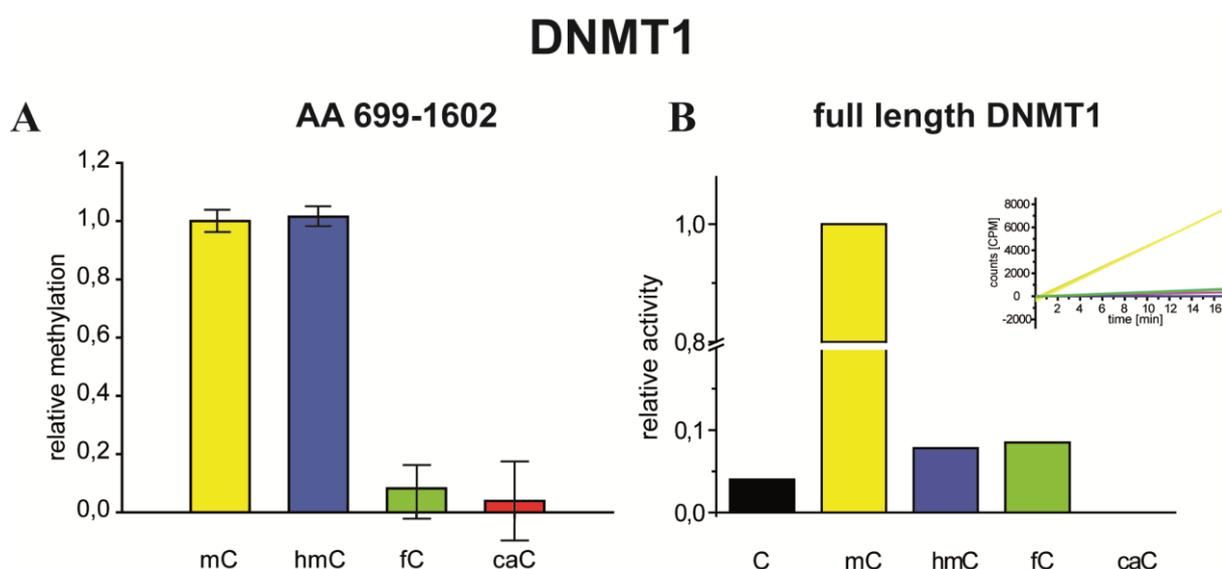
The inheritance of methylation marks is a major process in epigenetics. As described in Chapter I.2.2.1 it is carried out by DNMT1 in a conserved mechanism on hemi-methylated CpG loci after DNA replication.<sup>[1]</sup> The correct replication of a methylation mark is very complex and is tightly regulated. One of the most important factors is UHRF1, which acts as a molecular chaperone and activator for DNMT1.<sup>[2]</sup> As DNA methylation is involved in many biological processes such as transcription and cellular development, aberrant methylation patterns can have pathological consequences for a cell (see Chapter I.2.2).<sup>[3]</sup>

The recently discovered oxidized cytidine derivatives are supposed to be intermediates in an active demethylation process, however it is suggested that these bases might also be involved in epigenetic processes such as transcription regulation(see Chapter I.2.3).<sup>[4]</sup> Therefore, it might be necessary for a cell to maintain certain patterns of modifications in the CpG loci. There are 15 distinct patterns resulting from a DNMT/TET enzymatic cascade, who might alter biological processes or the epigenetic status of a distinct loci.<sup>[5]</sup> Investigations whether this cascade can occur on hemi-modified DNA or these modification are lost by dilution after cell replication are necessary.

While the potential of DNMT1 to methylate opposite of a hemi-hydroxymethylated CpG was already investigated *in vitro* and suggested to be rather low, an *in vivo* study showed significant methylation activity.<sup>[6-7]</sup> Additionally, the binding of UHRF1 to hmC containing DNA using its SRA domain was shown by EMSA studies and proteomic approaches, however this issue remains controversial as another study stated the opposite.<sup>[7-9]</sup> In the latter study FACS was also used to determine the dilution of hmC after DNA replication in the S-phase in mESCs.<sup>[7]</sup> A significant reduction of hmC levels was observed and a model was suggested that connects hmC predominantly with *de novo* methylation. A model for replication dependent dilution of hmC in PGCs was shown as well as the dilution of fC and caC in embryonic development.<sup>[10-11]</sup> Unlike hmC, fC and caC are suggested to be intermediates in active DNA demethylation. However as the exact biological functions of these new epigenetic players are yet not fully understood, it might be necessary for a cell to copy a formylated or carboxylated mark under certain circumstances.

## Results and discussion:

To investigate the methyltransferase activity in the context of the newly discovered cytidines, different oligonucleotides containing the new cytidine derivatives hmC, fC and caC were incubated with the three DNA methyltransferases DNMT1, DNMT3a and DNMT3b in presence of radiolabeled SAM using a multiwell assay.<sup>[12]</sup> The resulting kinetics (**Fig. 1**) showed a strong preference of DNMT1 for hemi-methylated DNA and only 1/10 of its activity towards hemi-hydroxymethylated and-formylated DNA. Unmethylated DNA showed an even slower activity and hemi-carboxylated DNA did not act as a substrate. This result is consistent with earlier studies and supports the idea that demethylation by dilution after DNA replication is enhanced by oxidation of mC. However DNMT1 is a tightly regulated enzyme and establishing the methylation system *in vitro* is very challenging. Especially interactions of several domains in DNMT1s N-terminal region can influence substrate binding and activity as described in Chapter I.2.3.<sup>[1]</sup>



**Figure 1: Methylation activity of truncated (A) and kinetics of full length (B) DNMT1 in different modification contexts. All datapoints are presented in table 1 and table 3 A) Quantification of the change of mC in different contexts by LC-MS/MS. Each context was measured in triplica and the mean values were taken, SD is illustrated in black bars. Relative methylation of 1.0 represents full conversion of the single CpG to mCpG. This was observed in the mC and hmC context while no significant change in the fC and caC context could be detected B) Relative activity of full length DNMT1 in different contexts determined by radiolabeling using <sup>3</sup>H-SAM. Plotting radio counts and time resulted in a kinetic plot of the enzyme. The calculated slope of the hemi-methylated substrate was set as 1 and the slopes of the other modifications set relative to it. A clear preference for the mC context was detected, while a CpG in the caC context did not act as substrate.**

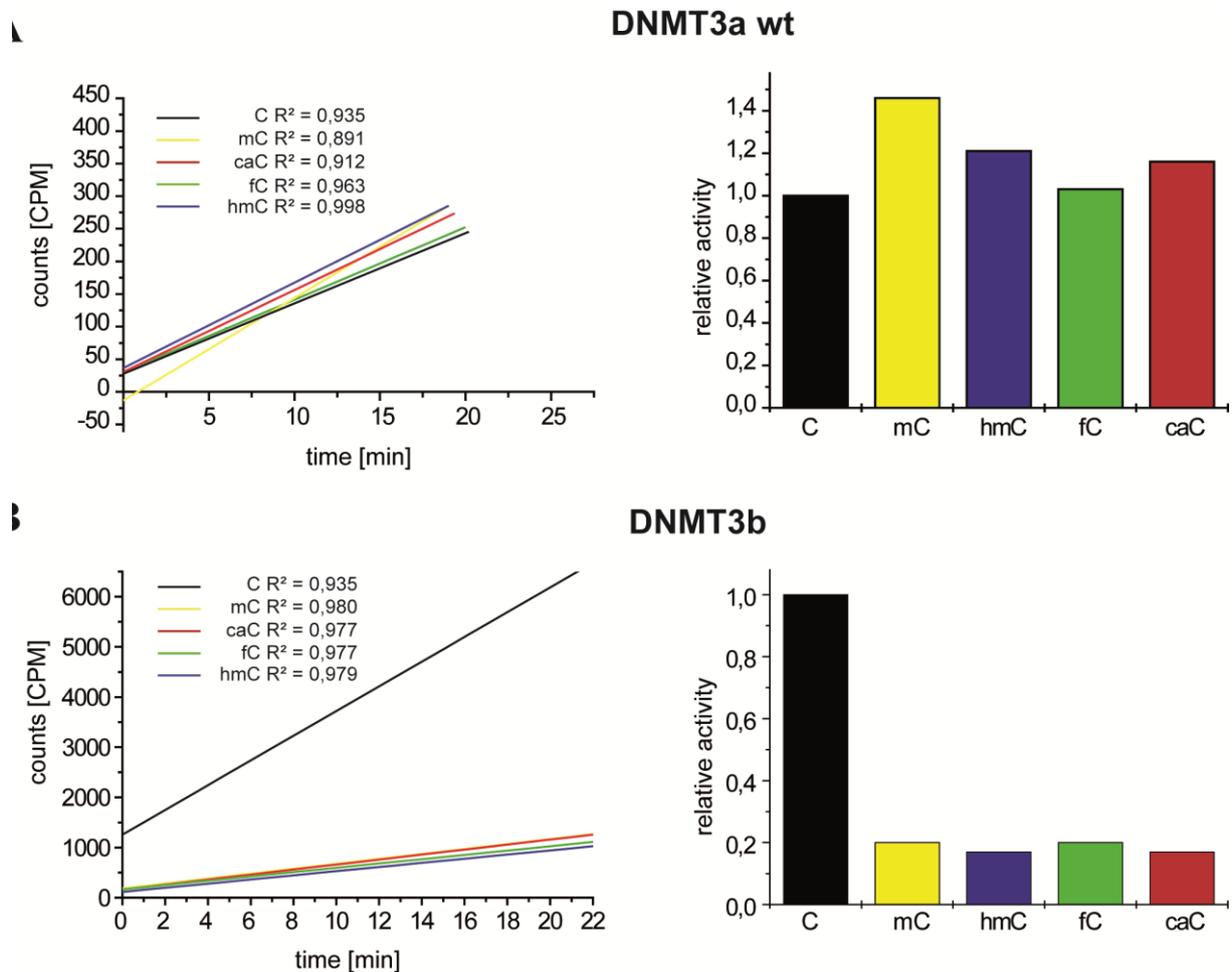
To address this issue a truncated version of DNMT1 lacking the N-terminal region was expressed by *Caterina Brandmayr* and a methylation assay was performed to determine the activity in the context of the newly discovered cytidine derivatives.<sup>[13]</sup> Surprisingly the truncated enzyme was capable to convert 100% of the available CpG loci into mCpG in an hmC context (**Fig. 1**).

In contrast, in an fC or caC context only about 10% of conversion was observed, even though the kinetics of the full length protein in fC and hmC context are similar. It is therefore possible, that a switch for the recognition of hemi-hydroxymethylated CpGs exists, as the catalytic site is in principle capable of establishing mC in an hmC context. Interestingly UHRF2, a homolog of UHRF1 which is supposed to be responsible for such a switch in the mC context, was identified as a specific reader for hmC, which repels mC, and found to enhance hmC levels *in vivo*.<sup>[9]</sup> In contrast to the hmC context, the CpG loci in the fC and caC context remained in an unmethylated state, which supports their role as intermediates in DNA demethylation. This observation might result from the different chemical properties of fC and caC that exclude binding to DNMT1.

Even though the low activity of the full length DNMT1 on hmC containing CpG loci could be confirmed, this study revealed that in principle hmC can fit unlike fC or caC into the active site of the enzyme and contribute to methylation. By using a regulatory switch for DNMT1 followed by oxidation by one of the TET enzymes, a fully hydroxymethylated CpG locus could be maintained. Therefore further investigations have to be done to gain insights into the role of DNMT1 in the maintenance of methylation/hydroxymethylation patterns.

In the kinetic study DNMT3a did not show a preference for any modification type (**Fig. 2A**), On the other hand DNMT3b showed a strong preference for unmethylated DNA and about 0.2 times the activity on the other types of CpGs (**Fig. 2B**). The results for unmodified, hemi-methylated and -hydroxymethylated CpG loci are consistent with previous reports. These reports suggested a role of DNMT3a/b in maintenance of hydroxymethylation.<sup>[6]</sup> This study could extend the role of DNMT3a to fC and caC CpG sites and offers an enzymatic route to the 15 distinct patterns of CpG modifications by a DNMT3a/TET cascade. These enzymes have already been linked and the distinct patterns could be connected to different biological functions.<sup>[5]</sup> However this study was only performed *in vitro* and can of course not mimic the complex processes of the DNA methylation/demethylation machinery. Additionally, oxidation of mC is suggested to take place predominantly at hemi-methylated CpG loci, therefore the biological relevance of the distinct patterns is yet not known.<sup>[14]</sup> Very recently *Ji et al.* were able to show similar effects of the new epigenetic bases on DNMT1 and DNMT3a methylation

using LC-MS/MS on ODN containing two repetitive CpG sites.<sup>[15]</sup> Interestingly the modification status of one CpG site has great influence on the methylation of the second site. Further studies have to be performed using more complex *in vitro* and *in vivo* approaches.



**Figure 2: Relative activity of DNMT3a (A) and DNMT3b (B) in different contexts determined by radiolabeling using <sup>3</sup>H-SAM. Plotting radio counts and time resulted in a kinetic plot of the enzyme. The calculated slope of the unmethylated substrate was set as 1 and the slopes of the other modifications set relative to it. Datapoints were collected in duplica and the mean value was used to calculate the slope. A) No preference of DNMT3a is visible. B) A clear preference for unmethylated DNA of DNMT3b was detected.**

## Materials and Methods:

### *Enzymatic digestion of DNA samples*

DNA samples isolated after *in vitro* experiments were resuspended in 25  $\mu\text{L}$   $\text{H}_2\text{O}$ . The first digestion step was carried out in 100  $\mu\text{M}$   $\text{ZnSO}_4$  with 42 units Nuclease S1 (*Aspergillus oryzae*, Roche) and 5 units antarctic phosphatase (*New England Biolabs*) and incubated at 37 °C for 3 h. The second digestion step was performed by addition of  $[\text{Na}]_2\text{-EDTA}$  (final concentration 100  $\mu\text{M}$ ) and 0.2 units snake venom phosphodiesterase I (*Crotalus adamanteus venom*, UBS Corporation) and incubation at 37 °C for 3 h.

### *LC-ESI-MS/MS:*

Acetonitrile (LC-MS grade) was purchased from *Carl Roth*; formic acid (p.a. grade for MS) was purchased from *Fluka*; water was purified by a Milli-Q Plus system (*Merck Millipore*); general HPLC and source settings for LC-ESI-MS/MS were as previously described.<sup>[16]</sup>

### *Methylation assay using truncated DNMT1*

Oligonucleotides (ODN7 – ODN11) bearing a single CpG site (modified or unmodified) were annealed by heating to 95 °C for 4 min followed by cooling to 4 °C over a period of 45 min. Methylation assay was performed together with *Caterina Brandmayr* in 1 x Dnmt1 Reaction Buffer (*New England Biolabs*), with double stranded DNA (1  $\mu\text{M}$ ), purified truncated mDnmt1 (1  $\mu\text{M}$ , *Caterina Brandmayr*), Bovine Serum Albumin (0.1 mg/mL) and with or without SAM (160  $\mu\text{M}$ ). All experiments were performed in triplicate in a total reaction volume was 20  $\mu\text{L}$ . Samples were incubated at 37 °C for 3 h, followed by extraction using Roti-Phenol/ $\text{CHCl}_3$ /isoamylalcohol (*Roth*) and by desalting using ZipTip (*Millipore*). Purified oligonucleotides were subsequently digested to single nucleosides and analyzed by LC-MS/MS. Mass spectrometry analysis was performed using a method established by *Toni Pfaffeneder* as described.<sup>[16]</sup>

### *DNA-Synthesis:*

DNA-Synthesis was performed as described by *Schröder et al.* or purchased from metabion and verified by MALDI-MS and HPLC.<sup>[17]</sup>

*Data of methylation assay:*

XpG:CpG	$\Delta$ mCpG (%)	St.Dev. (%)
mC	104,52	5,50
hmC	106,07	5,43
fC	8,55	10,72
caC	4,08	14,58

**Table 1: Change of mC in percent upon incubation with truncated mDnmt1 (with or without SAM) per modified CpG site (X = mC, hmC fC or caC). All experiments were performed in triplicate.**

*Oligonucleotides used:*

ODN	Sequence	source
1	5'-Bio-TTGCACTCTCCTCCCGGAAGTCCCAGCTTC-3'	metabion
2	5'-GAAGCTGGGACTTCCGGGAGGAGAGTGCAA-3'	metabion
3	5'-GAAGCTGGGACTTcmCGGGAGGAGAGTGCAA-3'	metabion
4	5'-GAAGCTGGGACTTChmCGGGAGGAGAGTGCAA-3'	Arne Schröder
5	5'-GAAGCTGGGACTTcfCGGGAGGAGAGTGCAA-3'	Arne Schröder
6	5'-GAAGCTGGGACTTccaCGGGAGGAGAGTGCAA-3'	Arne Schröder
7	5'-CCTAGCGCATTAC-3'	metabion
8	5'-GTAATGmCGCTAGG-3'	metabion
9	5'-GTAATGhmCGCTAGG-3'	Dr. Martin Muenzel
10	5'-GTAATGfCGCTAGG-3'	Dr. Martin Muenzel
11	5'-GTAATGcaCGCTAGG-3'	Dr. Martin Muenzel

**Table 2: Oligonucleotides used for methylation assays and kinetics, XpG in red.**

*Methylation kinetics:*

Methylation kinetics were performed as described by *Dr. Renata Jurkowska and colleagues*.<sup>[12]</sup> Briefly, methylation reactions using DNMT1 (0.2  $\mu$ M) were carried out with DNA substrate (2  $\mu$ M), [methyl-<sup>3</sup>H]AdoMet (1.0  $\mu$ M; PerkinElmer) at 37.8 °C in methylation buffer (20 mM HEPES, pH 7.5), EDTA (1 mM), KCl (50 mM)). Methylation reactions using DNMT3b/a (2.5  $\mu$ M) were carried out with DNA substrate (1  $\mu$ M), [methyl-<sup>3</sup>H]AdoMet (1.0  $\mu$ M; PerkinElmer) at 37.8 °C in methylation buffer (20 mM HEPES, pH 7.5), EDTA (1 mM), KCl (50 mM)). DNA substrates for the assay were prepared by annealing an ODN1 with a complementary oligonucleotide ODN2-6, to form unmethylated and hemimethylated substrates. Initial rates

were determined by linear regression. Methylation assays were calibrated by complete methylation of the substrates with M.SssI (CpG methyltransferase; *NEB*).

*Data of methylation kinetics:*

DNMT1:

Time [min]	C	mC	hmC	caC	fC
2	54	637	53	14	78
5	128	1920	149	15	192
10	224	4162	304	19	393
15	296	6762	539	13	591
Slope	19	470	37	0	40
Rel. activity	1.0	25.4	2.0	0.0	2.1

**Table 3: Methylation kinetics measurement results for DNMT1 in radio counts/minute (CPM) for different methylation contexts. Unmethylated DNA corresponds to C-context, hemi-methylated DNA to mC-context, hemi-hydroxymethylated to hmC-context, hemi-formylated DNA to fC-context, hemi-carboxylated DNA to caC.**

DNMT3a:

Time [min]	C-1	C-2	C Average			mC-1	mC-2	mC Average		
3	79	78	79			78	nd	78		
6	106	89	97			76	82	79		
12	129	100	115			113	124	118		
18	234	221	228			226	243	234		
24	260	341	300			485	342	414		
Slope	10.8					15.7				
Rel. activity	1.00					1.45				

Time [min]	hmC-1	hmC-2	hmC Average		caC-1	caC-2	caC Average		fC-1	fC-2	fC Average
3	78	81	80		106	80	93		57	62	60
6	94	125	110		94	83	88,167			117	117
12	218	173	195		154	155	154,5		141	127	134
18	278	266	272		277	275	276,1665		240	240	240
24	nd	nd	nd		nd	nd	nd		298	301	300
Slope	13.1					12.6					
Rel. activity	1.21					1.17					

**Table 4: Methylation kinetics measurement results for DNMT3a in radio counts/minute (CPM) for different methylation contexts. Unmethylated DNA corresponds to C-context, hemi-methylated DNA to mC-context, hemi-hydroxymethylated to hmC-context, hemi-formylated DNA to fC-context, hemi-carboxylated DNA to caC.**

**DNMT3b:**

Time [min]	C-1	C-2	C Average			mC-1	mC-2	mC Average		
3	1250	1185	1217			218	232	225		
6	2348	2144	2246			398	381	390		
12	4212	3685	3949			663	632	648		
18	5877	4949	5413			894	1122	1008		
24	5636	5913	5775			nd	1119	1119		
Slope	246.3					49.2				
Rel. activity	1.00					0.20				

Time [min]	hmC-1	hmC-2	hmC Average	caC-1	caC-2	caC Average	fC-1	fC-2	fC Average
3	180	211	251	160	163	161	200	204	202
6	383	367	337	260	288	274	354	402	378
12	552	564	509	655	523	589	614	603	608
18	816	790	852	673	782	728	1124	915	1019
24	1185	898	1042	978	932	955	1020	1164	1092
Slope	43.0			40.8			49.7		
Rel. activity	0.17			0.17			0.20		

**Table 6: Methylation kinetics measurement results for DNMT3b in radio counts/minute (CPM) for different methylation contexts. Unmethylated DNA corresponds to C-context, hemi-methylated DNA to mC-context, hemi-hydroxymethylated to hmC-context, hemi-formylated DNA to fC-context, hemi-carboxylated DNA to caC.**

## Literature

- [1] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *Chembiochem* **2011**, *12*, 206-222.  
*Structure and Function of Mammalian DNA Methyltransferases*
- [2] J. Sharif, *et al.*, *Nature* **2007**, *450*, 908-912.  
*The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA*
- [3] Z. M. Svedruzic, *Curr. Med. Chem.* **2008**, *15*, 92-106.  
*Mammalian Cytosine DNA Methyltransferase Dnmt1: Enzymatic Mechanism, Novel Mechanism-Based Inhibitors, and RNA-directed DNA Methylation*
- [4] M. Iurlaro, *et al.*, *Genome Biol.* **2013**, *14*, R119.  
*A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation*
- [5] H. Wu, Y. Zhang, *Cell* **2014**, *156*, 45-68.  
*Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions*
- [6] H. Hashimoto, *et al.*, *Nucleic Acids Res.* **2012**, 10.1093/nar/gks155.  
*Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation*
- [7] J. Otani, *et al.*, *PLoS One* **2013**, *8*, e82961.  
*Cell Cycle-Dependent Turnover of 5-Hydroxymethyl Cytosine in Mouse Embryonic Stem Cells*
- [8] C. Frauer, *et al.*, *PLoS One* **2011**, *6*, e21306.  
*Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain*
- [9] Cornelia G. Spruijt, *et al.*, *Cell* **2013**, *152*, 1146-1159.  
*Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives*
- [10] A. Inoue, L. Shen, Q. Dai, C. He, Y. Zhang, *Cell Res.* **2011**, *21*, 1670-1676.  
*Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development*
- [11] J. A. Hackett, *et al.*, *Science* **2013**, *339*, 448-452.  
*Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine*
- [12] R. Jurkowska, A. Ceccaldi, Y. Zhang, P. Arimondo, A. Jeltsch, in *Epigenetics Protocols*, Vol. 791 (Ed.: T. O. Tollefsbol), Humana Press, **2011**, pp. 157-177.
- [13] C. Brandmayr, doctoral thesis, LMU Munich **2013**.
- [14] M. Yu, *et al.*, *Cell* **2012**, *149*, 1368-1380.  
*Base-Resolution Analysis of 5-Hydroxymethylcytosine in the Mammalian Genome*
- [15] D. Ji, K. Lin, J. Song, Y. Wang, *Molecular BioSystems* **2014**, 10.1039/C4MB00150H.  
*Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation*
- [16] S. Schiesser, *et al.*, *JACS* **2013**, *135*, 14593-14599.  
*Deamination, Oxidation, and C-C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine*
- [17] A. S. Schröder, *et al.*, *Angewandte Chemie International Edition* **2014**, *53*, 315-318.  
*Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2'-Deoxycytidine*

## **XIII. Dynamics of pyrimidines during methyltransferase treatment**

### **Introduction:**

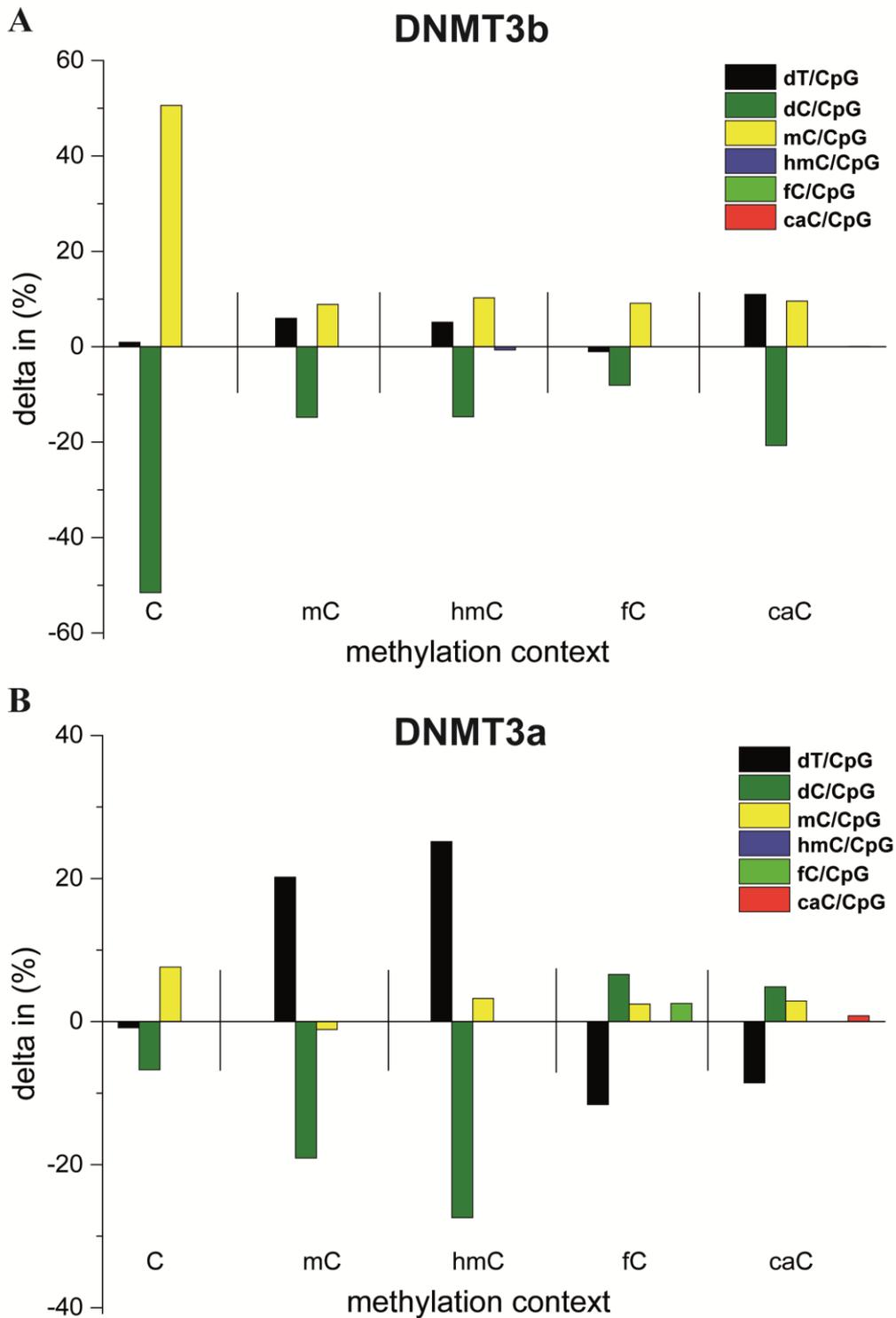
As described in Chapter I.2.3 deamination of mC to dT is speculated to be a pathway in active demethylation. DNMT3a/3b are supposed to be involved in this pathway, as upon saturation of the C5-C6 double bond in the pyrimidine ring a nucleophilic attack is possible by for example a surrounding water molecule. The resulting in deamination is possible before methylation resulting in dU, or after methylation resulting in dT. Next to this conversion of cytidines into uridines, DNMTs have been reported to act as demethylases by C-C bond cleavage and also a mechanism has been proposed and investigated chemically. The dynamics of pyrimidines is therefore a combination of deamination and C-C bond cleavage, which can in principle both be catalysed by DNMTs.

### **Results and discussion:**

For a first investigation of the dynamics of pyrimidine derivatives, the oligonucleotides that were used for the kinetics of DNMT3s were incubated in the presence and absence of SAM by *Dr. Renata Jurkowska and colleagues*. The re-isolated DNA was then analysed using LC-MS/MS and dynamics of pyrimidines evaluated by calculating the change of each modification per CpG (**Fig. 1**).

The transformation of C to mC catalysed by DNMT3b in C-context correlates with the kinetic study, as the amount of C is decreasing and mC is rising by 50% per CpG (**Fig. 1a**). This observation corresponds to methylation and is weaker for the other substrates, which only show methylation of about 10% per CpG. Additionally, the amounts of the other pyrimidine-derivatives did not show significant changes. The dynamics for pyrimidines incubated with DNMT3a were controversial (**Fig. 1b**). While the methylation activity of the enzyme was very low, a loss of C of about 20% per CpG and a gain of dT was observed in roughly the same amount. This could refer to deamination and could be observed in mC-context (hemi-methylated DNA) and hmC-context (hemi-hydroxymethylated DNA). Again no change in the other pyrimidine derivatives could be observed.

Data from chapter XII using truncated DNMT1 was also evaluated in respective to deamination but no indications could be found (not shown).



**Fig.1: Dynamics of pyrimidines in different CpG modification contexts after incubation with DNMT3b (A) or DNMT3a (B) as detected by single measurements via LC-MS/MS in the presence of SAM compared to the absence of SAM. A) A gain in mC with a combined decrease of C displays in the C context the preference of DNMT3b for unmethylated DNA as reported in chapter XII. All other pyrimidines in the other contexts did not show dynamic changes. B) No changes of mC, hmC, fC and caC could be detected in any context. In the mC and hmC context a loss C and a gain of T could be detected after incubation of DNMT3a in the presence of SAM.**

These results therefore do not support the idea of C-C bond cleavage by DNMTs, but suggest a deamination activity of DNMT3a under the given conditions. However the absolute changes are in the range of the error of the applied method and only single samples were measured.

In principle two sides can be deaminated in hemi-methylated DNA: the template strand already harbouring an mC, or the counterstrand after successful methylation using SAM. In hemi-hydroxymethylation, deamination from mC to dT can only occur after methylation. To validate the suggestion of deamination an incubation under the same conditions using D<sub>3</sub>-SAM was performed to check if deamination occurs after successful methylation, which would result in D<sub>3</sub>-dT. However this was not the case using DNMT3a and DNMT3b in unmethylated or hemimethylated DNA (data not shown). To further investigate the dynamics and to exclude the possibility of deamination in the strand already harbouring mC, DNA containing dU instead of dT was synthesized, as dU did not show up in any of the earlier experiments and deamination of mC to dT would be easier detectable.

Incubation of these oligonucleotides did not show any deamination activity of DNMT3a (data not shown). As the methylation activity was with a maximum of 1.3% per CpG also very low, it is possible that the enzyme was not active at all and therefore the experiment has to be repeated in the future.

## **Materials and Methods:**

### *Enzymatic digestion of DNA samples*

DNA samples isolated after *in vitro* experiments were resuspended in 25  $\mu\text{L}$   $\text{H}_2\text{O}$ . The first digestion step was carried out in 100  $\mu\text{M}$   $\text{ZnSO}_4$  with 42 units Nuclease S1 (*Aspergillus oryzae*, Roche) and 5 units antarctic phosphatase (*New England Biolabs*) and incubated at 37 °C for 3 h. The second digestion step was performed by addition of  $[\text{Na}]_2\text{-EDTA}$  (final concentration 100  $\mu\text{M}$ ) and 0.2 units snake venom phosphodiesterase I (*Crotalus adamanteus venom*, UBS Corporation) and incubation at 37 °C for 3 h.

### *LC-ESI-MS/MS:*

Acetonitrile (LC-MS grade) was purchased from *Carl Roth*; formic acid (p.a. grade for MS) was purchased from *Fluka*; water was purified by a Milli-Q Plus system (*Merck Millipore*); general HPLC and source settings for LC-ESI-MS/MS were as previously described.<sup>[1]</sup>

### *In vitro reactions:*

*In vitro* reactions were performed as described by *Dr. Renata Jurkowska*.<sup>[2]</sup> Briefly, methylation reactions were carried out with DNA substrate (1 $\mu\text{M}$ ), 320 $\mu\text{M}$  SAM or D<sub>3</sub>-SAM and recombinant DNMTs (2.5  $\mu\text{M}$ ) at 37.8 °C in methylation buffer (20 mM HEPES, pH 7.5), EDTA (1 mM), KCl (5.0 mM)) for 1 h. DNA substrates for the assay were prepared by annealing an ODN containing one unmethylated CpG site with a complementary oligonucleotide, to form unmethylated and hemimethylated substrates. After incubation the samples were frozen at -20 °C and then purified by spin columns after addition of 1 x volume isopropanol (GeneJET, *Thermo Fisher*), followed by enzymatic digestion as described above.

*Data of methylation reactions:*

	$\Delta T/CpG$ island	$\Delta C/CpG$ island	$\Delta mC/CpG$ island	$\Delta hmC/CpG$ island	$\Delta fC/CpG$ island	$\Delta caC/CpG$ island
C DNMT3b	0,94%	-51,57%	50,62%	0,00%	0,01%	0,00%
mC DNMT3b	5,95%	-14,80%	8,88%	0,00%	-0,01%	-0,02%
hmC DNMT3b	5,11%	-14,71%	10,25%	-0,67%	0,01%	0,00%
fC DNMT3b	-1,05%	-8,07%	9,10%	0,03%	0,00%	-0,01%
caC DNMT3b	10,99%	-20,72%	9,58%	0,05%	0,03%	0,07%
C DNMT3a	-1%	-6,75%	7,63%	0,00%	0,01%	-0,01%
mC DNMT3a	20%	-19,09%	-1,11%	0,00%	0,01%	-0,01%
hmC DNMT3a	25%	-27,41%	3,24%	-1,03%	0,02%	0,00%
fC DNMT3a	-12%	6,59%	2,44%	0,02%	2,54%	0,00%
caC DNMT3a	-9%	4,85%	2,87%	0,04%	0,00%	0,81%

**Table 1: Change of pyrimidines in (%) per CpG in different methylation contexts after incubation using DNMT3a or DNMT3b.**

*DNA-Synthesis:*

DNA-Synthesis was performed as described by *Schröder et al.* or purchased from *metabion* and verified by MALDI-MS and HPLC.<sup>[3]</sup> The phosphoramidite for dU was purchased from *Glen Research*.

*ODNs used:*

ODN	Sequence	Source
1	5'-Bio-TTGACTCTCCTCCCGGAAGTCCCAGCTTC-3'	Metabion
2	5'-GAAGCTGGGACTTCCGGGAGGAGAGTGCAA-3'	Metabion
3	5'-GAAGCTGGGACTTcmCGGGAGGAGAGTGCAA-3'	Metabion
4	5'-GAAGCTGGGACTTChmCGGGAGGAGAGTGCAA-3'	Arne Schröder
5	5'-GAAGCTGGGACTTcfCGGGAGGAGAGTGCAA-3'	Arne Schröder
6	5'-GAAGCTGGGACTTccaCGGGAGGAGAGTGCAA-3'	Arne Schröder
7	5'-Bio-UUGCACUCUCCUCCCGGAAGUCCCAGCUUC-3'	Benjamin Hackner
8	5'-Bio-UUGCACUCUCCUCCmCGGAAGUCCCAGCUUC-3'	Benjamin Hackner
9	5'-GAAGCUGGGACUUCCGGGAGGAGAGUGCAA-3'	Benjamin Hackner
10	5'-GAAGCUGGGACUUCmCGGGAGGAGAGUGCAA-3'	Benjamin Hackner
11	5'-Bio-TTGACTCTCCTCCcaCGGAAGTCCCAGCTTC-3'	Benjamin Hackner

**Table 2: Oligonucleotides used for dynamics of pyrimidines, XpG sites marked in red.**

## Literature

- [1] S. Schiesser, *et al.*, *JACS* **2013**, *135*, 14593-14599.  
*Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine*
- [2] P. Bashtrykov, *et al.*, *Chembiochem* **2014**, 10.1002/cbic.201300740, n/a-n/a.  
*Targeted Mutagenesis Results in an Activation of DNA Methyltransferase 1 and Confirms an Autoinhibitory Role of its RFTS Domain*
- [3] A. S. Schröder, *et al.*, *Angewandte Chemie International Edition* **2014**, *53*, 315-318.  
*Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2'-Deoxycytidine*

## XIV. Abbreviations

AA	Amino acid
AdoMet	S-Adenosylmethionine
AP-site	Apurinic/Apyrimidinic site
ARP	Aldehyde reactive probe
BER	Base excision repair
BS	Bisulphite sequeence
caC	5-carboxylcytidine
CGI	Cytidine-Guanosine dinucleotide island
CpG	Cytidine-Guanosine dinucleotide
dA	2'desoxy adenine
dC	2'desoxy cytidine
dG	2'desoxy guanidine
DNA	Desoxy-ribonucleic acid
DNMT	DNA 5-Methyltransferase
dsDNA	Double stranded DNA
dT	2'desoxythymidine
dU	2'desoxyuridine
EDTA	Ethylenediaminetetraacetaic acid
EMSA	Electromobility shift assay
ESI	Electrion spray ionization
FACS	Fluorescence assisted cell sorting
fC	5-formylcytidine
GK repeats	Glycine-Lysine repeat
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonic acid
hmC	5-hydroxymethylcytidine
HPLC	High performance liquid chromatography
IDCase	Isoorotate decarboxylase
iPSC	Induced pluripotent stem cell
IUPAC	International Union of Pure and Applied Chemistry
LC-MS	Liquid chromatography- mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
MALDI	Matrix associated LASER desorption ionization
mC	5-methylcytidine
mESC	Mouse embryonic stem cell
mRNA	Messenger ribonucleic acid
ODN	Oligonucleotide
OG	2-oxogluterate
oligo	Oligonucleotide
PCR	Polymerase chain reaction
PGC	Primordial germ cells
PWWP-domain	proline-tryptophane motif
QQQ	Triple quadropole
SAM	S-adenosylmethionine
ssDNA	Single stranded DNA
TDG	Thymidine DNA glycosylase

TET  
UV

Ten eleven translocation  
Ultra-Violet