Dissertation zur Erlangung des Doktorgrades

der Fakultät Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Synthese und Analyse von natürlichen DNA-Modifikationen zur Aufklärung des epigenetischen DNA-Metabolismus

Toni Michael Pfaffeneder

aus

Gräfelfing (Deutschland)

2014

"Das Leben jedes Menschen ist ein Weg zu sich selber hin, der Versuch eines Weges, die Andeutung eines Pfades. Kein Mensch ist jemals ganz und gar er selbst gewesen; jeder strebt dennoch, es zu werden, einer dumpf, einer lichter, jeder wie er kann. Jeder trägt Reste von seiner Geburt, Schleim und Eischalen einer Urwelt, bis zum Ende mit sich hin. Mancher wird niemals Mensch, bleibt Frosch, bleibt Eidechse, bleibt Ameise. Mancher ist oben Mensch und unten Fisch. Aber jeder ist ein Wurf der Natur nach dem Menschen hin. Uns allen sind die Herkünfte gemeinsam, die Mütter, wir alle kommen aus demselben Schlunde; aber jeder strebt, ein Versuch und Wurf aus den Tiefen, seinem eigenen Ziele zu. Wir können einander verstehen; aber deuten kann jeder nur sich selbst."

Hermann Hesse, Demian, die Geschichte einer Jugend (1919).

Danksagung

An dieser Stelle möchte ich einigen Menschen danken, die zum Gelingen dieser Arbeit beigetragen haben.

An vorderster Stelle danke ich meinem Doktorvater *Prof. Dr. Thomas Carell* für das herausfordernde, hochspannende Thema, die exzellenten Arbeitsbedingungen und seine stete Unterstützung.

Herrn *PD Dr. Stylianos Michalakis* danke ich für die Übernahme des Koreferates, sowie allen weiteren Mitgliedern der Prüfungskommission für ihre Mitwirkung an der mündlichen Prüfung.

Dr. Markus Müller, *Michael Ehrlich*, *Dr. Benjamin Hackner, Dr. Stefan Schiesser* und *Dr. Johannes Harder* danke ich für ihr engagiertes Korrekturlesen meiner Doktorarbeit.

Frau *Slava Gärtner* danke ich für ihre Hilfe bei allen organisatorischen und bürokratischen Angelegenheiten. Frau *Sabine Voß* und Frau *Kerstin Kurz* danke ich für alle chemischen und labortechnischen Belange. Besonders freute mich die Mitwirkung von Frau *Kerstin Kurz* im Massenspektrometrie-Team. Ich danke dir für schier endlose "DNA-Verdaus". Niemand pipettiert so genau wie du!

Herrn *Dr. David Stephenson* und *Frau Claudia Dubler* danke ich für die Aufnahme meiner NMR-Spektren, auch wenn es manchmal zu wenig Substanz war. Bei Herrn *Dr. Werner Spahl* und Frau *Sonia Kosak* möchte ich mich für die gute Zusammenarbeit in der Massenspektrometrie bedanken.

Dem Fonds der chemischen Industrie danke ich für ein Doktorandenstipendium.

Dr. Benjamin Hackner gebührt mein besonderer, herzlicher Dank. Er war für mich in den letzten Jahren der beste Kollege und Freund in vielen Lebenslagen. Vieles wäre ohne dich nicht möglich gewesen oder weniger schön. Merci!

Meinem Laborplatznachbarn *Michael "Mischa" Ehrlich* danke ich für unser kommunistisches Prinzip – was mein ist, ist auch dein – und natürlich für die vielen lustigen und anregenden Momente. An dieser Stelle möchte ich auch meinem Labor F4.004 für die gute Stimmung und den steten Zusammenhalt danken; gerade bei so monotonen Arbeiten wie dem wöchentlichen Spitzenstecken. Insbesondere danke ich *Jessica Steinbacher*, *Olesea Kosmatchev*, *René*

Publikationsliste

Teile dieser Dissertation wurden bereits publiziert oder auf Konferenzen vorgestellt.

Publikationen während der Doktorarbeit:

[17] T. Pfaffeneder[#], F. Spada[#], M. Wagner[#], C. Brandmayr, S. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574–581. *Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.*

[16] D. B. Bucher, B. M. Pilles, T. Pfaffeneder, T. Carell, W. Zinth, *ChemPhysChem* **2014**, *15*, 420–423. *Fingerprinting DNA oxidation processes: IR characterization of the 5-Methyl-2'-deoxycytidine cation.*

[15] A. S. Schröder, J. Steinbacher, B. Steigenberger, F. A. Gnerlich, S. Schiesser, T. Pfaffeneder, T. Carell, *Angew. Chem. Int. Ed.* **2013**, *53*, 315–318. *Synthesis of a DNA Promoter Segment Containing all four Epigenetic Nucleosides 5-methyl-, 5-hydroxymethyl-, 5-formyl- and 5-carboxy-dC.*

[14] S. Schiesser[#], T. Pfaffeneder[#], K. Sadeghian, B. Hackner, B. Steigenberger, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599. *Deamination, Oxidation and C–C bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxycytosine.*

[13] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P.W.T.C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146–1159. *Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives.*

[12] B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *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.*

[11] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516–6520. *Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing.*

[10] T. F. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Münzel, M. Müller, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schuller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, *131*, 1577–1590. *Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors.*

[9] M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chem. Eur. J.* **2011**, *17*, 13782–13788. *Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine.*

[8]   <u>T. Pfaffeneder</u>[#], B. Hackner[#], M. Truss[#], M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012. *The discovery of 5-formylcytosine in embryonic stem cell DNA.*

[#] geteilte Erstautorenschaft.

Publikationen während des Studiums:

[7]   J. Cordes, F. Calo, K. Anderson, <u>T. Pfaffeneder</u>, S. Laclef, A. J. White, A. G. Barrett, *J. Org. Chem.* **2012**, *77*, 652–657. *Total syntheses of angelicoin A, hericenone J, and hericenol A via migratory prenyl- and geranylation-aromatization sequences.*

[6]   K. Anderson, F. Calo, <u>T. Pfaffeneder</u>, A. J. White, A. G. Barrett, *Org. Lett.* **2011**, *13*, 5748–5750. *Biomimetic total synthesis of angelicoin A and B via a palladium-catalyzed decarboxylative prenylation-aromatization sequence.*

[5]   M. Seizl, L. Lariviere, <u>T. Pfaffeneder</u>, L. Wenzeck, P. Cramer, *Nucleic Acids Res.* **2011**, *39*, 6291–6304. *Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization.*

[4]   B. Weber, W. Bauer, <u>T. Pfaffeneder</u>, M. M. Dîrtu, A. D. Naik, A. Rotaru, Y. Garcia, *Eur. J. Inorg. Chem.* **2011**, *2011*, 3193–3206. *Influence of Hydrogen Bonding on the Hysteresis Width in Iron(II) Spin-Crossover Complexes.*

[3]   W. Bauer, <u>T. Pfaffeneder</u>, K. Achterhold, B. Weber, *Eur. J. Inorg. Chem.* **2011**, *2011*, 3183–3192. *Complete Two-Step Spin-Transition in a 1D Chain Iron(II) Complex with a 110-K Wide Intermediate Plateau.*

[2]   <u>T. M. Pfaffeneder</u>, S. Thallmair, W. Bauer, B. Weber, *New J. Chem.* **2011**, *35*, 691–700. *Complete and incomplete spin transitions in 1D chain iron(II) compounds.*

[1]   <u>T. Pfaffeneder</u>, W. Bauer, B. Weber, *Z. Anorg. Allg. Chem.* **2010**, *636*, 183–187. *X-ray Structure and Magnetic Properties of Two New Iron(II) 1D Coordination Polymers with Bis(imidazolyle)methane as Bridging Ligand.*

Konferenzbeiträge:

[4]   **Vortrag:** „*Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.*" 4th Nucleic Acids Conference, Cancún, Mexico, Dezember 2014.

[3]   **Posterpräsentation:** T. Pfaffeneder, B. Hackner, S. Schiesser, M. Müller, M. Truss, T. Carell, „*Discovery of 5-formylcytosine in stem cells and mechanistic insights into decarboxylation, deformylation and dehydroxymethylation.*" IMB Conference, Institute of Molecular Biology, Mainz, Oktober 2012.

[2]   **Vortrag:** „*5-Hydroxymethylcytosin, die sechste Base im Genom.*" VAAM-Fachgruppentreffen „Qualitätssicherung & Diagnostik", Köln, September 2012.

[1]   **Vortrag:** „*The chemistry of genome development.*" 3rd International Gene Center Symposium, Ludwig-Maximilians-Universität, München, Oktober 2011.

Inhaltsverzeichnis

Zusammenfassung

Die Information, wie Organismen aufgebaut sind, wird in der DNA-Sequenz durch einen Code aus vier Buchstaben von 2′-Desoxyadenosin (dA), 2′-Desoxyguanosin (dG), 2′-Desoxythymidin (dT) und 2′-Desoxycytidin (dC) gespeichert. Obwohl alle Zellen eines multizellulären Organismus genetisch gesehen identisch sind, erfüllen sie sehr unterschiedliche Funktionen (z.B. im Herzen oder im Gehirn). Aus diesem Grund müssen bestimmte Gene zu definierten Zeitpunkten während der zellulären Entwicklung aktiviert oder inaktiviert werden. Um Gene zu inaktivieren, kann die DNA kovalent durch DNA-Methyltransferasen (Dnmt) modifiziert werden, wobei 5-Methyl-2′-desoxycytidin gebildet wird (5mdC).[1] 2009 entdeckte man 5-Hydroxymethyl-2′-deoxycytidin (5hmdC) als weitere DNA-Modifikation, welche aus der Oxidation von 5mdC durch sogenannte Tet-Enzyme (*ten eleven translocation*) entsteht. 5hmdC wurde als ein lange gesuchtes Intermediat der aktiven DNA-Demethylierung angesehen, welche die Reaktivierung von Genen ermöglicht.[2-5] Die zentrale Frage, wie 5mdC/5hmdC auf molekularer Ebene aktiv aus der DNA entfernt werden könnte, wurde in dieser Dissertation adressiert.

Zu diesem Zweck wurden verschiedene LC-MS-basierende Analysenmethoden entwickelt, um die in **Schema 1** dargestellten, möglichen Demethylierungswege zu untersuchen. Ein denkbarer Mechanismus wäre die iterative Oxidation von 5mdC bis hin zu 5cadC durch die Tet-Enzyme. Dazu wurde die isolierte DNA von pluripotenten embryonalen Stammzellen der Maus (mES-Zellen) analysiert, da die Suche nach höheren Oxidationsprodukten in adulten (differenzierten) Geweben in einer früheren Studie ergebnislos war.[3] Das Oxidationsprodukt 5-Formyl-2′-desoxycytidin (5fdC) konnte in mES-Zellen per hochauflösender Massenspektrometrie in der Tat nachgewiesen werden und dessen Generierung ausgehend von 5mdC gezeigt werden. Die Identität von 5fdC konnte zweifelsfrei durch MS^3-Fragmentierung bewiesen werden und wurde mit einer synthetischen Referenzverbindung verifiziert. Die Reaktivität der Formylgruppe ausnutzend, wurde eine Derivatisierungsmethode mit einem Biotin-Hydroxylamin etabliert, die zu einer verbesserten Nachweisempfindlichkeit führte. Mit dieser Methode wurden abnehmende 5fdC-Level während der mES-Zelldifferenzierung detektiert, woraus ein globaler oxidativer DNA-Demethylierungsmechanismus abgeleitet wurde.[6] Währenddessen wurde 5-Carboxy-2′-desoxycytidin (5cadC), das Oxidationsendprodukt, in mES-Zellen durch zwei andere Forschungsgruppen entdeckt und die direkte Umwandlung von 5hmdC zu 5fdC und 5cadC durch die Tet-Enzyme, wie postuliert, gezeigt.[7-8] Durch die Entdeckung dieser Intermediate konnte in der Folge ein plausibler Demethylierungsweg über

die Basenexzisionsreparatur (BER) mittels der Thymin-DNA-Glykosylse (Tdg) etabliert werden.[7, 9]

In einer anschließenden Studie wurde ein verbesserter Phosphoramidit-Baustein von 5fdC synthetisiert und in Oligodesoxynukleotide (ODNs) per Festphasensynthese inkorporiert. Mit Hilfe dieser ODNs wurde *in vitro* gezeigt, dass 5fdC und 5cadC nur marginal mutagen sind, was eine Grundvoraussetzung der neu entdeckten Modifikationen ist, um als epigenetische Kontrolleinheiten zu fungieren.[10]
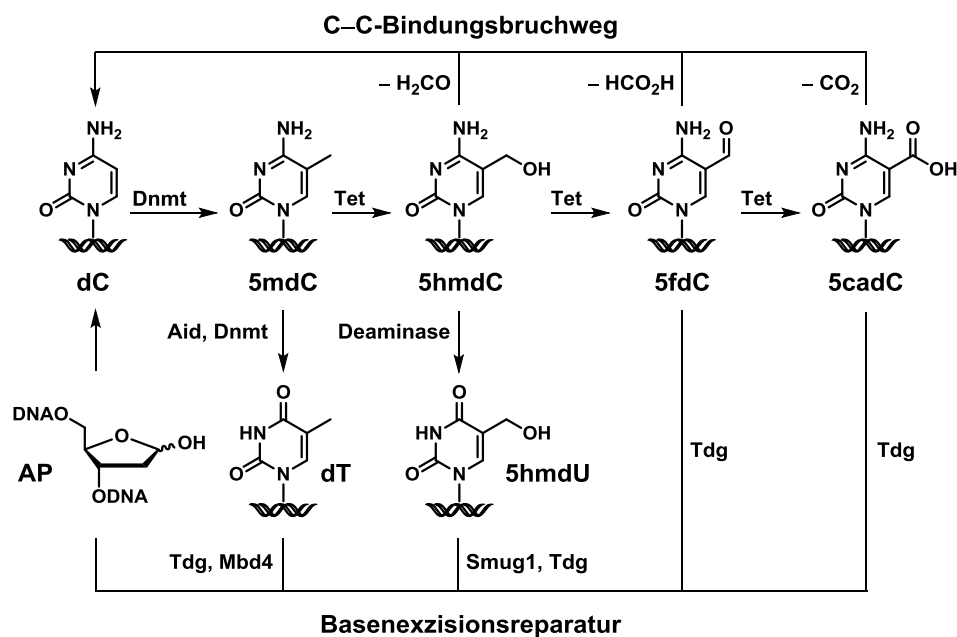
Schema 1 | Untersuchte Stoffwechselwege der aktiven DNA-Demethylierung.

Da wir an einem tieferen Einblick in die molekularen Mechanismen des epigenetischen DNA-Metabolismus interessiert waren, wurden empfindliche MS-basierte Methoden der akkuraten Quantifizierung und der Isotopenverfolgung aller möglicher Intermediate etabliert. Aus diesem Grund wurde zunächst eine modulare Syntheseroute von Isotopen-markierten Derivaten benötigt. **Schema 2A** gibt einen Überblick über die Moleküle, die als Standards für die Massenspektrometrie synthetisiert wurden. Die meisten Moleküle wurden ausgehend von einer Silyl-*Hilbert-Johnson*-Glykosylierung von [$^{15}N_2$]-Uracil erhalten. Die Schlüsselschritte waren Pd(0)-katalysierte Funktionalisierungen von 5-Iodo-2′-desoxy-[$^{15}N_2$]-cytidin (**Schema 2B**) oder 5-Iodo-2′-desoxyuridin (**Schema 2C**).

IV
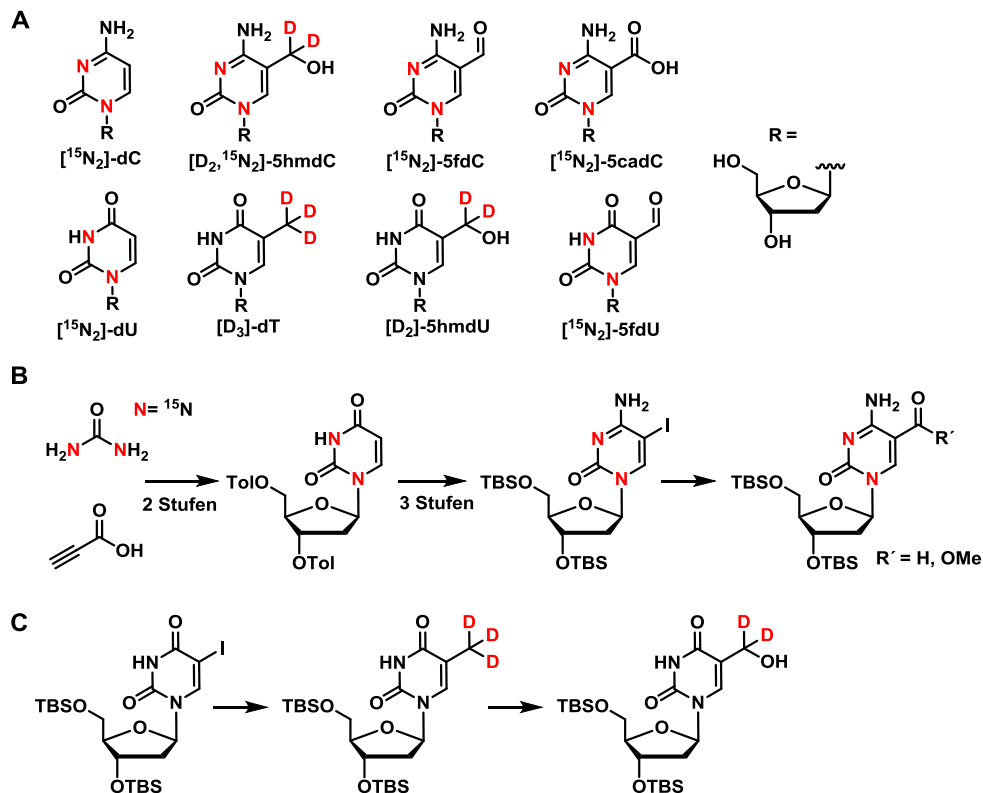
Schema 2 | Isotopen-markierte Verbindungen für die Massenspektrometrie.

Wir postulierten, dass die Decarboxylierung von 5cadC eine attraktive Alternative der DNA-Demethylierung darstellt (**Schema 1**), da sie die Generierung von abasischen Stellen (apyrimidinische Stellen; AP) vermeiden würde, die ein Problem für die genomische Integrität darstellen könnten. Die oben erwähnte modulare Synthese ausnutzend, wurde Isotopen-markiertes [^{15}N$_2$]-5cadC in ODNs inkorporiert. Die Inkubation dieser ODNs mit Zellkernextrakten von mES-Zellen und anschließende Analyse per hochauflösender Massenspektrometrie zeigte die Bildung von markiertem Cytidin ([^{15}N$_2$]-dC), was den ersten Beweis einer Decarboxylierungsaktivität lieferte.[11]

Um die chemischen und enzymatischen Reaktivitäten der neuen Cytidin-Modifikationen im Detail zu beleuchten, wurde eine umfassendere und sensitivere MS-basierte Methode (UHPLC-UV-ESI-MS/MS) mit einem Triple-Quadrupol-Massenspektrometer entwickelt. Diese Methode ermöglichte die akkurate und schnelle Quantifizierung des gesamten Produktspektrums im unteren fmol-Bereich innerhalb eines Analysenlaufs. Die exakte Quantifizierung der natürlichen Modifikationen bzw. der Reaktionsprodukte wurde mit den Isotopen-markierten internen Standards in **Schema 2A** durchgeführt.

Zuerst wurde die chemische Reaktivität der Cytidin-Derivate in chemischen Modellstudien verglichen. Es konnte gezeigt werden, dass 5hmdC im Vergleich zu 5mdC und 5fdC leichter zu oxidieren ist. Im Gegensatz dazu traten die Desaminierungsreaktionen der exozyklischen Aminogruppen unter physiologischen Bedingungen in einem geringfügigen Umfang auf. Die C–C-Bindungsbruchreaktionen (Dehydroxymethylierung von 5hmdC, Deformylierung von 5fdC und Decarboxylierung von 5cadC) benötigten die Anwesenheit hoher Konzentrationen von Thiolen als biologisch relevante (Organo)-Katalysatoren und waren Säure-katalysiert. Während 5hmdC sehr langsam reagierte, war die Umsetzung von 5fdC und insbesondere von 5cadC zu dC beträchtlich schneller. Die Reaktionen verlaufen vermutlich über ein kovalentes Enamin-Intermediat ab (**Schema 3**).[12] Thiole kommen als funktionelle Reste in aktiven Taschen von DNA-modifizierenden Enzymen, wie den Dnmts, vor. Die Identifizierung bzw. der Nachweis einer entsprechenden enzymatischen Aktivität steht *in vivo* allerdings noch aus.
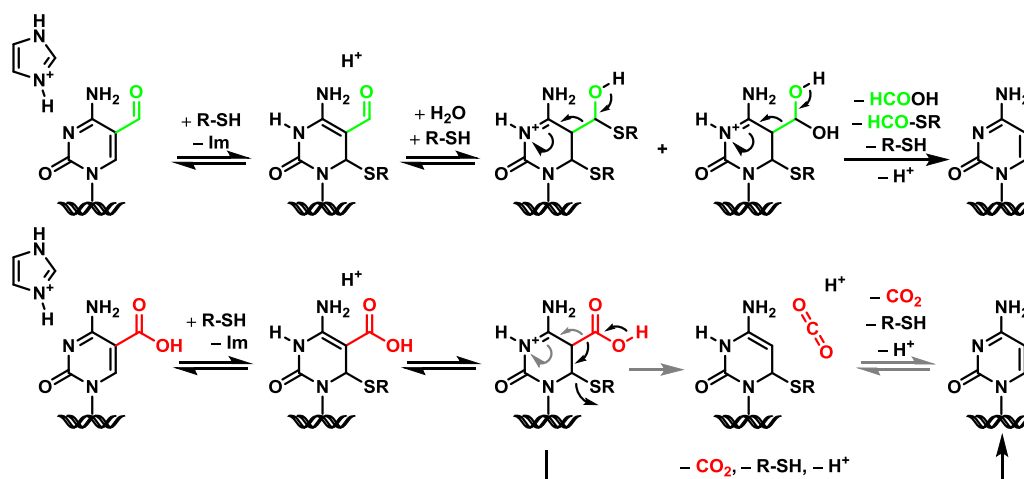
Schema 3 | Vorgeschlagener Mechanismus der Thiol-vermittelten und Säure-katalysierten C–C-Bindungsbruchreaktionen von 5fdC und 5cadC zu dC.

In einer weiteren Studie wurden enzymatische Oxidations- und Desaminierungsreaktionen in der DNA von mES-Zellen mit der UHPLC-UV-ESI-MS/MS-Methode untersucht. Zu diesem Zweck wurde eine Kombination aus quantitativer MS und Isotopenverfolgung durch metabolische Markierung angewendet. Es konnte gezeigt werden, dass 5-Hydroxymethyl-2′-desoxyuridin (5hmdU) in mES-Zellen aufgrund von Tet-vermittelter Oxidation von dT häufiger vorkommt als in differenziertem Gewebe (**Schema 4**). In mES-Zellen waren ca. 75% der globalen 5hmdU-Level auf eine Tet-vermittelte Generierung zurückzuführen, während die restlichen 25% von reaktiven Sauerstoffspezies (*reactive oxygen species*; ROS) herrührten. Die Desaminierung von 5hmdC zu 5hmdU, die als Teil eines alternativen DNA-

Demethylierungsmechanismus angesehen wurde (**Schema 1**), trug zu den Gleichgewichts-werten von 5hmdU nicht bei.[13]
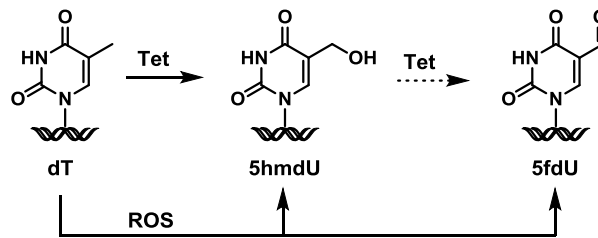
Schema 4 | In genomischer DNA wird dT neben der klassischen ROS-induzierten Oxidation auch durch Tet-Enzyme oxidiert.

Die globalen Level von 5-Formyl-2′-desoxyuridin (5fdU) in mES-Zellen und differenzierten Gewebezellen korrelierten hingegen lediglich mit oxidativer DNA-Schädigung. Durch ektopische Expression der katalytischen Domäne von Tet1 in HEK-293T-Zellen konnte gezeigt werden, dass Tet-Enzyme auch dieses höhere Oxidationsprodukt zu einem geringen Teil bilden können (**Schema 4**). Ein quantitativer Vergleich der 5mdC- und dT-Oxidation durch Volllängen-Tet1 ergab, dass Tet-Enzyme eine hohe Selektivität für 5mdC haben. Die N-terminalen, regulativen Domänen tragen ca. zwei Größenordnungen zur Selektivität der Enzyme bei. Interessanterweise war die Selektivität der Enzyme während der mES-Zelldifferenzierung zwischenzeitlich um eine Größenordnung reduziert.

Mit Hilfe der in dieser Arbeit entwickelten UHPLC-UV-ESI-MS/MS-Methode konnte ferner der Einfluss von zuvor identifizierten Modifikations-bindenden Proteinen auf die genomischen Level der modifizierten Bausteine untersucht werden. So wurde beispielsweise Uhrf2 als 5hmdC- und 5hmdU-Binder entdeckt. Die Analyse einer gleichzeitigen ektopischen Expression von Tet1 und Uhrf2 in HEK-293T-Zellen zeigte, dass Uhrf2 die Aktivität von Tet1 steigerte.[13-14]

Da 5fdC, 5cadC und die möglichen Desaminierungsprodukte von 5mdC und 5hmdC Substrate z.B. von Tdg sind und diese in Folge dessen durch die BER unter zeitweiliger Ausbildung von AP-Stellen repariert werden (**Schema 1**), wurde eine MS-basierte Methode für die Quantifizierung von AP-Stellen entwickelt. Hierfür wurde ein Hydroxylamin-Reagenz für die milde und schnelle Derivatisierung von AP-Stellen in der DNA sowie für die sensitive und selektive LC-MS/MS-Detektion im sub-fmol-Bereich entwickelt (**Schema 5**). Das Reagenz und die entwickelte UHPLC-MS/MS-Methode können in weiterführenden Studien für die Untersuchung der Demethylierungswege verwendet werden, die zur Bildung von AP-Stellen führen, z.B. bei der Stammzellerhaltung oder der Differenzierung.
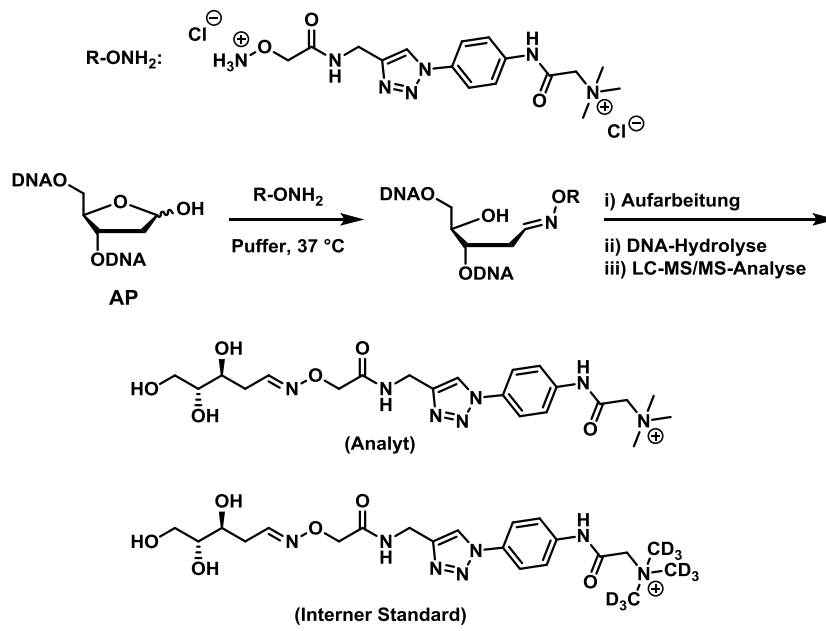
Schema 5 | Hydroxylamin-Reagenz für die sensitive Detektion und Quantifizierung von AP-Stellen per LC-MS/MS.

Summary

The information, of how living beings are constructed, is stored in the sequence of the four letter code of 2′-deoxyadenosine (dA), 2′-deoxyguanosine (dG), 2′-deoxythymidine (dT) and 2′-deoxycytidine (dC) that builds the DNA. Although all cells of a multicellular organism are genetically identical, their functions are divers (e.g. heart vs. brain). Therefore, distinct genes have to be switched on or off at certain time points during development. In order to inactivate genes, DNA can be covalently modified by DNA methyltransferases (Dnmt) to give 5-methyl-2′-deoxycytidine (5mdC).[1] In 2009, 5-hydroxymethyl-2′-deoxycytidine (5hmdC) was discovered as a further DNA modification, which results from oxidation of 5mdC by ten eleven translocation (Tet) enzymes. 5hmdC was considered to be the long searched for intermediate of active DNA demethylation enabling gene reactivation.[2-5] The central question of how 5mdC/5hmdC could be removed actively from DNA on a molecular basis, was addressed in this thesis.

For this purpose, several analytical techniques based on mass spectrometry were developed in this thesis, in order to dissect the putative pathways depicted in **Scheme 1**. A conceivable mechanism could involve iterative oxidation of 5mdC to 5cadC by the Tet enzymes. Owing to this, DNA isolated from pluripotent mouse embryonic stem cells (mESCs) was analysed, as higher oxidation products than 5hmdC could not been detected in adult (differentiated) tissue previously.[3] The oxidation product 5-formyl-2′-deoxycytidine (5fdC) was indeed detected in mESCs by high resolution mass spectrometry and its generation from 5mdC was shown. The identity of 5fdC in mESCs was unequivocally proven by MS^3-fragmentation and was verified by the synthetic reference compound. Exploiting the reactivity of the formyl group, a derivatization method for 5fdC with a biotin-hydroxylamine was established, which resulted in an improved detection sensitivity. With this method, declining 5fdC-levels could be observed upon mESC differentiation and a globally acting oxidative DNA demethylation mechanism was deduced.[6] Meanwhile, 5-carboxy-2′-deoxycytidine (5cadC), the final 5mdC oxidation product, was discovered in mESCs by two other groups and the direct conversion of 5hmdC to 5fdC and 5cadC was shown to be catalyzed by Tet enzymes, as postulated.[7-8] As a result of the discoveries of these intermediates a plausible DNA demethylation mechanism via the base excision repair (BER) by the thymine DNA glycosylase (Tdg) was established.[7, 9]

In a subsequent study an improved phosphoramidite building block for 5fdC was synthesized and incorporated into oligodeoxynucleotides (ODNs) by solid phase synthesis. Using these

ODNs, it was shown *in vitro* that 5fdC and 5cadC are only marginally mutagenic, which is a prerequisite for the newly identified modifications to function as epigenetic control units.[10]

C–C bond cleavage pathway
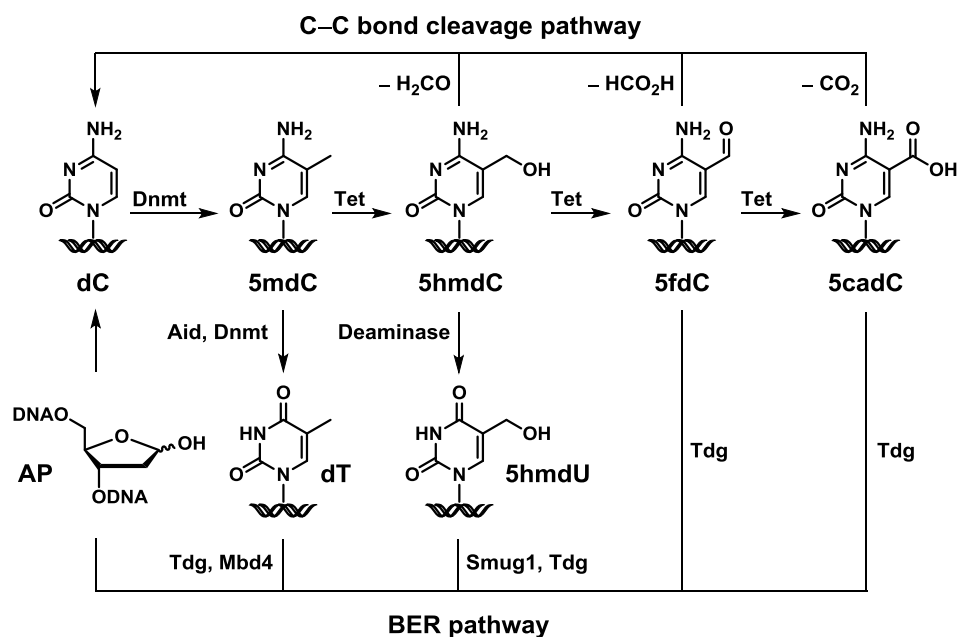
Scheme 1 | Investigated DNA demethylation mechanisms.

Since we were interested to get deeper insights into the molecular mechanisms of the epigenetic DNA metabolism, sensitive MS-based methods for accurate quantification and isotope tracing of all putative intermediates were established. For this, a modular synthesis route of isotopically labelled derivatives was required. An overview of the molecules, which were prepared as standards for mass spectrometry is given in **Scheme 2A**. Most molecules were obtained starting from silyl-*Hilbert-Johnson* glycosylation of [$^{15}N_2$]-uracil. The key steps were Pd(0) catalyzed functionalizations of 5-iodo-2′-deoxy-[$^{15}N_2$]-cytidine (**Scheme 2B**) or 5-iodo-2′-deoxyuridine (**Scheme 2C**).

We reasoned that decarboxylation of 5cadC could be an attractive alternative for DNA demethylation (**Scheme 1**), as it would avoid the production of abasic sites (apyrimidinic sites; AP) along with BER, which could pose a problem for genome integrity. Exploiting the abovementioned modular synthesis, isotopically labelled [$^{15}N_2$]-5cadC was incorporated into ODNs. Incubation of these ODNs with nuclear extracts from mESCs and analysis by high resolution mass spectrometry showed the formation of labelled cytidine ([$^{15}N_2$]-dC), which provided the first evidence for a decarboxylation activity.[11]
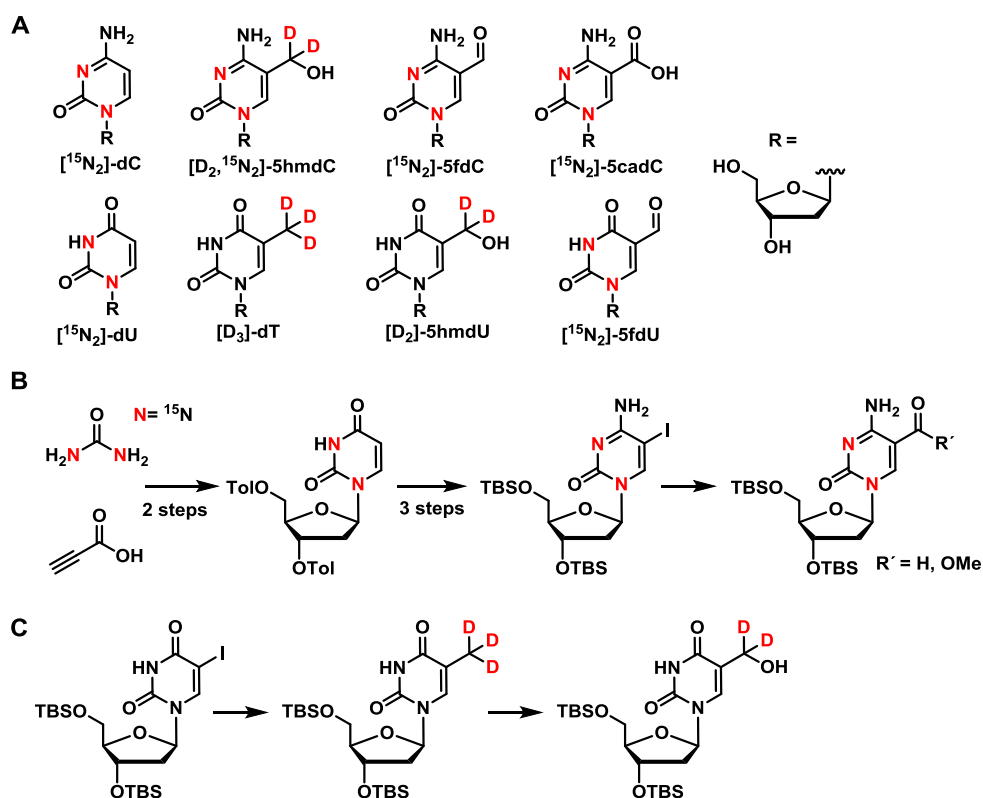
Scheme 2 | Isotopically labelled compounds for mass spectrometry.

In order to shed light on the chemical and enzymatic reactivity of the new cytidine modifications in detail, a comprehensive and more sensitive MS-based method (UHPLC-UV-ESI-MS/MS) using a triple quadrupole mass spectrometer was developed. The method allowed accurate and fast quantification of the whole product spectrum in the low fmol range in one single analysis. The exact quantification of the natural modifications or reaction products was conducted using the isotopically labelled internal standards in **Scheme 2A**.

Firstly, the chemical reactivity of the cytidine derivatives was compared in chemical model studies. It was revealed that 5hmdC is particularly prone to oxidation in comparison to 5mdC and 5fdC. In contrast, under physiologically conditions deamination reactions of the exocyclic amino group occurred only to a minor extent. C–C bond cleavage reactions (dehydroxy-methylation of 5hmdC, deformylation of 5fdC and decarboxylation of 5cadC) require the presence of high concentration of thiols as biologically relevant (organo)catalysts and are acid catalyzed. While 5hmdC reacted very slowly, 5fdC and especially 5cadC reacted considerably faster to dC. These reactions were postulated to proceed via a covalent enamine intermediate (**Scheme 3**).[12] Thiols are active site residues in DNA modifying enzymes such as the Dnmts. However, the identification of an enzymatic activity *in vivo* is still ongoing.
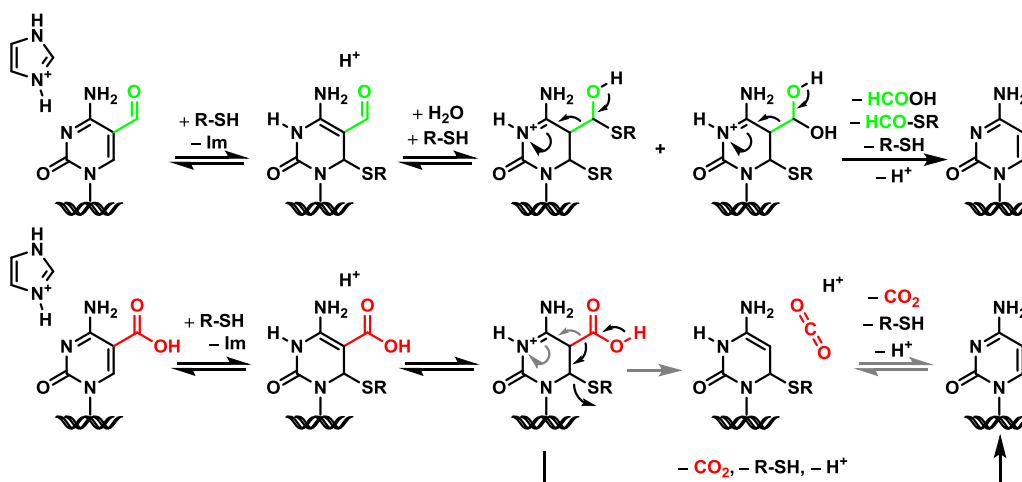
Scheme 3 | Proposed mechanisms of the thiol-mediated and acid-catalyzed C–C bond cleavage reactions of 5fdC and 5cadC to dC.

In a further study, enzymatic oxidation and deamination reactions in mESC DNA were investigated by the UHPLC-UV-ESI-MS/MS method. For this, a combination of quantitative MS and isotope tracing by metabolic labelling was performed. It could be shown, that 5-hydroxymethyl-2′-deoxyuridine (5hmdU) is more frequent in mESCs compared to differentiated tissue due to Tet-induced oxidation of dT (**Scheme 4**). In mESCs about 75% of the global 5hmdU levels were found to be produced by Tet and 25% by reactive oxygen species (ROS). Deamination of 5hmdC to 5hmdU, which was considered being part of an alternative DNA demethylation mechanism, did not contribute to the steady state level of 5hmdU.[13]
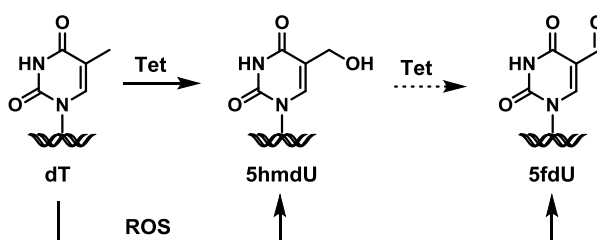
Scheme 4 | In genomic DNA dT is oxidized by Tet enzymes along with the classical ROS-induced oxidation.

The levels of the higher oxidation product 5-formyl-2′-deoxyuridine (5fdU) was found to correlate only with ROS-induced DNA-lesions in mESC and adult tissue. However, in HEK-293T-cells over-expressing the catalytic domain of Tet1, it could be shown that 5fdU is generated by the Tet enzyme to a less extent as well (**Scheme 4**). A comparison of the amount of 5mdC and dT oxidation events by full-length Tet1 revealed that Tet enzymes have a high selectivity for 5mdC. The N-terminal regulatory domains of Tet contribute about two orders

of magnitude to this selectivity. Interestingly, the selectivity of Tet enzymes is reduced by one order of magnitude during mESC differentiation.

Furthermore, with the help of the developed UHPLC-UV-ESI-MS/MS-method the influence of previously identified modified DNA-binding proteins on the genomic level was investigated. For instance, Uhrf2 was identified as a 5hmdC and 5hmdU binder. The analysis of HEK-293T-cells in which Tet1 and Uhrf2 were ectopically expressed, showed that Uhrf2 increased the activity of Tet1.[13-14]

Since 5fdC, 5cadC and the putative deamination products of 5mdC and 5hmdC are substrates of e.g. Tdg and are therefore possibly repaired by BER (**Scheme 1**), a MS-based method for the quantification of AP sites was developed. For this, a hydroxylamine reagent was designed for the mild and fast derivatization of AP sites in DNA as well as for the sensitive and selective LC-MS/MS-detection in the sub-fmol range (**Scheme 5**). The reagent and the developed UHPLC-MS/MS method can be used in further studies to dissect the pathways that lead to AP sites in genomic DNA, e.g. during stem cell maintenance or differentiation.
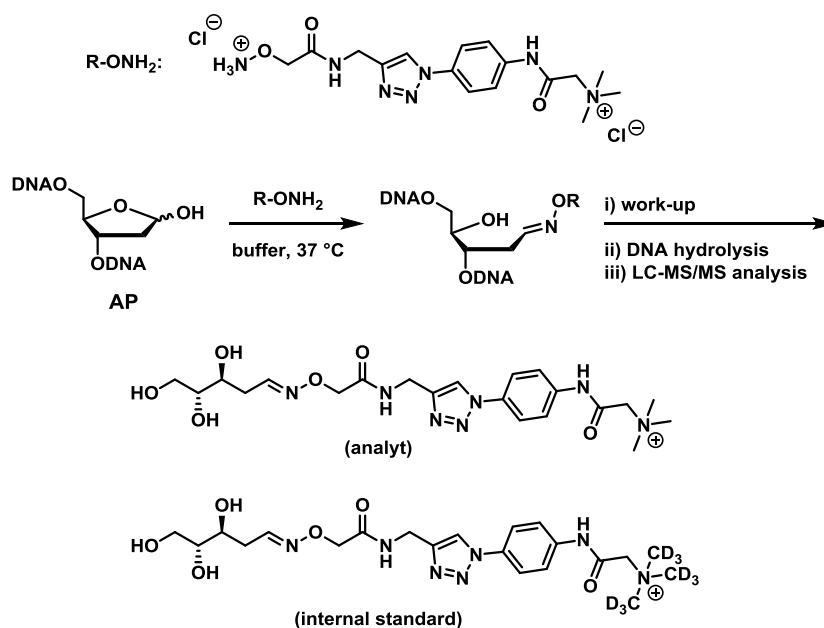
Scheme 5 | Hydroxylamine-reagent for sensitive AP site detection and quantification by LC-MS/MS.

1 Einleitung

1.1 Epigenetik

Der Bauplan eines jeden Lebewesens ist in der Desoxyribonukleinsäure (DNA) gespeichert und wird an die Nachkommen vererbt.[15-16] Alle Zellen eines multizellulären Organismus, wie dem Menschen, besitzen die gleiche genetische Information, unterscheiden sich aber in hohem Maße in Gestalt und Funktion. Um die Spezialisierung von Zellen innerhalb von Organstrukturen (z.B. im Gehirn, Herz, Leber, etc.) zu erreichen, müssen während der Entwicklung Mechanismen vorhanden sein, die die Regulation der Genexpression räumlich und zeitlich steuern. Nicht benötigte Gene müssen zu einer definierten Zeit deaktiviert und benötigte Gene aktiviert werden. Ferner müssen diese Genaktivitätsmuster an die entsprechenden Tochterzellen vererbt werden. Mit diesem Thema befasst sich das Forschungsgebiet der Epigenetik.[17] Die heute allgemeingültige, wissenschaftlich anerkannte Definition bezieht sich auf mitotisch oder meiotisch vererbbare Merkmale und Veränderungen, die die Genexpression beeinflussen, aber nicht auf die zugrundeliegenden DNA-Sequenzen zurückzuführen sind.[18-20] Chemisch gesehen wird dies in den Zellen durch kovalente Modifikationen der DNA und der DNA-assoziierten Proteine realisiert (siehe Abschnitt 1.2 und folgende). Die Gesamtheit dieser Mechanismen und Informationen – das Epigenom – beeinflussen sich in einem komplexen Netzwerk gegenseitig. Aberrante Prozesse und Veränderungen innerhalb dieses Systems spielen eine große Rolle bei der Entstehung von Krebs, Autoimmun- und neurodegenerativen Erkrankungen. Aus diesem Grund ist die Untersuchung des Epigenoms Gegenstand intensiver Forschung; ein detailliertes Verständnis wird in Zukunft neue Diagnostik- und Therapieansätze ermöglichen.[1, 21-24] Dabei stehen speziell embryonale Stammzellen (kurz ES-Zellen) im Fokus der Forschung, da diese Zellen über einzigartige Eigenschaften der Selbsterneuerung und Differenzierung in fast alle Zellarten verfügen. Das Epigenom unterliegt bei diesen Differenzierungsprozessen starken Veränderungen. Deshalb sind diese Zellen nicht nur für grundlegende Entwicklungsstudien interessant, sondern auch für die regenerative Medizin, wenn es um die Erneuerung krankhaften Gewebes geht.[25] Die Forschung an pluripotenten ES-Zellen stellt die Menschheit allerdings vor ein grundsätzliches, ethisches Problem, da deren Gewinnung mit der Tötung von Embryonen einhergeht. Deshalb ist in Deutschland die Gewinnung menschlicher ES-Zellen nach dem Embryonenschutzgesetz verboten. In der Grundlagen-forschung bedient man sich deshalb embryonaler Stammzellen aus Modellorganismen, wie

der Maus. Eine Lösung dieses Dilemmas würde die Reprogrammierung von körpereigenen Gewebezellen oder adulten Stammzellen in pluripotente Stammzellen liefern. Dies gelang erstmals *John B. Gurdon* und *Shinya Yamanaka*, die dafür 2012 mit dem Nobelpreis für Medizin ausgezeichnet wurden. Die Generierung von sogenannten induzierten pluripotenten Stammzellen (kurz iPS-Zellen) aus Gewebezellen gelang durch gentechnische Überexpression von vier Transkriptionsfaktoren, den *Yamanaka*-Reprogrammierungsfaktoren Oct4, Sox2, Klf4 und c-Myc, die Teil des Pluripotenznetzwerkes sind.[26-27] Die iPS-Zellen sind den ES-Zellen zwar ähnlich, aber sie lassen sich zum einen nicht uneingeschränkt differenzieren und zum anderen ist deren Gewinnung noch sehr ineffizient.[28] Aus diesem Grund besteht ein großer Bedarf, die zugrundeliegenden epigenetischen Reprogrammierungs-prozesse im Detail zu verstehen und die Generierung von iPS-Zellen zu optimieren.

1.2 Chromatin

Das Genom von Eukaryoten liegt in mehreren getrennten DNA-Molekülen, den Chromosomen, vor. Da das Genom je nach Organismus insgesamt eine Länge von mehreren Milliarden Basenpaaren haben kann (Genom der Maus: 2.7 Milliarden Basenpaare), wird die DNA im Zellkern mit Hilfe von basischen Proteinen (Histonen) zu höher geordneten Strukturen gepackt. Die Gesamtheit an DNA, Histonen und anderen DNA-bindenden Proteinen bildet das sogenannte Chromatin. Auf der ersten Ordnungsstufe werden je 147 Basenpaare um einen flachen zylinderförmigen Histonkomplex, bestehend aus je zwei Untereinheiten der Histone H2A, H2B, H3 und H4, gewunden. Diese bilden zusammen ein Nukleosom mit einem Durchmesser von 10 nm. Die Nukleosomen sind im Abstand von etwa 10–90 Basenpaaren zu einer Polynukleosomenkette angeordnet, die die räumlich zugäng-lichste und ausgedehnteste DNA-Struktur im Zellkern darstellt. Man findet diese Struktur in Bereichen, in denen Gene transkribiert werden (Euchromatin).[29] Mit Hilfe des Linker-Histons H1, wird die Polynukleosomenkette weiter zur sogenannten 30 nm Faser aufgewickelt, bei der die zylinderförmigen Nukleosomen zickzackförmig wie Münzen übereinander gestapelt werden.[29-30] Nichtaktive oder stillgelegte Chromatinbereiche liegen hauptsächlich in dieser kondensierten Form vor und können in mehreren Stufen noch dichter gepackt werden (Heterochromatin). Kondensierte Bereiche blockieren den räumlichen Zugang von RNA-Polymerasen oder anderen Faktoren, die für die Transkription von Genen nötig sind.[29]

Die Interaktion von Histonen mit der DNA ist in erster Linie elektrostatischer Natur. Histone weisen einen hohen Anteil basischer Aminosäuren auf, die unter physiologischen

Bedingungen einem Protonierungsgleichgewicht unterliegen und mit dem negativ geladenen Zucker-Phosphat-Rückgrat der DNA wechselwirken. Darüber hinaus besitzen Histone N- und C-terminale, längliche Fortsätze, die von dem Nukleosomenkern herausragen und u.a. für die Bildung der 30 nm Faser benötigt werden.[29] Diese sind Gegenstand diverser posttrans-lationaler Modifizierungen wie der Acetylierung, Methylierung, Phosphorylierung, Ubiquitin-ylierung, Sumoylierung, ADP-Ribosylierung und vieler weiterer. Deren Metabolismus und Funktion ist noch zu großen Teilen unverstanden.[31]

Durch diese chemischen Veränderungen wird nicht nur die Affinität der Histone zur DNA moduliert und damit die Kompaktheit des Chromatins. Die Modifikationen bilden auch einen sogenannten „Histon-Code", der die Chromatin-Funktion beeinflusst, indem Bindungsstellen für andere Proteine (z.B. Transkriptionsfaktoren) geschaffen oder entfernt werden.[32-35] Als Beispiel für Modifizierungen, die die Affinität der Histone zur DNA erhöhen, sei die gut verstandene Trimethylierung der Lysine K9 und K27 in Histon H3 genannt (H3K9me3, H3K27me3). Diese Modifikationen werden durch Histon-Methyltrans-ferasen (HMT) eingeführt und sind ein Signal für DNA-Kondensierung bzw. Gen-Inaktivier-ung. Eine prominente Ausnahme stellt das trimethylierte H3K4 dar, welches entgegen der Affinitätstheorie in transkriptionell aktiven Chromatin-Bereichen gefunden wird. H3K4me3 ist eine Bindestelle für Histon-Acetyltransferasen (HAT), die die Acetylierung der Lysine K9 und K14 in Histon H3 bewirkt (H3K9ac, H3K14ac).[29, 34] Mechanistisch versteht man die Acetylierung von Lysinen als eine "Neutralisierung" der vormals positiven Ladung von protonierten oder trimethylierten Lysinen, wodurch die Wechselwirkung der Histone mit dem negativen Zucker-Phosphat-Rückgrat der DNA geschwächt wird. Auf diese Weise nimmt das Chromatin eine losere Struktur ein, wodurch die Transkriptionsmaschinerie räumlichen gesehen einen besseren Zugang hat.[36-38]

Vergleicht man die Struktur des Chromatins von differenzierten Zellen (Gewebezellen) mit ES-Zellen, stellt man fest, dass das Chromatin in ES-Zellen insgesamt in einem weniger kondensierten Zustand vorliegt. ES-Zellen besitzen einen sehr großen Anteil lose gebundener Histone. Auf diese Weise haben Histon- und DNA-modifizierende Enzyme, sowie Regulationsfaktoren einen leichten Zugang zu Chromatin-Abschnitten, die bei der komplexen Zelldifferenzierung in schneller und definierter Weise epigenetisch markiert bzw. modifiziert werden müssen.[39] Histonmodifikationen unterliegen generell gesehen einem dynamischen Metabolismus, um den Aktivitätszustand von Genen bedarfsweise zu verändern.[40] Es sind neben den bereits angesprochenen modifizierenden Enzymen (z.B. HAT oder HMT) in der

Regel auch die entsprechenden demodifizierenden Enzyme bekannt (z.B. Histon-Deacetylasen (HDAC) oder -Demethylasen).[34, 39]

Einige Histonmodifikationen, die den Status des Eu- bzw. Heterochromatins festlegen, sind epigenetisch. Sie werden somit bei der Zellteilung an die Tochterzellen vererbt. Während der DNA-Replikation werden die ursprünglichen, modifizierten Nukleosome auf die Tochter-stränge verteilt und es schließt sich die Chromatin-Replikation an. Spezielle Proteinkomplexe binden an die Modifikationen der „alten" Nukleosome und induzieren die Etablierung der gleichen Modifikationen in benachbarten „neuen" Nukleosomen. In diesem Zusammenhang sei das antagonistische Trithorax/Polycomb-System zur dauerhaften Vererbung des Expressionsstatus von Hox-Genen erwähnt. Diese Gene werden während der frühen Embryonalentwicklung durch H3K4-Trimethylierung aktiviert oder durch H3K27-Trimethylierung inaktiviert. Der Trithorax-Proteinkomplex bindet nach jeder Zellteilung H3K4me3 der alten Nukleosomen, repliziert die Histonmodifikation in den neuen Nukleosomen, rekrutiert HATs und hält die Gene somit in einem aktiven Zustand.[29] Das Pendant hierzu bildet der Polycomb-Proteinkomplex, der H3K27me3 bindet, repliziert und dadurch die Stilllegung der entsprechenden Gene aufrechterhält.[29, 41] Polycomb-Proteine sind für die Erhaltung der Pluripotenz von ES-Zellen wichtig: Mehr als 1000 Gene, darunter viele Zelllinien- und differenzierungsspezifische Gene, werden von ihnen in einem inaktiven Zustand gehalten.[39] Bekannt sind Polycomb-Proteine auch für die dauerhafte X-Chromosom-Inaktivierung von weiblichen Säugetieren.[29]

Zur langfristigen Stilllegung der Gene wird das Chromatin nicht nur an Histonen, sondern auch an der DNA kovalent modifiziert. DNA besteht im wesentlichen aus den vier kanonischen Nukleosiden 2′-Desoxyadenosin (dA), -thymidin (dT), -cytidin (dC) und -guanosin (dG), die über Phosphorsäurediester zu einer eindimensionalen Sequenz assembliert sind (**Abbildung 1A**).[42-44] Über spezifische Wasserstoffbrückenbindungen gehen die Nukleobasen komplementäre Paarungen mit einem DNA-Gegenstrang ein.[43, 45-47] Das Wasserstoffbrückenbindungsmuster sowie die Oberflächen-Komplementarität der Basenpaare sind nicht nur für die Ausbildung der DNA-Doppelhelix verantwortlich, sondern bilden ferner auch die Grundlage für die DNA-Replikation[48-49] und die Genexpression (Transkription und Translation).[50-52] Die Oberfläche der DNA-Doppelhelix ist in kleine und große Furchen gegliedert (**Abbildung 1**). Hierbei stehen die Nukleobasen in Kontakt zur flüssigen Phase, wobei in der große Furche sequenzspezifische Interaktionen mit Proteinen wie z.B. den Transkriptionsfaktoren stattfinden, während in der kleinen Furche nur kleine Moleküle binden

können.[50, 53] Neben den kanonischen Nukleosiden kennt man seit 1948 die Modifikation 5-Methyl-2′-desoxycytidin (5mdC, **Abbildung 1A**), die durch postreplikative Methylierung von dC gebildet wird.[54] Die Methylgruppe von 5mdC befindet sich in einer exponierten Position in der großen Furche (**Abbildung 1B**) und beeinflusst deshalb Proteininteraktionen, wodurch eine Diskriminierung von Genomabschnitten bei der Regulation der Genexpression möglich ist. 5mdC gehört zu den wichtigsten Markern in der Epigenetik und ist generell mit transkriptionell inaktiven Chromatinbereichen assoziiert. Auf dieses Thema wird im den nächsten Kapitel näher eingegangen.[1, 22-23, 55-56]
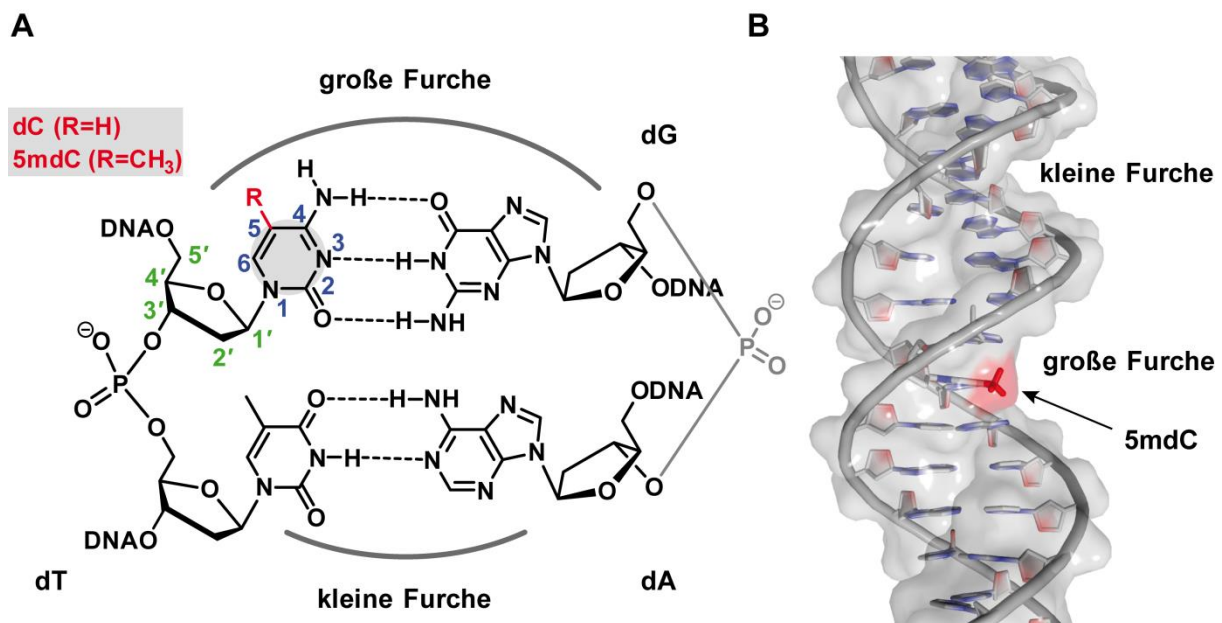
Abbildung 1 | Die doppelsträngige DNA-Struktur. (**A**) Basenpaarungen in doppelsträngiger DNA. dC und dG bilden drei Wasserstoffbrückenbindungen aus, dT und dA hingegen nur 2 (gestrichelte Linien). In grün und blau sind die Ringnummerierungen der Desoxyribose- bzw. Pyrimidineinheit dargestellt. Atome und Reste der Nukleobasenpaare (Wassenstoffbrückenbindungsakzeptoren, -donatoren, Methylgruppen), die an der Oberfläche der kleinen und großen Furche der B-förmigen DNA exponiert sind, wurden mit einem grauen Halbbogen hervorgehoben. (**B**) Die Methylgruppe der epigenetischen DNA-Modifikation 5mdC zeigt in die große Furche der DNA-Doppelhelix und beeinflusst daher Proteinen-DNA-Wechselwirkungen.

1.3 DNA-Methylierung

Als Bestandteil der genomischen DNA höherer Organismen gehört 5-Methyl-2′-desoxycytidin (5mdC) heute zu den am besten verstandenen epigenetischen Modifikationen.[1, 22-23, 55-56] Die Modifizierung wird postreplikativ durch DNA-Methyltransferasen (Dnmt) mit dem Cofaktor *S*-Adenosylmethionin (SAM) als Methylgruppendonor eingeführt. In Säugetieren kommen drei katalytisch aktive Dnmts vor (Dnmt1, 3a und 3b) sowie eine verwandte, aber katalytisch inaktive Form (Dnmt3l).[57-58] Die DNA-Methylierung ist für die Entwicklung von Säugetieren unentbehrlich. Dnmt1-, Dnmt3a- oder Dnmt3b-Knockout-

Mäuse sterben in frühen oder späten Phasen ihrer Entwicklung.[58-60] Dnmt1 ist für die Erhaltung von 5mdC-Mustern nach jeder Zellteilung verantwortlich und macht 5mdC damit zu einer vererbbaren, d.h. epigenetischen, Modifikation. Dnmt3a und Dnmt3b führen als *de novo*-Methyltransferasen neue 5mdC-Muster bei Differenzierungsprozessen ein, wobei Dnmt3l ein positiver Regulator ist.[57-58] Dnmt3a und 3b sind trotz hoher Sequenzhomologie nicht redundant, obwohl sie teilweise die Funktion des jeweils anderen Enzymes übernehmen können und sind beide für eine normale Entwicklung und Lebensfähigkeit essentiell.[58-60]

1.3.1 Katalytischer Mechanismus von DNA-Methyltransferasen

Der kinetische Mechanismus von Dnmts wird als ein geordneter bi-bi-Mechanismus beschrieben, bei dem sich durch Bindung von DNA und anschließend von SAM zunächst ein ternärer Komplex bildet, bevor die chemische Reaktion eintritt (**Abbildung 2**).[61]
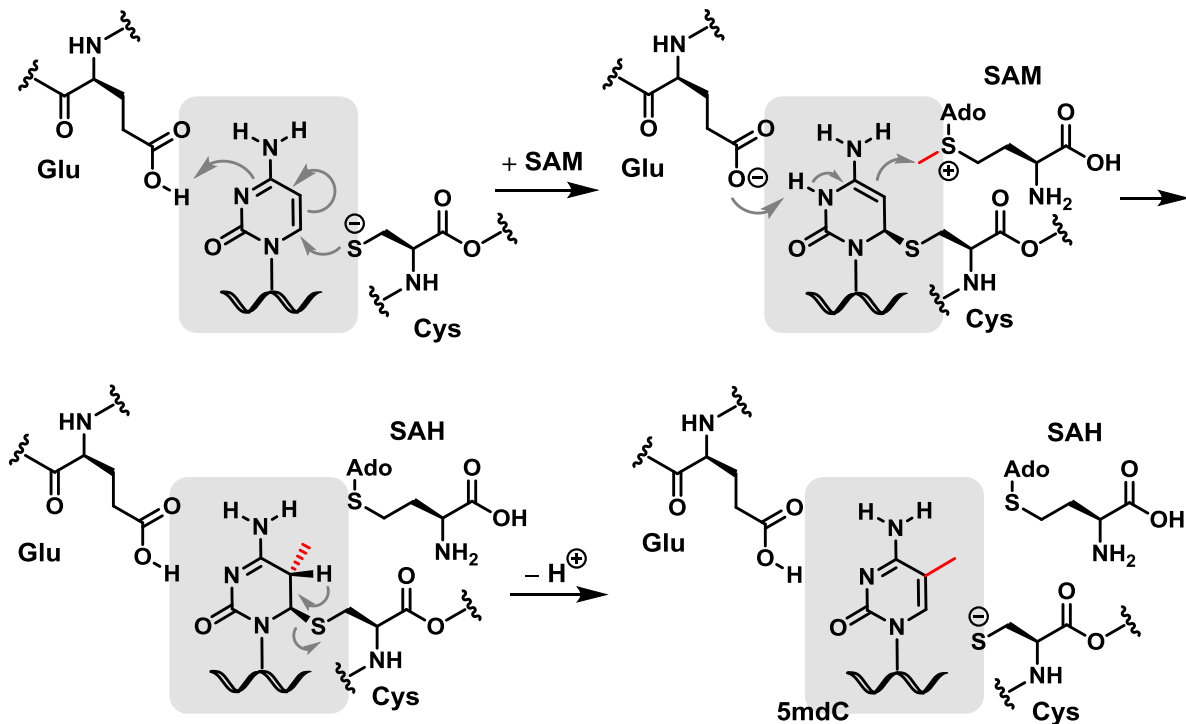
Abbildung 2 | Katalytischer Mechanismus von DNA Methyltransferasen.[62] SAM: *S*-Adenosylmethionin; SAH: *S*-Adenosylhomocystein.

Der Mechanismus beginnt mit dem Herausdrehen von Cytosin aus der DNA-Doppelhelix (*flipping*-Mechanismus) und der transienten Absättigung der C5=C6-Doppelbindung durch das Schwefelnukleophil Cystein in einer *Michael*-ähnlichen 1,4-Addition, bei der ein Glutaminsäurerest die N3-Position protoniert. Das resultierende Enamin-Intermediat greift dann den Methylgruppendonor SAM in einer S_N2-ähnlichen Reaktion an. Durch die

Abstraktion des H-Atoms in C5-Position wird die Eliminierung des Cysteins eingeleitet und abschließend die Nukleobase rearomatisiert.[57, 62]

1.3.2 Häufigkeit und Verteilung von 5mdC in Säugetierzellen

In somatischen Säugetierzellen beträgt der 5mdC-Gehalt zwischen 4–6% aller Cytidine,[3, 63-64] wobei die Methylgruppe hauptsächlich an palindromischen CpG-Dinukleotid-Stellen (CpG-Dyade) auf beiden Seiten der Helix zu finden ist. Methylierte CpH-Dinukleotide (H = A, C oder T) sind zwar häufig in Pflanzen, aber generell selten in Säugetierzellen.[56] Ausnahmen findet man in Oozyten, ES-Zellen und adulten Neuronen (hier bis zu 53% bezogen auf die globalen 5mdC-Gesamtlevel).[64-66] Methylierte CpH-Dinukleotide kompensieren vermutlich den Mangel von 5mCpG in CpG-armen Genomabschnitten.[65] Dies ist insofern von Relevanz, da CpG-Dinukleotide in Säugetier-Genomen evolutionär gesehen unterrepräsentiert sind, was auf Punktmutationen (C→T Transitionen) durch Desaminierung von 5mdC zurückzuführen ist.[58, 67-69] Im menschlichen Genom liegt der G/C-Gehalt bei ca. 40%, daher würde man theoretisch CpG-Dinukleotide mit einer Häufigkeit von 4% erwarten. Tatsächlich existieren allerdings nur ca. 0.9%,[70] was einer Anzahl von 56 Millionen CpG-Dinukleotiden entspricht, von denen 60–80% zelltypspezifisch methyliert sind.[1] In Promotorregionen findet man häufig sogenannte CpG-Inseln (im Menschen ca. 60% aller Promotoren). Das sind durchschnittlich 1000 Basenpaar-lange DNA-Abschnitte, die sich durch eine höhere Dichte an CpG-Dinukleotiden im Vergleich zum restlichen Genom auszeichnen.[68-69, 71-72] Diese machen ca. 10% aller CpG-Dinukleotide aus, sind in der Regel hypomethyliert, d.h. vor Methylierung geschützt, und stellen wichtige Elemente der Transkriptionskontrolle dar.[1, 73]

1.3.3 5mdC während der zellulären Entwicklung

Alle Zellen eines multizellulären, bilateral-symmetrischen Organismus (z.B. Säugetiere) stammen von einer befruchteten Eizelle ab, der Zygote. Diese teilt sich zunächst durch Furchung und differenziert sich schrittweise über die sogenannten 2-Zell-, 4-Zell-, 8-Zell-Stadien, zur Morula und Blastozyste weiter, während sie entlang des Eileiters Richtung Gebärmutter wandert (**Abbildung 3A**). Die Zygote ist totipotent, d.h. sie kann sich in alle Zellen eines Organismus entwickeln bzw. einen gesamten Organismus erschaffen. Die frühe Blastozyste besteht aus den Trophoblasten, die eine schützende Hüllschicht bilden und sich zur Plazenta weiterentwickeln, und der inneren Zellmasse (IZM), die sich in alle drei Keimblätter (Ento-, Meso-, Ektoderm) des Embryos weiterentwickelt. Die embryonalen Stammzellen (ES-Zellen) der IZM können sich zwar zu allen Gewebezellen eines Organismus

entwickeln, aber keinen gesamten Organismus mehr erschaffen und sind deshalb nur noch pluripotent.[39, 74-76] Aus der IZM dieser frühen prä-implantiven Embryoblasten können die ES-Zellen gewonnen und unter geeigneten Bedingungen kultiviert werden. Mit Hilfe von speziellen Medien kann ihr pluripotenter Status erhalten werden, die Zellen in unterschiedliche Entwicklungsstadien versetzt bzw. in viele Zelltypen ausdifferenziert werden (siehe **Abbildung 3B** und Legende) [74, 77-83]
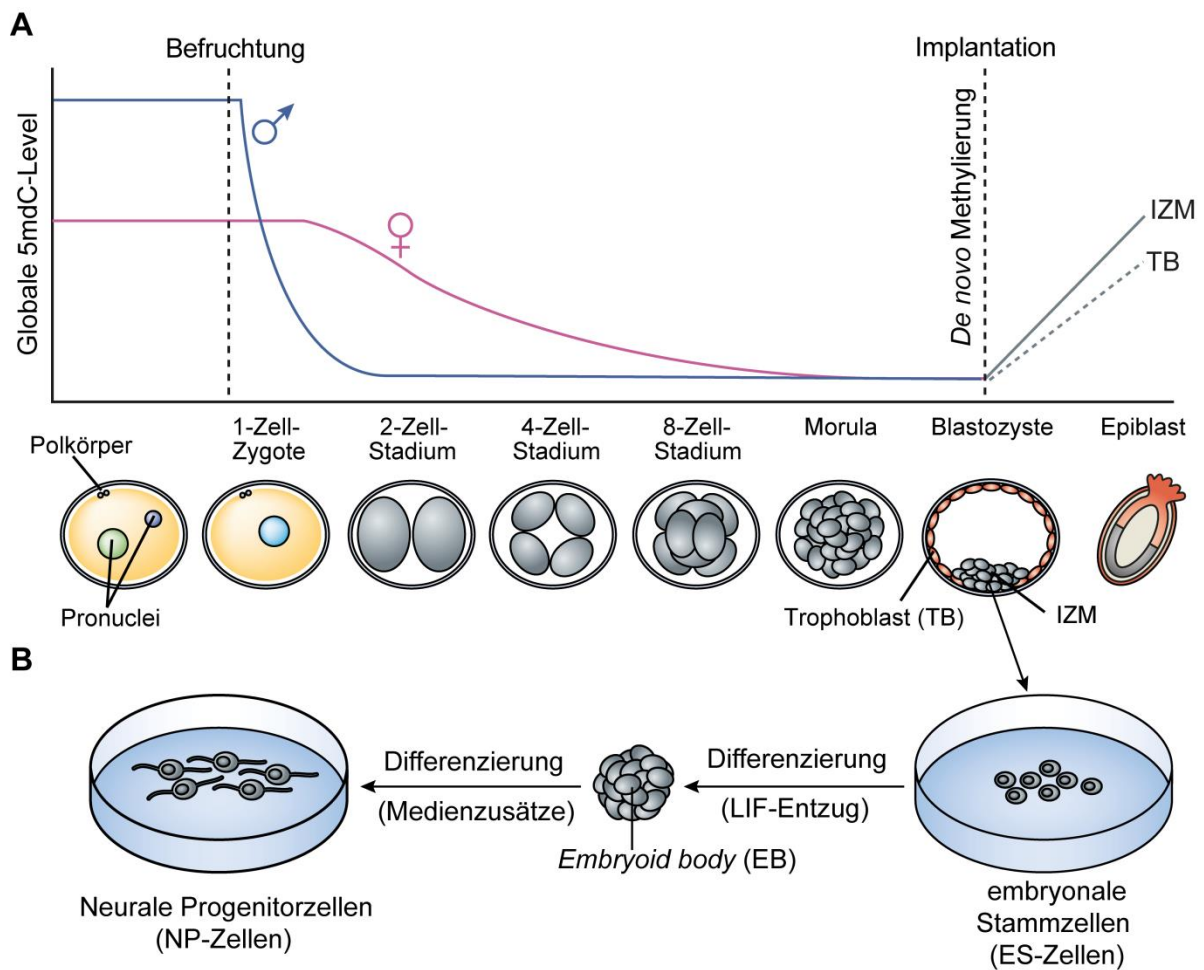
Abbildung 3 | (**A**) Dynamik der DNA-Methylierung während der frühen embryonalen Entwicklung; siehe Text. (**B**) Embryonale Stammzellen (ES-Zellen) werden aus der inneren Zellmasse (IZM) im frühen Blastozysten-Stadium gewonnen. Diese können unter geeigneten Bedingungen kultiviert werden, z.B. in Anwesenheit des Pluripotenz-erhaltenden Faktors LIF (*leukemia inhibitory factor*). Unter LIF-Entzug differenzieren ES-Zellen *in vitro* zu sogenannten *embryoid bodies* aus, welche in speziellen Medien in viele Zelltypen ausdifferenziert werden können. Mittlerweile gibt es eine Reihe von Arbeiten, die geeignete Kulturbedingungen beschreiben, mit denen ES-Zellen in Zuständen kultiviert und differenziert werden können, die den natürlichen Gegebenheiten in der Bastozyste sehr nahe kommen.[74, 77-83] Die Abbildung wurde von *Meshorer* und *Misteli* sowie *Wu* und *Zhang* adaptiert.[39, 76, 84]

Bei all diesen zellulären Differenzierungsprozessen wird ein komplexes Zusammenspiel zwischen den epigenetischen Chromatin-Modifizierungen beobachtet.[1, 23] Nicht nur die

Histon-Modifizierung, sondern auch die DNA-Methylierung ist ein hochkontrollierter und dynamischer Prozess. Während sich die Zygote zur Blastozyste differenziert, erfolgt eine nahezu vollständige globale Demethylierung des männlichen und weiblichen Genoms (siehe **Abbildung 3A**), um den pluripotenten Status des frühen Embryos zu etablieren.[85] Dabei verläuft die Demethylierung des Genoms des männlichen Elternteils schneller als die Demethylierung des weiblichen Genoms. Bei der Demethylierung des weiblichen Genoms ging man in der Vergangenheit bisher nur von einem rein passiven Demethylierungs-mechanismus aus, d.h. dem Verlust von 5mdC durch Verdünnung bei der DNA-Replikation.[86-94] Eine neue Studie weist aber auch hier auf aktive bzw. Enzym-vermittelte Demethylierungsanteile hin.[95] Näheres zu den zugrundeliegenden molekularen Mechanismen wird später in Abschnitt 1.4 und 1.6 abgehandelt.

Nachdem die Blastozyste den Eileiter verlassen hat, geschieht dann um den Zeitpunkt der Einnistung in die Gebärmutterschleimhaut (Implantation) eine programmierte *de novo* Methylierung spezifischer DNA-Bereiche.[84-85, 88-90, 93, 96-98] Dnmt3a und Dnmt3b werden hierbei durch den Histon-Code an die entsprechenden Stellen im Genom rekrutiert. In den differenzierenden ES-Zellen müssen beispielsweise die CpG-Inseln der Pluripotenzgene wie z.B. *Oct3/4* und *Nanog* für die dauerhafte Stilllegung methyliert werden. Diese liegen in ES-Zellen unmethyliert vor und sind mit acetylierten (H3 und H4) und methylierten Histonen (H3K4) assoziiert. Sie befinden sich daher in einem aktiven Zustand (siehe Abschnitt 1.2). Mit dem Beginn der ES-Zelldifferenzierung wird ein multifunktioneller Proteinkomplex gebunden. Zunächst werden alles Lysine deacetyliert und H3K4 demethyliert, womit die Transkriptions-assoziierten Signale entfernt werden. Der Proteinkomplex enthält des weiteren die Histonmethyltransferase G9a, die eine SET-Domäne enthält. SET ist ein evolutionär konserviertes Sequenzmotif, steht für *Su(var)3-9*, *Enhancer-of-zeste* sowie *Trithorax*, und ist als Bestandteil vieler Histon-Methyltransferasen für deren katalytische Aktivität notwendig.[23, 99] G9a kann nach der Deacetylierung der Lysine nun H3K9 methylieren, wodurch die Bindestelle für das Heterochromatin-Protein 1 (HP1) geschaffen und damit die lokale Heterochromatin-Bildung eingeleitet wird. G9a rekrutiert darüber hinaus Dnmt3a/3b, wodurch die finale DNA-Methylierung der stillzulegenden Pluripotenzgene umgesetzt wird.[23] Nach der Differenzierung wird die *de novo* DNA-Methylierungsmaschinerie heruntergefahren.

Während der Histoncode in der frühen Zellentwicklung scheinbar generell für die Generierung von Methylierungsmustern in der DNA verantwortlich ist, gibt es aber auch

Hinweise, dass das Methylierungsmuster nach der Differenzierung als Templat für die Reproduzierung von Histonmodifikationen dient:

Bei jeder Zellteilung wird die Chromatinstruktur durch die DNA-Replikationsgabel zertrennt, deshalb sind wiederherstellende Mechanismen nötig. Hierfür kann das DNA-Methylierungs-muster den nötigen Anker liefern, da dieses bei der DNA-Replikation durch Dnmt1 erhalten wird.[23] Dnmt1 wird durch Uhrf1 (*ubiquitin-like, containing PHD and RING finger domains 1*) an die Replikationsgabel rekrutiert, welches kooperativ H3K9me2/3 sowie hemimethylierte DNA bindet, um eine zuverlässige Vererbung des 5mdC-Musters sicherzustellen.[100-102] Die 5mdC-Regionen sind als Bindestellen sogenannter Methyl-CpG bindender Proteine wie MeCP2 und Mbd2[103] bekannt, die Histondeacetylasen an die methylierten DNA-Regionen rekrutieren können. Desweiteren ist 5mdC vermutlich ein Signal für die Methylierung von H3K9, da Dnmt1 sowie Uhrf1 mit der H3K9-Methyltransferase G9a wechselwirken.[23, 102]

Daneben kennt man aber auch die 5mdC-unabhängige Reproduktion von Chromatinstrukturen nach der DNA-Replikation, wie z.B. die in Abschnitt 1.2 beschriebene Polycomb-vermittelte Reproduktion des epigentischen Repressions-Markers H3K27me3.[29, 41] Beide Prozesse sind somit für die Erhaltung von Heterochromatin-Strukturen und die Stilllegung von Genen verantwortlich. Ein bidirektionaler "*cross talk*" zwischen Histonmodifikationen und 5mdC wird vielfach beobachtet, womit die Natur denkbarerweise mehrere Signale integriert, um die Fehlerrate bei der epigenetischen Genregulation zu reduzieren. Die langfristige Stilllegung von Genen scheint aber nur durch die DNA-Methylierung gewährleistet zu sein. Als Beispiele seien dabei die oben erwähnte dauerhafte Stilllegung der Pluripotenzgene bei der zellulären Entwicklung, die X-Chromosom-Inaktivierung oder die genomische (elterliche) Prägung genannt.[22-23]

Nach der Befruchtung der Eizelle ist es deshalb für die epigenetische Reprogrammierung wichtig, die elterlichen 5mdC-Muster zu entfernen, damit die Entwicklung des Nachkommens aus nur einer Zelle gelingt. Die DNA-Methylierung ist nicht nur, wie oben skizziert, in der frühen embryonalen Entwicklung dynamisch (**Abbildung 3A**), eine aktive, globale DNA-"Demethylierungswelle" findet im Embryo auch später bei der epigenetischen Reprogrammierung in Urkeimzellen statt.[84-85, 104-106] Nach der zellulären Entwicklung sind im erwachsenen Individuum solche globalen Demethylierungswellen nicht mehr zu beobachten. Aktive Demethylierungsereignisse treten in bestimmten DNA-Loci nur noch vereinzelt auf; z.B. bei aktivierenden Prozessen in T-Zellen oder Neuronen als Reaktion auf einen Stimulus.[76]

1.4 DNA-Hydroxymethylierung

Im Jahr 2009 wurde 5-Hydroxymethyl-2′-desoxycytidin (5hmdC) von zwei unabhängigen Forschungsgruppen als weitere Modifikation in genomischer DNA im Gehirn und in Stammzellen nachgewiesen,[2, 5] die nun generell als die "sechste Base" des Genoms höherer Organismen angesehen wird.[107] Diese Modifikation wird durch Fe(II)- und 2-Oxoglutarat-abhängigen Dioxygenasen, den Tet-Enzymen (*ten eleven translocation*), durch Oxidation der Methylgruppe von 5mdC postreplikativ gebildet (**Schema 6**).[2, 5]
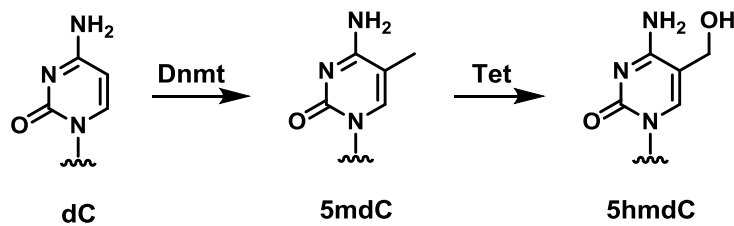
Schema 6 | In der genomischen DNA von Säugetieren kann Cytidin methyliert und hydroxymethyliert vorliegen.

Zu Beginn dieser Dissertation im Frühjahr 2011 deutete sich bereits an, dass 5hmdC wahrscheinlich eine weitere wichtige Rolle bei der epigenetischen Kontrolle der Zellentwicklung und Genexpression spielt, indem es die epigenetische Informationsdichte für die komplexen Kontrollmechanismen erhöht. Darüber hinaus vermutete man, dass es sich bei 5hmdC um ein Reaktionsintermediat eines lange gesuchten Mechanismus der DNA-Demethylierung handeln könnte. Dieser Mechanismus sollte für die aktive Demodifizierung und damit für die Reaktivierung von Genen verantwortlich sein, die durch Methylierung stillgelegt wurden, wie im vorangegangen Kapitel erläutert wurde.[76, 107] Seit dieser Zeit fokussierten sich die epigenetischen Forschungsgruppen weltweit auf diese neue Modifikation und erlangten Erkenntnisse über Bildung, Abbau und Funktion in rasanter Geschwindigkeit. Speziell die Untersuchung der aktiven DNA-Demodifizierung von 5hmdC war zentraler Bestandteil dieser Dissertation. Deshalb werden die entsprechenden Erkenntnisse anderer Arbeitsgruppen, die zu diesem Thema parallel oder anschließend publiziert wurden, in der nachfolgenden Einleitung nicht, aber im späteren Hauptteil dieser Arbeit berücksichtigt. Die nachfolgenden Abschnitte beschäftigen sich ausführlich mit den enzymologischen Grundlagen der 5hmdC-Generierung, der Häufigkeit und Verteilung von 5hmdC im Genom, dessen biologische Funktionen und der therapeutischen Relevanz.

1.4.1 Oxidation von Methylgruppen durch Tet und andere Dioxygenasen

Die Tet-Proteine zählen, wie bereits erwähnt, zu den Fe(II)- und 2-Oxoglutarat-abhängigen Dioxygenasen. Diese bewerkstelligen u.a. die thermodynamisch gesehen herausfordernde C–H-Aktivierung und Oxidation von Methylgruppen durch bimolekularen Sauerstoff.[108-112]

Bevor gezeigt wurde, dass Tet die Methylgruppe von 5mdC hydroxyliert (**Schema 6**),[5] kannte man bereits zwei ähnliche Mechanismen, bei denen die Dioxygenasen Jbp1/2 (J-bindendes Protein) und T7h (Thymin-7-Hydroxylase) die Methylgruppe von Thymin oxidieren (**Schema 7**):
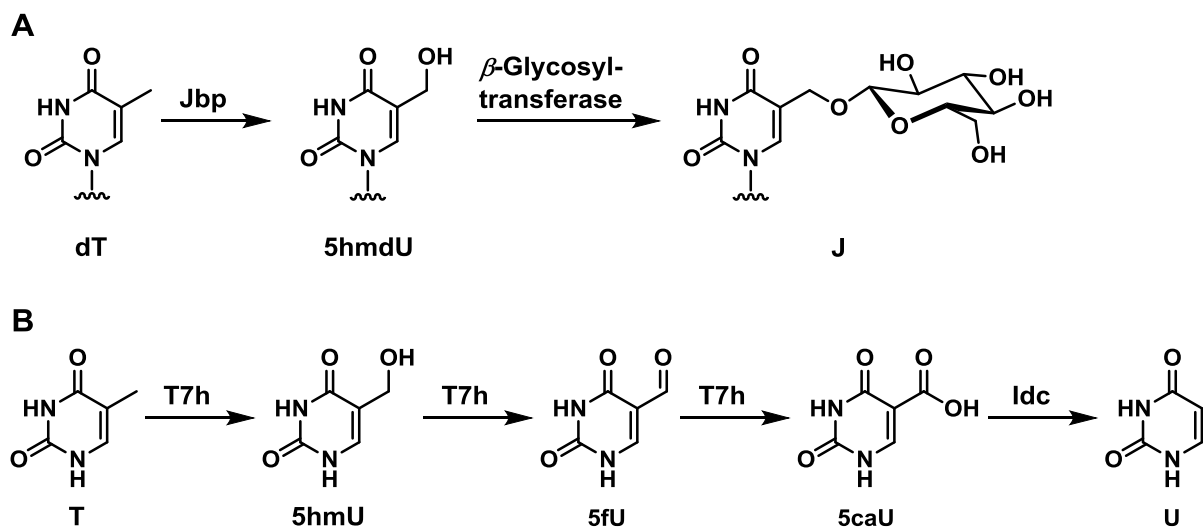
Schema 7 | Oxidation von Methylgruppen durch Dioxygenasen. (**A**) Der Parasit *Trypanosoma brucei* modifiziert sein Genom durch postreplikative Hydroxylierung der Methylgruppe von dT durch die Dioxygenasen Jbp1/2 und anschließender Glykosylierung der Hydroxy-Gruppe durch eine β-Glykosyltransferase. Dabei entsteht die sogenannte Nukleobase J (5-(β-D-Glukopyranosyl-oxymethyl)-2′-desoxyuridin). (**B**) Demethylierung von Thymin zu Uracil in *Neurospora crassa* mittels sukzessiver Oxidation durch die Thymin-7-hydroxylase (T7h) und anschließender Decarboxylierung durch die Isoorotatdecarboxylase (Idc).

Das Genom von Kinetoplasten, wie das des Parasiten *Trypanosoma brucei*, welches die afrikanische Schlafkrankheit auslöst, enthält die sogenannte Nukleobase J. Diese wird in einem zweistufigen Prozess postreplikativ durch Hydroxylierung der Methylgruppe von dT durch Jbp1 und Jbp2, sowie anschließender Glykosylierung der Hydroxy-Gruppe durch eine β-Glykosyltransferase gebildet (**Schema 7A**).[113-117] Bioinformatische Analysen von Jpb1/2 führten schließlich zur Vorhersage,[118] dass Tet-Proteine eine 5mdC-oxidierende Aktivität haben könnten und trieben diese Entdeckung voran.[5, 119] Neben Jbp und Tet gehören auch die entfernt verwandten AlkB-Reparaturenzyme, welche aberrante Methylgruppen geschädigter

Nukleobasen entfernen,[120-122] als DNA-bindende Proteine zur Superfamilie der Fe(II)- und OG-abhängigen Dioxygenasen.[108-112]

Desweiteren können manche Pilze, wie *Neurospora crassa*, Thymin demethylieren, um Uracil zurückzugewinnen. Hierzu katalysiert T7h die iterative Oxidation zu 5-Hydroxymethyl-, 5-Formyl- und 5-Carboxyuracil (5hmU, 5fU und 5caU);[109] ein sich anschließender Decarboxylierungsschritt wird durch die Isoorotatdecarboxylase (Idc) verwirklicht (**Schema 7B**).[123-125] Dieser Mechanismus inspirierte die Forschungsarbeit dieser Dissertation und anderer Arbeiten[3, 76, 110] bei der Untersuchung der aktiven DNA-Demethylierung (siehe Kapitel 1.6, 3 und 4).

Tet-Proteine: Vorkommen, Struktur, Substrat- und DNA-Erkennung

Bevor man die Funktion der Tet-Enzyme entdeckte, kannte man das humane Tet1-Protein bereits als aberranter Fusionspartner der H3K4-Methyltransferase MLL (*Mixed Lineage Leukemia*) in Leukämie-Patienten. Das entsprechende *Tet1*-Gen translozierte bei diesen Patienten von Chromosom 10 zu Chromosom 11 (*Ten Eleven Translocation*). Dieser Sachverhalt wurde daher namensgebend für das Protein.[126-127] Durch bioinformatische Methoden identifizierte man im Menschen zwei weitere paraloge Gene, *Tet2* und *Tet3*,[127] und orthologe Gene im gesamten Reich vielzelliger Tiere, deren genomische DNA 5mdC enthält.[5, 119]

Alle drei Tet-Proteine haben eine C-terminale katalytische Domäne, die sich aus einer Cystein-reichen Region und einer doppelsträngigen β-Helix-Region (DSBH) zusammensetzt (**Abbildung 4**).[5, 119] Die Kristallstrukturanalyse der katalytischen Domäne von humanem Tet2 im Komplex mit 5mdC-enthaltender DNA zeigte, dass die Cystein-reiche Region keine separate strukturelle Einheit ist, sondern sich in eine N- und C-terminale Subdomäne unterteilt, die sich um den DSBH-Kern winden.[128] Hierbei unterstützen drei Zn(II)-Zentren (Zinkfinger) den Zusammenhalt der katalytischen Domäne und damit die Stabilisierung der gebogenen DNA-Struktur. Dies schafft die Grundlage für das Herausdrehen von 5mdC aus dem DNA-Duplex durch einen *flipping*-Mechanismus und die Positionierung der Methylgruppe in direkter Nähe zum katalytischen Fe(II)-Zentrum in der DSBH-Region (**Abbildung 5**).[128-129]
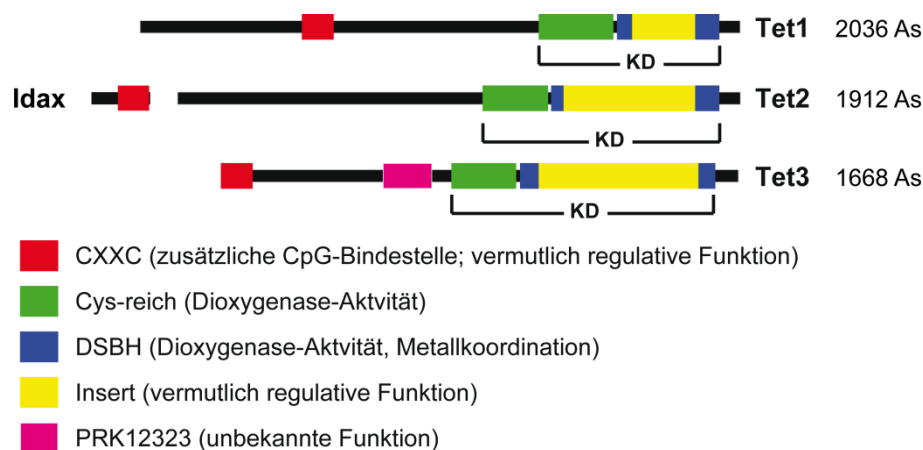
Abbildung 4 | Schematische Darstellung der Domänen innerhalb der Primärstruktur von murinen Tet-Proteinen. Alle drei haben eine C-terminale katalytische Domäne (KD), die sich aus einer Cys-reichen Region und einer doppelsträngigen β-Helix-Region (DSBH) zusammensetzt. Innerhalb der DSBH-Sequenz befindet ein wenig konserviertes Insert (gelb), dessen Länge unter den Tet-Proteinen variiert und die katalytische Aktivität kaum beeinflusst.[119, 128-130] Die N-Termini enthalten eine CXXC-Domäne, außer Tet2, welches mit Idax (CXXC4) assoziiert. Tet3 hat zudem eine PRK12323 Domäne (DNA-Polymerase III Untereinheit), deren Funktion noch ungeklärt ist. As = Aminosäuren. Die Abbildung wurde von *Tan* und *Shi* adaptiert.[131]

Die DSBH-Regionen fast aller Dioxygenasen enthalten mit der Sequenz His-X-Asp/Glu-X_N-His ein hochkonserviertes dreizähniges Fe(II)-Bindungsmotiv, das eine Fläche des oktaedrischen Fe-Zentrums bildet (**Abbildung 5**). Zusätzlich enthält die DSBH-Region ein konserviertes Arginin, das OG zusammen mit zwei Fe(II)-Koordinationsstellen bindet.[108-109, 111-112, 128, 132-135] Die Bindung bzw. Erkennung von 5mC erfolgt über zwei Wasserstoffbrückenbindungen zwischen der N3- und N4-Position der Nukleobase und den Seitenketten von Histidin und Asparagin (**Abbildung 5**). Interessanterweise indizieren Sequenzvergleiche, dass in Jbp-Proteinen anstelle dieser Aminosäurereste Aspartat und Arginin für die molekulare Erkennung von Thymin verantwortlich sein könnten. Die Bindung der zu oxidierenden Nukleobase wird daneben noch zusätzlich durch *Stacking*-Effekte mit einem konservierten Tyrosin-Rest unterstützt.

Tet-Proteine haben darüber hinaus eine Substratpräferenz für 5mCpG-Dinukleotide.[128-129] In der Tet2-Kristallstruktur wurden außer zu dem CpG-Dinukleotid keine weiteren Kontakte zu flankierenden Nukleobasen beobachtet, sondern nur Wasserstoffbrückenbindungen zu einigen Phosphaten (**Abbildung 5**). Hierdurch kann strukturell gesehen eine höhere Sequenzspezifität ausgeschlossen werden. Die molekulare Erkennung von CpG-Dinukleotiden ist scheinbar für die Positionierung der Methylgruppe am Reaktionsort essenziell, sodass 5mCpH-Dinukleotide (H= A, T, C) mit geringeren Reaktionsraten oxidiert werden.[128] Die Methylgruppe von 5mC

geht im aktiven Zentrum keine Proteinkontakte ein, so dass die umgebende Proteintasche genügend Platz für das Oxidationsprodukt bietet.[128]
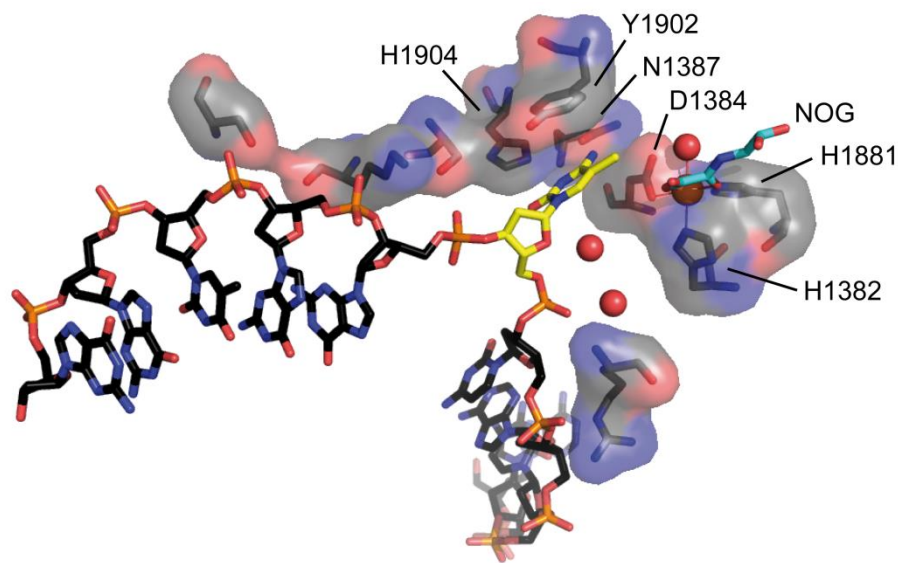
Abbildung 5 | Kristallstrukturausschnitt von humanem Tet2 im Komplex mit 5mdC-enthaltender DNA. Dargestellt sind der zu oxidierende DNA-Strang und alle Aminosäurereste die an dessen direkter Protein-Wechselwirkung beteiligt sind, sowie die Aminosäurereste der Fe(II)-Koordination. Die Wassermoleküle bzw. das Fe(II)-Atom sind mit roten bzw. orangen Sphären, das OG-Analogon *N*-Oxalylglycin (NOG) in cyan dargestellt.[128]

Zusätzlich zur C-terminalen katalytischen Domäne besitzen Tet-Enzyme zwei weitere Domänen, über die wahrscheinlich die genomische Lokalisation und die enzymatische Aktivität reguliert wird. Zum einen besitzen alle Tet-Enzyme eine variable weniger konservierte Einfügung (Insert) innerhalb der Primärstruktur der C-terminalen DSBH-Domäne (**Abbildung 4**). Das Insert ist je nach Tet-Protein unterschiedlich lang[119, 130-131] und steht in der Tertiärstruktur von der katalytischen Domäne weg.[128-129] In Tet1 weist das Insert zudem eine hohe Ähnlichkeit zur C-terminalen Domäne (CTD) der RNA-Polymerase II (*S. cerevisiae*) auf.[136] Darüber hinaus besitzen Tet1 und Tet3 je eine N-terminale CXXC-Domäne (~60 Aminosäuren), die vielen Chromatin-assoziierten Proteinen als gängiges DNA-Bindungsmotiv gemeinsam ist.[119, 127] Das *Tet2*-Gen durchlief evolutionär eine chromosomale Inversion, bei der der CXXC-Domänenabschnitt abgetrennt wurde und sich zu einem eigenständigen Gen namens *Idax* (oder *CXXC4*) entwickelte.[119, 130] Eine CXXC-Domäne koordiniert ein Zn(II)-Ion und enthält zwei Kopien des konservierten Motivs CGXCXXC(X)$_N$C, wobei X jede beliebige Aminosäure repräsentiert.[134] CXXC-Domänen von Chromatin-assoziierten Proteinen binden typischerweise bevorzugt an unmethylierte CpG-Stellen.[137-139] Dementsprechend zeigten Genom-weite Lokalisationsstudien, dass die Tet-Enzyme hauptsächlich an unmethylierten DNA-Bereichen zu finden sind. Biochemische

Studien indizierten aber auch, dass die CXXC-Domänen der Tet-Enzyme untypischerweise über eine erhöhte Flexibilität in ihrer Sequenzselektivität verfügen.[140-142] So bindet Tet1-CXXC möglicherweise neben unmethylierten auch methylierte CpGs,[143-144] und Tet3-CXXC bindet sowohl an unmethylierte CpGs und CpHs (H = A, C, T).[145]

Katalytischer Mechanismus von Fe(II)- und OG-abhängigen Dioxygenasen

Der postulierte Mechanismus von Tet beruht auf dem Konsensmechanismus Fe(II)- und OG-abhängiger Dioxygenasen und ist in **Schema 8** dargestellt.[108-109, 111-112, 128, 132-135] Es ist davon auszugehen, dass die katalysierte Reaktion nach einem Ping-Pong-Mechanismus[146] verläuft, d.h. das Enzym wird zuerst durch ein Cosubstrat chemisch verändert, bevor im Anschluss das eigentliche Substrat umgesetzt wird.[146] In diesem Sinne erfolgt in einem ersten Schritt die Oxidation von Fe(II) zu Fe(IV) durch das Cosubstrat OG und bimolekularem Sauerstoff. In sequenziell geordneter Weise findet an Fe(II) zunächst ein Ligandenaustausch von zwei H_2O-Molekülen durch OG statt. Anschließend oder gleichzeitig bindet das primäre Substrat.[147-149] Hierdurch wird die Bindung des axialen H_2O-Moleküls geschwächt und die Bindung von O_2 begünstigt, bei der Fe(II) zu Fe(III) oxidiert.[112, 132, 150-151]

In einem irreversiblen Schritt erfolgt jetzt die oxidative Decarboxylierung von OG zu Succinat und die Oxidation von Fe(III), wodurch bimolekularer Sauerstoff homolytisch gespalten wird und sich das Fe(IV)-Oxo-Intermediat ausbildet (**Schema 8**).[108-109, 111-112, 128, 132-135, 152] Dieser irreversible Schritt ist bei den entfernt verwandten Dioxygenasen T7h oder der Prolyl-4-Hydroxylase ratenbestimmend, wie es durch Analyse von kinetischen Isotopen-effekten demonstriert wurde: Bei Verwendung von deuterierten oder tritiierten Thymin- bzw. Prolin-Derivaten wurde kein oder ein vergleichsweise geringer Isotopeneffekt während der Gesamtreaktion beobachtet.[133, 135, 147, 149] Das reaktive Fe(IV)-Oxo-Intermediat oxidiert dann das primäre Substrat durch C–H-Abstraktion, wobei ein Substrat-Radikal als Intermediat entsteht, das mit dem gleichzeitig entstandenen Fe(III)-OH-Komplex mit einer Übertragung eines Hydroxy-Radikals erneut reagiert. Die Dissoziation der Produkte Succinat und 5hmdC komplettiert abschließend den katalytischen Zyklus.
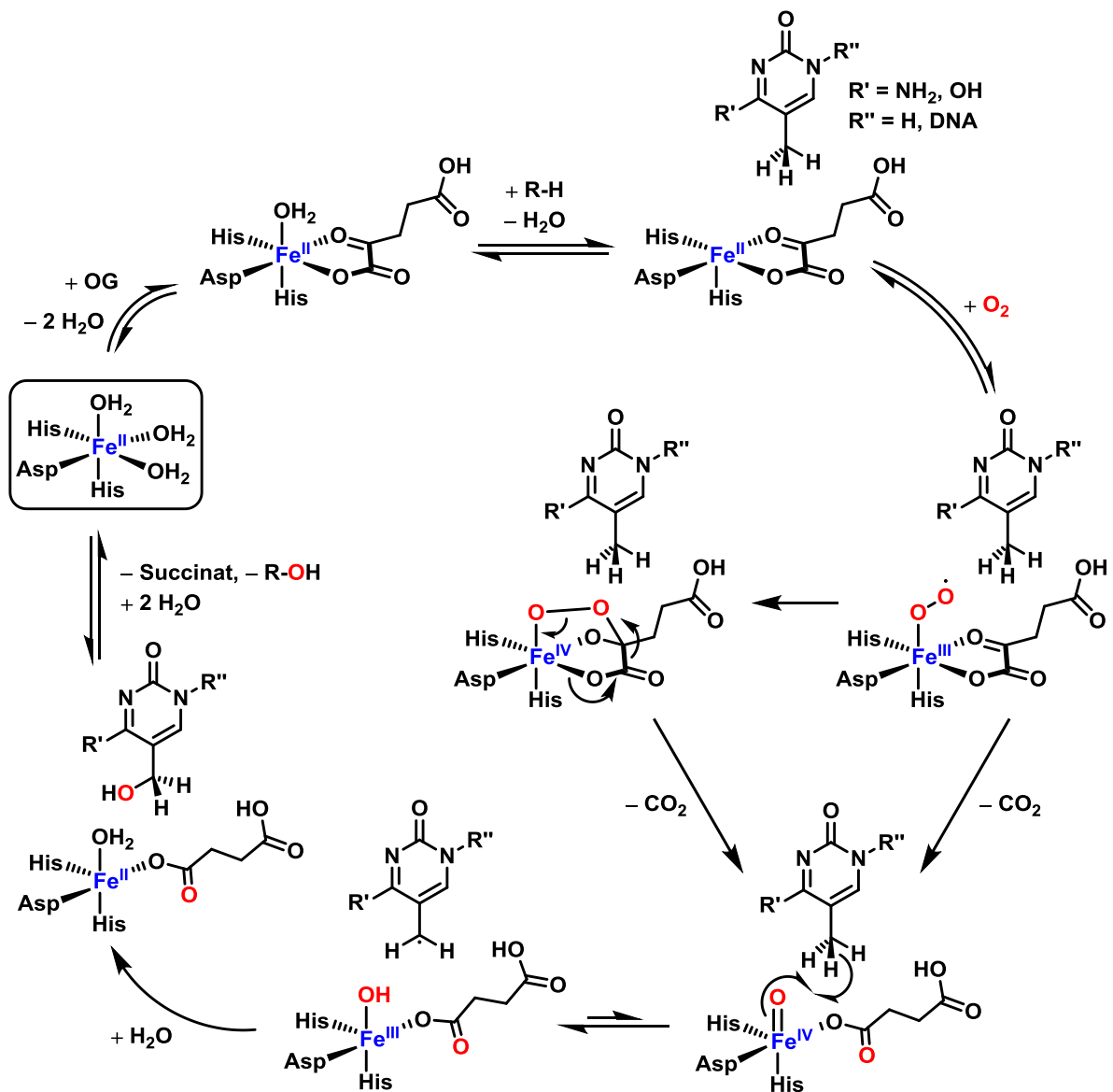
Schema 8 | Postulierter Reaktionsmechanismus der Fe(II)- und 2-Oxoglutarat (OG) abhängigen Dioxygenasen Tet und Jbp. Für alle gezeigten Intermediate gibt es bei verschiedenen Dioxygenasen Beweise, mit Ausnahme der O_2-Komplexe und des Fe(III)-OH-Intermediats.[108-109, 111-112, 128, 132-135] R-H: 5mdC oder dT als Substrate (R' = NH₂ bzw. OH; R'' = DNA). *Nota bene*: dT und seine Derivate (R' = OH) liegen ausschließlich als (C=O)-N(H)-(C=O)-Tautomer vor. Aus Gründen der Übersichtlichkeit wurde auf diese Darstellung verzichtet.

1.4.2 Häufigkeit und Verteilung von 5hmdC in Säugetierzellen

5hmdC wurde erstmals 1952 als Bestandteil des Genoms bestimmter Bakteriophagen entdeckt, die bei der DNA-Synthese in der Wirtszelle statt dem dC-Triphosphat das 5hmdC-Triphosphat verwenden und die Hydroxymethylgruppe anschließend postreplikativ glykosylieren.[153-159] In Säugetieren wurde genomisches 5hmdC zwar auch bereits 1972 per Papierchromatographie[160-161] in fragwürdig hohen Mengen entdeckt, aber nach Kontroversen über methodische Schwächen[162] zunächst nur noch als oxidativer Schaden von 5mdC

beschrieben (siehe Abschnitt 1.5).[163-164] Dank neuer analytischer Methoden, die sensitiver und exakter sind, wie der Kopplung von Hochdruck-Flüssigkeitschromatographie mit Massenspektrometrie (HPLC-MS), erfuhr die Identifizierung und Quantifizierung modifizierter Nukleoside eine Renaissance. 2009 entdeckten *Kriaucionis* und *Heintz* respektive *Tahiliani et al.* mit Hilfe radioaktiver Dünnschichtchromatographie (DC) 5hmdC im Mäusegehirn bzw. in embryonalen Stammzellen der Maus (mES-Zellen). Sie konnten den Gehalt mit 0.6% in Purkinje-Neuronen bzw. 0.15% pro dG in mES-Zellen beziffern. Zur Identifizierung ihrer DC-Signale nutzten sie massenspektrometrische Methoden und konnten die Identität von 5hmdC damit zweifelsfrei klären.[2, 5] Darüber hinaus wurde per DC gezeigt, dass 5hmdC im Gehirn im Vergleich zu diversen Organen angereichert ist.[2] Aufgrund mangelnder Sensitivität und Genauigkeit dieser Untersuchungsmethoden entwickelten *Carell* und Mitarbeiter eine quantitative Isotopen-basierte HPLC-MS-Methode und untersuchten die Verteilung von 5mdC und 5hmdC in der Maus.[3-4] Im Gegensatz zu den globalen Werten von 5mdC, welche einen durchschnittlichen Wert von 4.3±0.2% pro dG in allen untersuchten Geweben aufwiesen,[3] wurde für 5hmdC eine große Varianz zwischen den Geweben festgestellt (0.03–0.7% pro dG, siehe **Abbildung 6**).
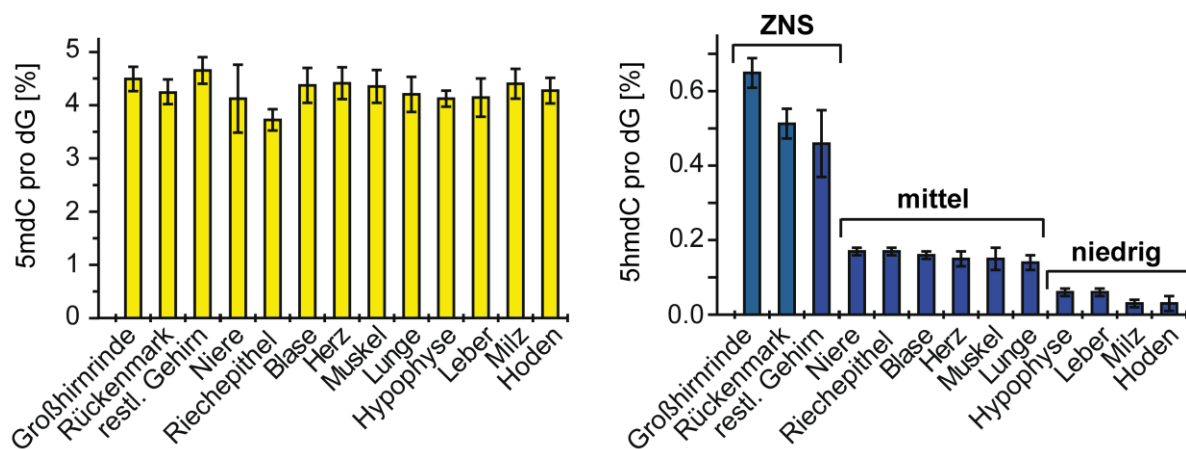
Abbildung 6 | Globale Verteilung von 5mdC (gelb) und 5hmdC (blau) in verschiedenen Mäusegeweben.[165] ZNS = zentrales Nervensystem.

Während z.B. in der Leber und im Hoden die Werte sehr gering waren (0.03–0.06%), fand man z.B. im Herzen und in der Niere mittlere Werte (0.16%). Im zentralen Nervensystem wurden in Übereinstimmung mit *Kriaucionis* und *Heintz* interessanterweise die höchsten 5hmdC-Level innerhalb der Mausgewebe gefunden,[2-4] die je nach Gehirnareal ebenfalls eine hohe Variabilität aufwiesen (0.3–0.7% pro dG).[3] So enthielt die Retina oder das Kleinhirn nur 0.3%, wohingegen bis zu 0.7% pro dG in der Mäuse-Großhirnrinde (Cortex) oder im Hypothalamus gefunden wurden.[3, 66] Da es sich bei diesen Werten um Mittelwerte aller

Zellen einer Gewebepräparation handelt, wurden immunohistologische Färbungen von Gewebeschnitten zur Visualisierung 5hmdC-positiver Zellen durchgeführt. Danach wiesen Neuronen im Vergleich zu umliegenden Zellen wie Gliazellen oder neuralen Stammzellen höchste Färbungsintensitäten auf. Dies macht deutlich, dass der 5hmdC-Gehalt in Neuronen noch höher sein muss, als die Gewebe-Mittelwerte der LC-MS-Messungen wiederspiegeln.[3] Im Menschen wurde der bisher höchste genomische 5hmdC-Gehalt gemessen: Dieser betrug in Gewebeschnitten des humanen Cortex (graue Substanz) 1.17% pro dG der fast doppelt so hoch wie der Vergleichswert der cerebralen weißen Substanz (0.7%) war. Mit einem durchschnittlichen 5mdC-Gehalt von 4.9% pro dG nahm 5hmdC mit ~20% einen hohen Anteil an der gesamten DNA-Modifizierung im humanen Cortex ein.[166] Die hohe Varianz der 5hmdC-Verteilung in Gewebezellen, die als Besonderheit im Vergleich zu 5mdC hervorzuheben ist, und der hohe Anteil von 5hmdC in neuronalen Geweben deuteten auf einen Zusammenhang mit Zelltyp-spezifischer Genregulation bzw. mit neuronaler Funktion hin (siehe Abschnitt 1.4.3).[3-4, 107]

Genomweite Lokalisations- und basenaufgelöste DNA-Sequenzierstudien zeigten, dass 5hmdC in Promotoren und Transkriptionsstartpunkten, aber auch im Genkörper angereichert ist.[66, 141, 167-174] Dabei residiert 5hmdC fast ausschließlich (>99%) in CpG-Dinukleotiden, obwohl in ES-Zellen und Neuronen 5mdC zu großen Teilen (bis zu 53%) auch in CpH-Dinukleotiden (H = A, T, C) vorkommt.[66, 167, 174] Dies könnte sich einerseits damit erklären lassen, dass 5mCpG in *in vitro*-Studien das bessere Substrat von Tet-Enzymen ist (siehe Abschnitt 1.4.1).[128-129] Da in diesen Studien kein dementsprechend deutlicher Unterschied in den Oxidationskinetiken der verschiedenen Dinukleotide vorlag, kann man spekulieren, dass 5hmCpH eventuell schneller umgesetzt wird. Des weiteren weiß man, dass die Tet-Enzyme spezifisch an CpG-reichen Stellen lokalisiert sind, wofür möglicherweise ihre CXXC-Domänen verantwortlich sind.[140-143, 175-176]

1.4.3 Biologische Funktionen von Tet und 5hmdC

Die Rolle von Tet und 5hmdC während der Embryonalentwicklung

Wie in Abschnitt 1.3.3 beschrieben wurde, ist die DNA-Demethylierung während der zellulären Entwicklung dynamisch. Insbesondere werden nach der Befruchtung der Eizelle die beiden parentalen Genome zunächst fast vollständig demethyliert, wobei die Demethylierung des männlichen Pronukleus schneller verläuft als die des Weiblichen. Bei den aktiven Anteilen dieser Demethylierungsprozesse handelt es sich tatsächlich um eine Oxidation von 5mdC durch Tet3. Die resultierenden 5hmdC-Level erreichen zwischenzeitlich einen

Höhepunkt und werden durch die ersten Zellteilungen dann hauptsächlich passiv reduziert.[91-95] Die Level von 5hmdC sind während der zellulären Entwicklung somit ebenfalls dynamisch reguliert und hängen von der Expression und Lokalisation der Tet-Proteine ab.[177-180] Hierbei spielen ferner auch "Schutzmechanismen" eine wesentliche Rolle; so ist das weibliche Genom in Zygoten durch das Protein Stella größtenteils vor der Tet3-Oxidation geschützt.[94, 181] Unklar ist bisher, warum die DNA-Demethylierung des männlichen Genoms schneller verlaufen muss. Ein möglicher Grund könnte sein, dass Spermien einen etwa doppelt so hohen 5mdC-Gehalt haben als Oozyten, weshalb neben der passiven Verdünnung durch Zellteilung ein aktiver Prozess vorhanden sein muss, damit beide Genome zum gleichen Zeitpunkt, also vor der Implantation, in gleichem Maße hypomethyliert sind.[84, 182] Im 2-Zell-Stadium verringert sich dann die Menge von Tet3 drastisch und die Expression von Tet1 und Tet2 nimmt zu. Diese erreicht im Blastozysten-Stadium ihren Höhepunkt und wird durch den Pluripotenzfaktor Oct4 positiv reguliert. Tet1 und Tet2 sind deshalb direkt mit dem Pluripotenz-Netzwerk von ES-Zellen verknüpft. Wenn sich die ES-Zellen dann ausdifferenzieren werden die Pluripotenz-assoziierten Gene methyliert und die Proteinlevel von Tet1 und Tet2 sinken entsprechend wieder, während Tet3 erneut hochreguliert wird.[177-180]

Die Tet-Proteine und die damit verbundene DNA-Modifizierung bzw. DNA-Demethylierung sind daher eng mit der Reprogrammierung des Epigenoms und einer normalen Entwicklung verknüpft.[91, 141-142, 183-186] *Tet1*[-/-] oder *Tet2*[-/-] Einzel-*Knockout*-Mäuse sowie *Tet1*[-/-]/*Tet2*[-/-]-Doppel-*Knockout*-Mäuse sind zwar überlebensfähig, diese wiesen jedoch anormale Phänotypen, Wachstumsdefekte und eine erhöhte pränatale Letalität auf.[184-186] Dagegen sind *Tet3*[-/-]-*Knockout*-Mäuse nicht überlebensfähig und sterben bereits pränatal, da Tet3 für die frühe globale Demethylierung des männlichen Genoms in der Zygote verantwortlich ist.[91]

Alle Tet-Proteine sind stark an Promotoren angereichert, insbesondere an CpG-reichen Promotoren.[140-143, 175-176] Dabei überlappt die Verteilung von Tet-Proteinen im Genom aber nicht nahtlos mit der Präsenz oxidierter 5mdCs.[187] Es konnte gezeigt werden, dass Tet-Enzyme duale Funktionen haben, d.h. sie erfüllen Regulationsaufgaben auch unabhängig von ihrer katalytischen Funktion.[141-142, 188-189] Eindrucksvoll konnte dies gezeigt werden, indem nicht nur die Überexpression von Tet1, sondern auch die einer katalytisch inaktiven Variante zu einer Hochregulierung von Gedächtnis-assoziierten Genen führte.[188] Tet1 besitzt in mES-Zellen neben einer aktivierenden Funktion von Pluripotenzgenen eine repressive Funktion in der Transkriptionskontrolle von Polycomb-assoziierten Genen,[141-143] während Tet2 generell

transkriptionsaktivierend wirkt.[175] In mES-Zellen rekrutiert Tet1 hierfür den 5hmdC-bindenden Mbd3/Nurd- (Nukleosom-Remodelling und Deacetylase)[190] sowie den Sin3a-Komplex,[141, 176, 191-192] wohingegen Tet2 keine Interaktion zu diesen repressiv wirkenden Proteinen herstellt.[176] Als weiteren Proteininteraktionspartner aller Tet-Proteine kennt man Ogt (*O-linked* β-D-*N*-Acetylglucosamin-Transferase), welches Tet und Histone an Transkriptionsstartpunkten glykosyliert und darüber hinaus auch zu einer Methylierung von H3K4 führt.[175-176, 191, 193-199]

Detaillierte Übersichtsartikel über die biologischen Rollen von Tet-Proteinen in der DNA-Demethylierung und Transkriptionskontrolle, speziell während der Embryonalentwicklung, wurden kürzlich von *Anjana Rao*, *Yi Zhang* und deren Mitarbeitern veröffentlicht.[84, 179, 187] Im Detail aufzuklären bleibt die Rolle von 5hmdC nicht nur bei der DNA-Demethylierung (siehe Abschnitt 1.6), sondern auch bei der Transkriptionskontrolle, da für 5hmdC Hinweise für die Aktivierung und Repression der Transkription je nach Gen-Lokalisation gefunden wurden.[141, 200-207] In mES-Zellen wurde berichtet, dass Tet1 hauptsächlich innerhalb von Promotoren bzw. an Transkriptionsstartpunkten 5mdC oxidiert, aber auch im Genkörper, während Tet2 nur im Genkörper oxidiert.[204] War 5hmdC dementsprechend in Promotoren und/oder Transkriptionsstartpunkten angereichert, so korrelierte dies negativ mit der Transkription, wohingegen 5hmdC im Genkörper positiv korrelierte. Diese Korrelation wurde nicht nur in ES-Zellen, sondern auch in entstehenden und ausdifferenzierten Gehirnzellen gefunden[202-204, 206] Insbesondere ist dabei von Interesse, welche Proteine spezifisch mit 5hmdC wechselwirken oder abgestoßen werden[14, 190, 201, 207-209] (siehe Abschnitt 1.4.4 und 3.5) und wie die zugrundeliegenden Mechanismen ablaufen.

Die Rolle von Tet und 5hmdC während der Gehirnentwicklung

5hmdC erreicht im Gehirn von Mensch und Maus die höchsten Level im Vergleich zum Rest des Körpers (siehe Abschnitt 1.4.2). Neben einer hohen Varianz dieser Level in unterschiedlichen Gehirnarealen, wurde ferner beobachtet, dass diese während der Gehirn-Entwicklung bzw. Alterung prä- und postnatal zunehmen.[4, 66, 161, 171, 210] Im menschlichen Cortex erreicht die altersbedingte Zunahme von 5hmdC mit etwa 20 Jahren ihr Maximum und ist auf einem Niveau von 1.1–1.2% pro dG bis ins hohe Alter stabil.[165]

Die altersbedingte Zunahme von 5hmdC im Mäusehirn geht mit einer Expressionssteigerung von *Tet2* und *Tet3* einher.[203, 211] Bei der Entwicklung von Neuronen reichert sich 5hmdC gerade in Genkörpern an, wodurch die Expression von Genen mit neuronalen Funktionen hochreguliert wird.[201-203] Die Zunahme von 5hmdC wird in diesen Genkörpern gleichzeitig

von der Demethylierung von H3K27me3 begleitet, wodurch sich die Chromatin-Zugänglichkeit erhöht. Ein *Knockdown* von Tet2/3 im embryonalen Cortex bewirkte neuronale Entwicklungsdefizite.[203] Sowohl eine intakte CXXC-Domäne als auch die katalytische Aktivität von Tet3 ist hierbei essenziell.[145] Im Gegensatz zu Tet2 und Tet3 ist Tet1 für eine normale Gehirnentwicklung zwar nicht notwendig, allerdings haben Tet1-defiziente Mäuse ausgeprägte Lern- und Gedächtnisschwierigkeiten.[212] Tet1 schützt Zielpromotoren vor Hypermethylierung und erhält so die Expression zahlreicher Gene, die Gedächtnis-assoziiert sind.[212-213] Interessanterweise ist die Beteiligung von Tet1 an der Gehirnfunktion zum Teil unabhängig von dessen katalytischer Aktivität. Tet1 erfüllt hier eine duale Funktion, wie zuvor bereits erwähnt wurde.[188]

Ist 5hmdC ein stabiler epigenetischer Marker?

Die hohe Abundanz von 5hmdC im Gehirn, dessen Zellen sich im Vergleich zu den übrigen Geweben kaum teilen, sowie die beobachtete altersbedingte Zunahme, implizierten früh, dass 5hmdC nicht nur ein transientes Intermediat der DNA-Demethylierung sein könnte, sondern im ausgewachsenen Individuum auch ein stabiler epigenetischer Marker sein müsste.[2-4, 66, 165-166, 171, 210] Die basenaufgelöste Sequenzierung von Genen, die bei der neuronalen Entwicklung mit 5hmdC angereichert werden, lieferte hier den ersten direkten Hinweis, da man neben der Oxidation von 5mdC in den meisten Fällen keine Reversion zu dC feststellen konnte.[203]

Per Definition ist eine Chromatin-Modifikation epigenetisch, wenn sie neben der Beteiligung an der Genregulation bei der Zellteilung vererbt werden kann (siehe Abschnitt 1.1 und 1.2). So ist 5mdC epigenetisch, da Dnmt1/Uhrf1 nach jeder Zellteilung hemimethylierte CpG-Dyaden (kurz M/C) erkennt, methyliert und damit die symmetrische Bismethylierung (M/M) instand hält (**Abbildung 7**).[100-102] In Abwesenheit dieses Enzymkomplexes wird das Genom passiv demethyliert, genauer gesagt der Gehalt von 5mdC wird mit jeder Zellteilung verdünnt (**Abbildung 7**). Dies geschieht z.B. mit dem mütterlichen Genom nach der Befruchtung der Eizelle (siehe Abschnitt 1.3.3) durch aktiven Ausschluss des Dnmt1/Uhrf1-Komplexes.[84, 214] In gleicher Weise wird 5hmdC bei der Teilung von Zygoten verdünnt,[95] allerdings muss Dnmt1/Uhrf1 nicht aktiv von hemimodifizierten 5hmCpG-Dyaden (H/C) ferngehalten werden, da diese kein geeignetes Substrat darstellen.[209, 215]
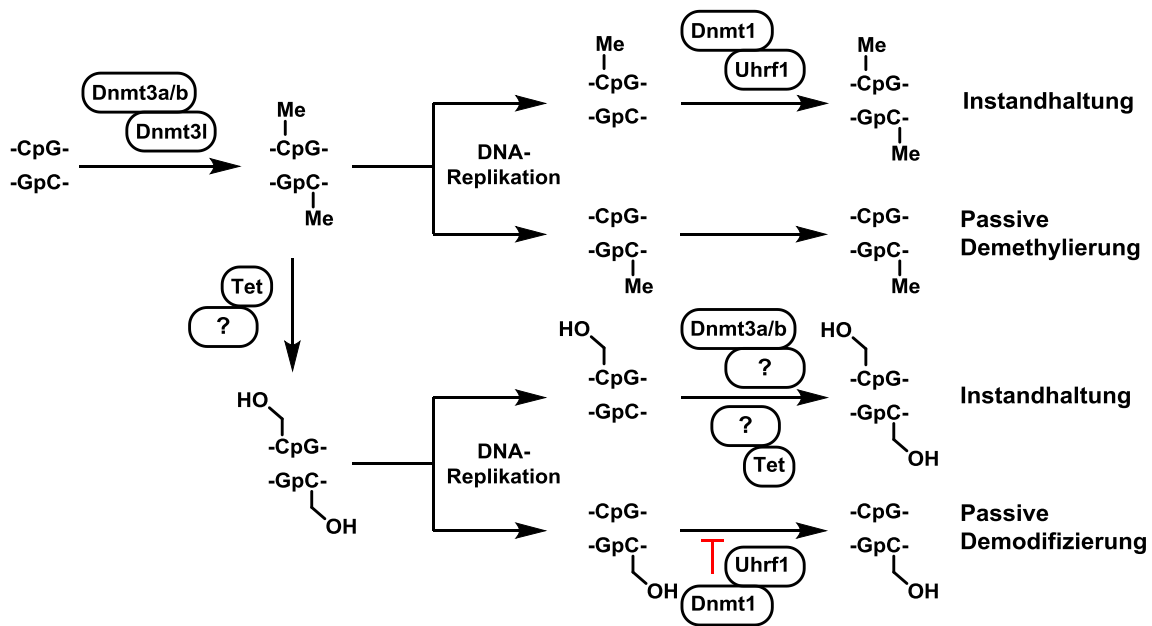
Abbildung 7 | Passive Demodifizierung sowie Instandhaltungsmechanismen des Modifizierungsmusters von CpG-Dyaden nach erfolgter Zellteilung.

Otani et al. konnten kürzlich den Nachweis erbringen, dass 5hmdC in mES-Zellen nach der DNA-Replikation aus der *de novo* Methylierung durch Dnmt3a/3b nachgebildet wird.[215] Für die Instandhaltung von H/H-Dyaden in mES-Zellen konnte ein Zusammenspiel aus der Methylierung von H/C-Dyaden durch DNMT3a/3b und anschließende Oxidation durch Tet1 gezeigt werden (**Abbildung 7**).

Per Isotopenverfolgung konnte kürzlich final bewiesen werden, dass 5hmdC im Mausgewebe sowie in mES-Zellen größtenteils eine stabile Modifikation ist und mit einer deutlichen Verzögerung nach der S-Phase (DNA-Replikation und -Methylierung) gebildet wird. Es wurde eine negative Korrelation zwischen der Höhe des 5hmdC-Gehaltes und der Proliferationsrate des jeweiligen Gewebes gefunden. Demnach war der 5hmdC-Gehalt in schnell proliferierenden Geweben wie der Milz gering, da hier 5hmdC laufend passiv verdünnt wird, wohingegen die Werte in langsam proliferierenden Geweben wie dem Herzen hoch waren. Die sehr hohen 5hmdC-Werte im Gehirn lagen sogar außerhalb dieser Korrelation, da sich die Zellen des Gehirns im Durchschnitt kaum teilen.[216]

Zusammenfassend demonstrierten die Untersuchungen, dass 5hmdC nun neben 5mdC als ein zusätzlicher stabiler epigenetischer Marker etabliert ist, der mit Hilfe von Dnmt3a/3b an die Tochterzellen vererbt werden kann.[203, 215-216]

1.4.4 Therapeutische Relevanz

Fehlerhafte DNA-Methylierungsmuster wurden mit einer Vielzahl von schwerwiegenden Krankheiten wie Krebs, Autoimmunerkrankungen, Diabetes und neuronalen Entwicklungsstörungen in Verbindung gebracht.[18, 24, 76, 217-225] So wurde eine anormale *de novo* Methylierung von CpG-Inseln von Tumorsuppressorgenen in vielen Krebszellen beobachtet, die zu aberranter Zellproliferation führte.[23] Involviert sind hier Gene fundamentaler Prozesse wie der DNA-Reparatur, der Zellzykluskontrolle, des p53-Netzwerks, der Ras- oder Vitamin-Signaltransduktion und der Apoptose.[24] Therapeutisch besteht daher ein hohes Interesse mit Hilfe von spezifischen Methyltransferase-Inhibitoren die zugrundeliegende DNA-Hypermethylierung rückgängig zu machen und die Reaktivierung dieser Gene zu bewirken.[226] Neben der Hypermethylierung von Tumorsuppressorgenen, wird in vielen Krebsarten auch die Hypomethylierung von zahlreichen sogenannten Onkogenen beobachtet.[217, 221-222, 224]

Derzeit verstärken sich die Hinweise, dass nicht nur eine fehlerhafte DNA-Methylierung sondern auch eine fehlerhafte DNA-Hydroxymethylierung mit der Pathogenese von Krebs zusammenhängt. Verdeutlicht wird dies durch eine Studie aus dem Arbeitskreis *Carell*, in der globale 5hmdC-Level in pathologischen Gehirngeweben von Tumorpatienten quantifiziert wurde.[166] Im gesunden humanen Cortex lag der 5hmdC-Spiegel bei durchschnittlich 1.17% pro dG. In Tumorgeweben war der 5hmdC-Gehalt dagegen mit durchschnittlich 0.16% pro dG deutlich reduziert.[166] Auch andere Studien haben in zahlreichen entarteten Geweben oder Zellen reduzierte 5hmdC-Level gefunden: Darunter fallen Bauchspeicheldrüsen-, Brust-, Darm-, Haut-, Leber-, Lungen- und Prostata-Krebs sowie Leukämie. 5hmdC könnte sich deshalb als ein genereller diagnostischer Krebsmarker (Biomarker) eignen.[227-237]

Als Ursache für reduzierte 5hmdC-Level kann man einerseits anführen, dass sich Krebsgewebe durch höhere Proliferationsraten auszeichnet, die eine passiven Verdünnung von 5hmdC zur Folge haben.[166, 216] Man findet als weitere Ursache in den entarteten Geweben aber häufig Mutationen, Deletionen oder reduzierte Expressionslevel von *Tet*-Genen, weswegen man *Tet*-Gene gewissermaßen zu den Tumorsurpressorgenen zählen kann.[126, 128, 131, 178, 186, 229-231, 234, 238-242] Interessanterweise hat man solche Veränderungen bisher nur bei Tet1 und Tet2 gefunden. Tet3 ist vermutlich der wichtigste Regulator unter den Tet-Proteinen und pathologische Veränderungen würden zu erhöhter Letalität führen.[131] Als weitere Ursache reduzierter 5hmdC-Level ist bekannt, dass Tet-Enzyme in Krebszellen durch die Anreicherung natürlicher Metabolite inhibiert werden. Die Isocitrat-Dehydrogenasen (Idh)

1 und 2 wandeln Isocitrat in 2-Oxoglutarat um, welches Tet-Enzyme und andere Dioxygenasen als Co-Substrat benötigen. In manchen Krebszellen weisen mutierte Idhs eine neue Enzymaktivität auf und reduzieren 2-Oxoglutarat zu 2-Hydroxyglutarat,[243-244] das als Onkometabolit alle 2-Oxoglutarat-abhänigen Dioxygenasen kompetitiv inhibiert.[166, 232, 237] In ähnlicher Weise werden Dioxygenasen durch eine krankhafte Anreicherung von Succinat oder Fumarat inhibiert.[112, 245]

Neben Störungen in der Synthese von 5hmdC, spielen auch Mutationen in den Proteinen eine Rolle, die an der Signalvermittlung zwischen Epigenom und Proteom beteiligt sind. Eine wichtige Rolle bei der Gehirnentwicklung und -funktion spielt u.a. das Methyl-CpG-bindende Protein MeCP2. Mutationen im menschlichen MeCP2 verursachen das Rett-Syndrom, eine tiefgreifende neuronale Entwicklungsstörung.[246] Kürzlich wurde MeCP2 auch als 5hmdC-Binder identifiziert. Obwohl die Affinität zu 5hmdC im Vergleich zu 5mdC geringer ist,[14, 201, 207, 209] konnte demonstriert werden, dass diese Wechselwirkung dennoch von entscheidender Bedeutung ist.[201] Eine Rett-Syndrom verursachende Punktmutation in MeCP2 führte zu einem vollständigen Verlust der Affinität zu 5hmdC, während die Affinität zu 5mdC nahezu unbeeinflusst blieb.[201] Es scheint offensichtlich zu sein, dass die Regulierung der Level von 5mdC und 5hmdC essenziell für die Zellintegrität ist und dass Mutationen in den wechselwirkenden Proteinen zu einer Störung in der epigenetischen Regulation und damit zu einer Störung der Zellentwicklung bzw. Zellhomöostase führt. Eine systematische Untersuchung bei der Entstehung von Krankheiten ist gefordert, damit ein tieferes Verständnis über die vielfachen Wechselwirkungen zu neuen Therapieansätzen führt.

1.5 Epigenetisch relevante DNA-Schäden

Die Integrität des Genoms und des Epigenoms wird vielfach durch exogene und endogene Substanzen sowie durch hochfrequente elektromagnetische Strahlung verletzt, die Ursache von Zelltod, Erbkrankheiten und Krebs sein können. Kanzerogene Substanzen kann man in zwei Klassen, in genotoxische und nicht-genotoxische unterteilen. Erstere bewirken in einer Dosis-abhängigen Weise kovalente DNA-Schäden, die direkt zu Mutationen oder Chromosomenschäden führen; letztere verursachen die Kanzerogenese durch Veränderung der Expressionsmuster der Zelle, d.h. sie verändern das Epigenom. Diese Substanzen verursachen u.a. aberrante DNA-Methylierungsmuster, wie im vorangegangen Abschnitt beschrieben wurde.[247-248] Dementsprechend können diese Substanzen auch für aberrante DNA-Hydroxymethylierungsmuster verantwortlich sein. Eine besondere Stellung unter den kanzerogenen Substanzen nehmen hierbei die reaktiven Sauerstoffspezies ein (englisch:

reactive oxygen species; ROS), die sowohl das Genom als auch das Epigenom oxidativ schädigen können und deren Herkunft endogener wie exogener Natur ist. Potenzielle endogene Quellen sind die oxidative Phosphorylierung in Mitochondrien, der P450-Metabolismus, Peroxisomen und Entzündungsprozesse durch Immunabwehrzellen; exogene Quellen von ROS sind im allgemeinen Umweltgifte, also diverse Redox-Verbindungen, Schwermetalle und hochfrequente Strahlung.[247-249]

Oxidative DNA-Schäden sind von allen Nukleosiden (dA, dG, dT, dC und 5mdC) bekannt. Darunter ist 8-oxo-7,8-dihydroguanosin (8oxodG) am prominentesten, welches als ubiquitärer Marker für fast jede Art von zellulärem Stress dient.[249] Diese oxidativen Nukleobasen-Schäden können mutagen sein, aber auch indirekt z.B. die Dnmt1/Uhrf1-vermittelte Instandhaltung[100-102] von 5mdC blockieren und so zu einer Hypomethylierung des DNA-Abschnitts führen.[248] 5mdC kann ferner direkt durch Oxidationsreaktionen geschädigt werden, für die ROS[164, 250] sowie andere OH-Radikal-basierte Prozesse[251] oder foto-chemische Ein-Elektronen-Oxidationsreaktionen[163, 250, 252-254] verantwortlich sind. OH-Radikale werden z.B. durch Fenton-ähnliche Reaktionen (Fe(II)/Cu(II)) aus Wasserstoff-peroxid generiert, welches während des oxidativen Metabolismus in Mitochondrien gebildet wird (**Schema 9A**). OH-Radikale können mit 5mdC sowie mit dT auf zwei unterschiedliche Arten reagieren. Zum einen können sie Addukte mit den C5=C6-Doppelbindungen bilden und zum anderen die in **Schema 9B** gezeigte Abstraktion eines H-Atoms der Methylgruppe bewirken.[249, 256-257] Die Methylradikale können dann mit bimolekularem Sauerstoff unter Bildung eines Peroxy-Radikals weiterreagieren und in anschließenden Zerfallsreaktionen die entsprechenden Alkohole und Aldehyde ausbilden. So wurde in Modellstudien 5hmdC und 5-Formyl-2′-desoxycytidin (5fdC) als oxidativer Schaden von 5mdC sowie 5hmdU und 5fdU als oxidativer Schaden von dT gefunden. Unter foto-oxidativen Bedingungen konnte ferner *in vitro* die Bildung von 5-Carboxy-2′-desoxycytidin (5cadC) festgestellt werden, die jedoch aufgrund einer geringen Effizienz wohl kaum biologisch signifikant ist.[253, 256] Im Gegensatz zu den Oxidationsschäden von dT steht der Nachweis für die analogen 5mdC-Schäden in zellulärer DNA unter natürlichen Bedingungen allerdings noch aus.[249]

In genomischer DNA wurden die Level des prominenten DNA-Oxidationsmarkers 8oxodG am besten untersucht. Diese sind mit durchschnittlich $1-4\times10^{-6}$ pro dN stets höher im Vergleich zu anderen Oxidationsprodukten (z.B.: 8oxodG > 5fdU > 5hmdU).[258-260] Das europäische Standardkomitee für oxidative DNA-Schäden (ESCODD) einigte sich bzgl. 8oxodG auf einen Hintergrundwert von $\sim 1\times10^{-6}$ pro dN.[259] Höhere 8oxodG-Werte bzw.

Unterschiede in den publizierten Messergebnissen von 8oxodG und anderen Oxidations-produkte sind dabei oft auf Effekte der Probenpräparation zurückzuführen, da vor allem beim Zellaufschluss endogene ROS freigesetzt werden, die zu artifiziell erhöhten Messwerten führen. Mittlerweile existieren eine Reihe von Protokollen, die diese Artefakte zu reduzieren versuchen, beispielsweise durch Zusatz von Antioxidanzien beim Zellaufschluss[259, 261-263]
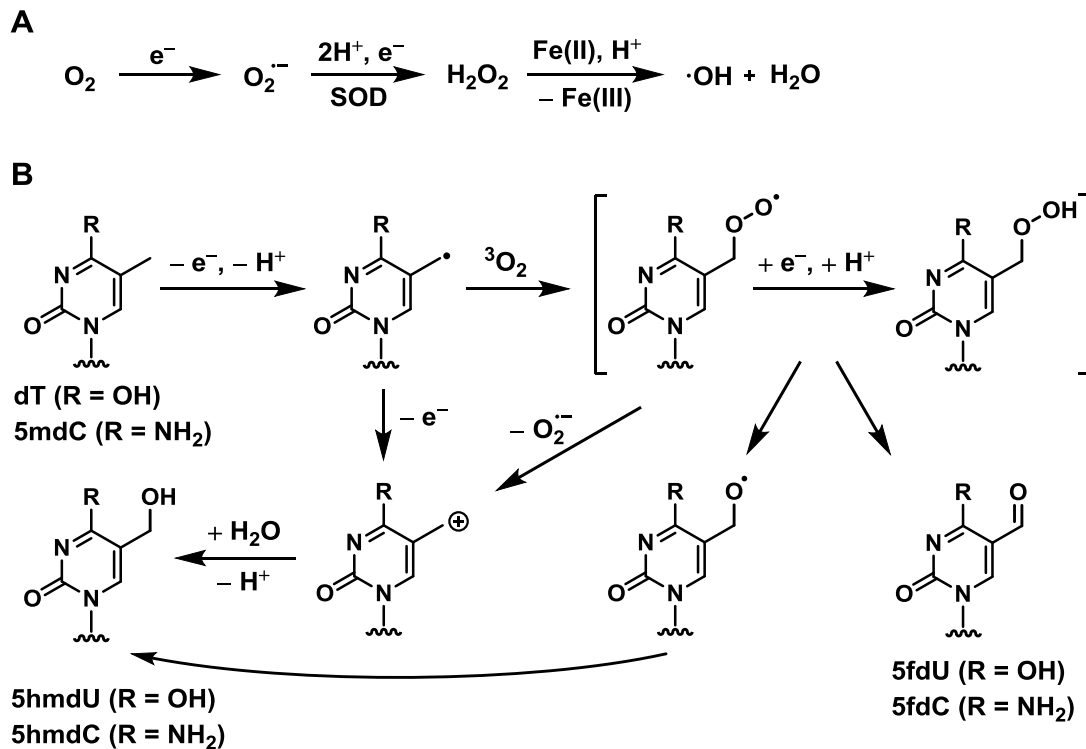
Schema 9 | Bildungswege epigenetisch relevanter DNA-Schäden durch reaktive Sauerstoff-spezies (engl.: *reactive oxygen species*, ROS). (**A**) ROS werden in Zellen durch exogene und endogene Prozesse erzeugt. SOD = Superoxiddismutase. (**B**) ROS verursachen unter anderem die Bildung von Methylradikalen von dT und 5mdC und sind deshalb für die Entstehung der DNA Oxidationsschäden 5hmdU, 5fdU beziehungsweise 5hmdC und 5fdC verantwortlich.[255] *Nota bene*: dT und seine Derivate (R = OH) liegen ausschließlich als (C=O)-N(H)-(C=O) Tautomer vor. Aus Gründen der Übersichtlichkeit wurde auf diese Darstellung verzichtet.

Genomisches 5hmdU und 5fdU kann nicht nur von der Oxidation von dT in genomischer DNA herrühren, sondern auch von der Oxidation von dT, dTMP und dTTP im Cytosol, bevor dieses per DNA-Replikation eingebaut wird.[248, 264] Obwohl 5hmdU und 5fdU in einer Watson-Crick-Basenpaarung mit dA nur geringfügig mutagen sind,[264-265] besitzen höhere Organismen eine signifikante Reparaturkapazität gegenüber diesen Oxidationsprodukten. Zellkulturen, deren Medien mit 5hmdU oder 5fdU versetzt wurden, starben in Folge der Inkorporation dieser Nukleoside in die DNA.[264, 266-270] Diese Sensitivität ging bei einer "5hmdU-Fütterung" vollständig verloren, wenn ein *Knockout* der Glykosylase Smug1 zuvor durchgeführt wurde.[266, 268] Die Toxizität von 5hmdU rührt in diesen Zellen also nur indirekt

von dessen Einbau in DNA gegenüber dA her, aber direkt von anschließender Basen-Exzisionsreparatur. Warum könnte es für den Organismus wichtig sein, 5hmdU und 5fdU zu beseitigen, wenn doch diese im Kontext einer dA-Basenpaarung nicht zwingend mutagen sind? Aus der Sicht von Protein-DNA-Interaktionen scheint es offensichtlich zu sein, dass eine Hydroxygruppe bzw. Formylgruppe von 5hmdU bzw. 5fdU, die in der großen Furche der DNA exponiert ist, die physikochemischen Wechselwirkungen zu Proteinen stark beeinflusst. In diversen Studien konnte in diesem Zusammenhang ein negativer Einfluss von 5hmdU auf Bindungsaffinitäten von Transkriptionsfaktoren oder TATA-Box/TBP-Komplexen gezeigt werden.[271-276] In Analogie hierzu demonstrierten *Valinluck et al.*, dass Oxidationsschäden von 5mCpG-Stellen wie 5hmdC und 8oxodG *in vitro* in der Lage waren die DNA-Protein-Interaktion zu dem Methyl-CpG-bindenden Protein MeCP2 zu stören. Wie bereits erwähnt ist MeCP2 u.a. für die Rekrutierung von Histon-Deacetylasen wichtig, um die Erhaltung von Heterochromatinstrukturen zu gewährleisten.[207] Des Weiteren kann die Anwesenheit von 5hmdC in Genkörpern die Transkription aktivieren, wie in Abschnitt 1.4.3 bereits beschrieben wurde.[202-204, 206] Zusammenfassend können also die Oxidationsschäden von dT und 5mdC den Chromatinstatus sowie die Expressionsmuster der Zelle aus dem Gleichgewicht bringen und die Krebsentstehung begünstigen.

Neben dieser klassischen Sicht, könnte sich die Zelle aber auch die Anwesenheit von Oxidationsschäden zu Nutze machen, denn das Leben in Abhängigkeit von Luftsauerstoff macht deren Vermeidung praktisch unmöglich.[261] In einigen kürzlich erschienen Arbeiten wurde gezeigt, dass 8oxodG womöglich die Rezeptor-vermittelte Genaktivierung von Östrogen-, Myc- und Hypoxie-Genen reguliert.[277-280] Es wurde ein Modell entwickelt, in dem die Bindung der entsprechenden Transkriptionsfaktoren die Rekrutierung der Histon-Demethylase LSD1 (Lysin-spezifische Demethylase) auslöst.[261, 281] Diese FAD-enthaltende Demethylase soll für die lokale Produktion von H_2O_2 während der Histondemethylierung verantwortlich sein, welches naheliegende Guanine in der Promotorregion oxidiert und damit die BER auslöst. 8oxodG wird durch die 8-Oxoguanin-Glykosylase (OGG1) ausgeschnitten, die gleichzeitig mit einer Topoisomerase rekrutiert wird. Über den resultierenden Einzelstrangbruch kann schließlich die kondensierte DNA-Region relaxiert und die Transkription aktiviert werden.[261, 277, 280-281] Passend zu diesem Zusammenhang wurde berichtet, dass 8oxodG speziell in transkriptionell aktivem Chromatin etwa 5mal im Vergleich zu Heterochromatin angereichert ist (4.4 bzw. 0.9×10^{-6} pro dN) und damit als epigenetisch relevant betrachtet werden könnte.[261] Auch aus einer anderen Perspektive wurden mehrere DNA-Modifikationen, darunter 8oxodG, dU und abasische Stellen, die Substrate von BER

sind, kürzlich als epigenetisch relevant eingestuft, da durch diese die Affinität des Transkriptionsfaktors CREB um mehr als zwei Größenordnungen ab- oder zunahm. CREB ist ein Vermittler, durch den mehr als 25% des Transkriptoms von Säugetieren reguliert wird. Durch die Anwesenheit von BER-Substraten bestünde für die Säugetierzelle die Möglichkeit den Grad der Transkriptionsaktivierung entsprechend fein zu justieren.[282] Da jedoch noch sehr wenig über die möglichen zugrundeliegenden Mechanismen bekannt ist, müssen weitere Forschungen in diese Richtung angestrebt werden, bevor eine epigenetische Relevanz dieser Modifikationen neben der klassischen, Krebs-verursachenden Eigenschaft etabliert ist.

1.6 Aktive DNA-Demethylierungsmechanismen

In einem komplexen Organismus, der sich der kovalenten Modifizierung der DNA zur Stilllegung von Genen bedient, müssen Mechanismen vorhanden sein, diese bei Bedarf wieder zu entfernen und Gene zu reaktivieren. Die Aufklärung der molekularen Mechanismen ist von höchstem Interesse, da ein detailliertes Verständnis die Chance bietet, die Steuerung der Zellentwicklung vollständig zu verstehen, sowie die Verbindung zwischen fehlerhafter Genexpression und diversen Krankheitsbildern im Detail aufzuklären. Insbesondere ist die gezielte Reprogrammierung von ausdifferenzierten Zellen in der regenerativen Medizin von großem Interesse.[76]

Nach der Befruchtung von Eizellen und zu bestimmten Zeitpunkten in der embryonalen Entwicklung werden große Teile des DNA-Methylierungsmusters sowohl aktiv als auch passiv entfernt und die Zellen somit reprogrammiert (siehe Abschnitt 1.3.3 und 1.4.3).[85, 88-90, 93, 96-98, 104-106] Im erwachsenen Individuum finden dagegen nur noch vereinzelte, Lokus-spezifische, aktive Demethylierungsprozesse statt, wie z.B. in Neuronen.[76, 283-288]

Auf der Suche nach den zugrundeliegenden Mechanismen durchlief man in der Vergangenheit einige Rückschläge (**Schema 10**): Während in Pflanzen 5-Methylcytosin durch spezielle Glykosylasen (Dme/Ros1) direkt ausgeschnitten und per Basenexzisionsreparatur (BER) in unmodifiziertes dC umgewandelt werden kann,[56, 289-292] wurde in Säugetieren keine entsprechende Aktivität bis dato gefunden. Die dabei untersuchte Thymin-DNA-Glykosylase (Tdg) und das Methyl-CpG-bindende Protein Mbd4 hatten eine ~30 fach geringere Aktivität im Vergleich zu T:G-Fehlpaarungen.[293-297] Schlussfolgernd wurde die gezielte Desaminierung von 5-Methylcytosin untersucht, die zur Bildung des besseren Substrats von Tdg und Mbd4 führen würde. Als Cytosin-Deaminasen kamen Aid (*activation-induced deaminase*), die Apobec-Familie (*apolipoprotein B pre-mRNA editing enzyme, catalytic*

polypeptide) und sogar Dnmts in Frage.[298-304] Tatsächlich konnte gezeigt werden, dass Aid/Mbd4 in Zebrafisch-Embryos für DNA-Demethylierungsprozesse verantwortlich ist.[303] Obwohl Aid und Apobec ferner auch in Oozyten, ES-Zellen und Urkeimzellen der Maus exprimiert werden,[300] sind sie hier eher nur für einen kleinen Teil der Demethylierung verantwortlich.[302] *Aid⁻/⁻*- und *Apobec⁻/⁻-Knockout*-Mäuse zeigen zudem keine Entwicklungs-störungen, sind lebensfähig und fruchtbar.[305-308] Parallel hierzu konnte gezeigt werden, dass Dnmts in der Lage sind 5mdC zu desaminieren.[298-299] Dies soll mechanistisch über eine Absättigung der C(5)=C(6)-Doppelbindung und anschließendem nukleophilen Angriff von Wasser auf die C(4)-Position geschehen. Anzuzweifeln ist bei dieser Studie jedoch, dass die Desaminierung nicht durch direkte Messverfahren beobachtet wurde und die Reaktion nur dann erfolgte, wenn geringe Konzentrationen an SAM vorlagen, die im biologischen Kontext irrelevant sind.[76, 299]
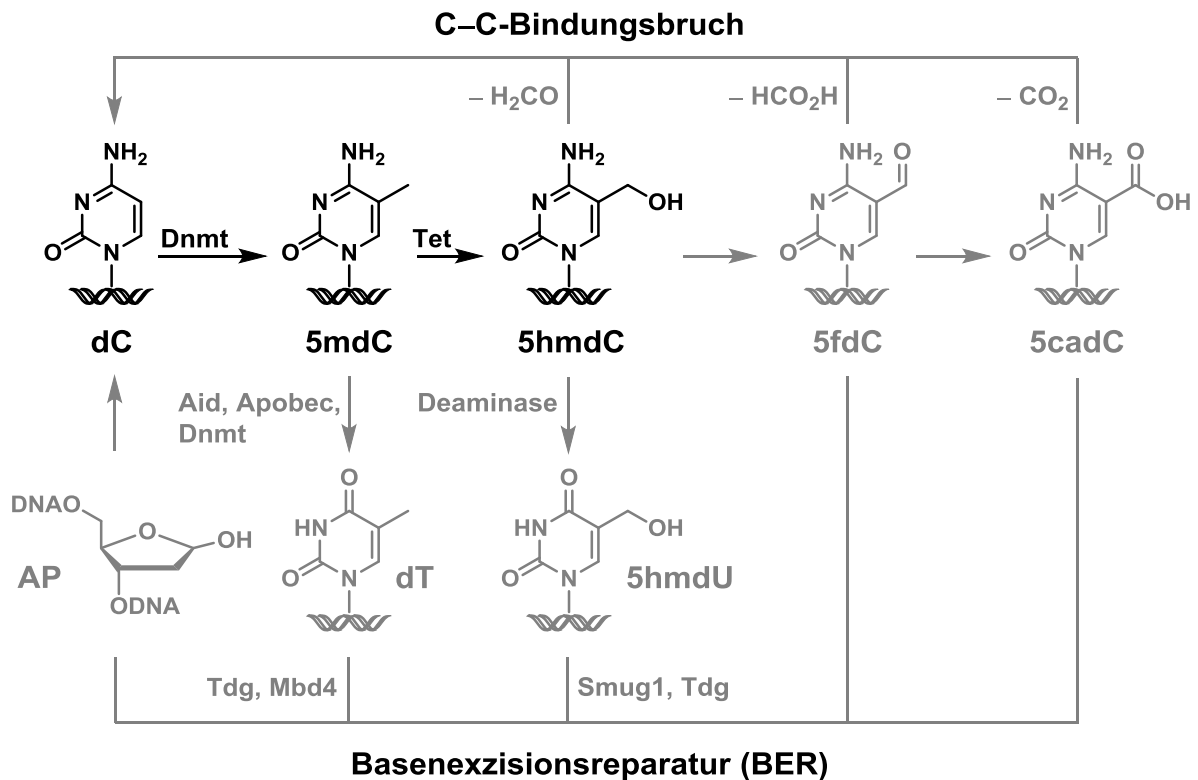
Schema 10 | Mögliche aktive DNA-Demethylierungswege (grau hinterlegt). AP = apyrimidinische(/apurinische) Stelle.

Eine hohe Durchschlagskraft hatte deswegen 2009 die Entdeckung von *Tahiliani et al.*, dass Tet-Enzyme die Methylgruppe von 5mdC effizient hydroxylieren können (siehe Abschnitt 1.4).[5] Sofort wurden die Möglichkeiten untersucht, wie 5hmdC als Intermediat weiter abgebaut werden könnte. Eine Möglichkeit ergäbe sich in Analogie zu Pilzen, wie *Neurospora crassa*, die Thymin demethylieren können, um Uracil zurückzugewinnen.[3, 76, 110]

Die verantwortliche Dioxygenase T7h oxidiert Thymin nicht nur zu 5hmU sondern weiter zu 5fU und 5caU.[109] Dementsprechend könnte Tet die iterative Oxidation von 5hmdC zu 5fdC und 5cadC bewerkstelligen (**Schema 10**). Alle drei Oxidationsprodukte könnten dann anschließend durch einen C(5)–C(exo)-Bindungsbruch direkt, ohne Beteiligung von Reparaturenzymen, demodifiziert werden. Dies wurde *in vitro* für 5hmdC bereits gezeigt. Eine mutierte bakterielle DNA-Methyltransferase (HhaI) war hierzu in der Lage, die über die typische transiente Absättigung der C(5)=C(6)-Doppelbindung Formaldehyd freisetzte.[309-310] In Analogie könnten dementsprechend auch 5fdC und 5cadC unter Freisetzung von Ameisensäure bzw. Kohlenstoffdioxid demodifiziert werden, wobei die Decarboxylierung von 5cadC chemisch gesehen am plausibelsten erscheint.[3, 76, 110] So wird in *Neurospora crassa* 5caU durch die Isoorotatdecarboxylase (Idc) decarboxyliert.[123-125]

Alternativ könnte 5hmdC, 5fdC und 5cadC auch über DNA-Reparaturmechanismen wie der BER entfernt werden (**Schema 10**). Da für 5hmdC keine Glykosylaseaktivität in Säugetieren gefunden wurde,[311-313] untersuchten *Guo et al.* den Zusammenhang der Desaminierung von 5hmdC zu 5hmdU durch Aid sowie Apobec und fand positive Hinweise für diesen Mechanismus im Gehirn.[313] Die anschließende Reparatur der Basenfehlpaarung soll dann durch die Glykosylase Smug1 oder Tdg eingeleitet werden.[312-313] Die Bedeutung von Tdg ist dabei durch die embryonale Letalität von entsprechenden *Knockout*-Mäusen untermauert.[312] Der vorgeschlagene Desaminierungsmechanimus von *Guo et al.* konnte allerdings nicht als etabliert angesehen werden, da ein endgültiger Beweis durch direkte Verfolgung der Reaktionsintermediate und Produkte beispielsweise durch Massenspektrometrie-basierte Methoden ausstand.

2010 suchten *Globisch et al.* in Mäusegeweben nach 5fdC, 5cadC und 5hmdU per Umkehrphasenchromatographie gekoppelt mit hochauflösender Massenspektrometrie (HPLC-HRMS). Jedoch konnte selbst im Gehirngewebe, welches sich durch höchste 5hmdC-Level auszeichnet, auch unter Verwendung größter DNA-Mengen keines dieser Intermediate detektiert werden. Dies wies zum einen auf die Stabilität von 5hmdC bzw. auf den transienten Charakter von 5fdC, 5cadC oder 5hmdU hin.[3] Demzufolge mussten in dieser Doktorarbeit sensitivere Methoden etabliert werden, um sich auf die Spur des DNA-Demethylierungs-mechanismus zu begeben.

2 Aufgabenstellung

2009 wurde 5hmdC – als sechste Base des Genoms – in genomischer DNA im Gehirn und in Stammzellen gefunden. Diese Modifikation wird durch Tet-vermittelte Oxidation der Methylgruppe von 5mdC postreplikativ gebildet.[2, 5] Zu Beginn dieser Arbeit im Frühjahr 2011 deutete sich bereits an, dass 5hmdC eine weitere wichtige Rolle bei der epigenetischen Kontrolle der Zellentwicklung und Genexpression spielt. Darüber hinaus spekulierte man, dass 5hmdC ein Reaktionsintermediat eines lange gesuchten Mechanismus der aktiven DNA-Demethylierung darstellt. Dieser Mechanismus soll für die Demodifizierung und damit zur Reaktivierung von Genen verantwortlich sein, die durch Methylierung stillgelegt wurden.[107]

Das erste Ziel dieser Arbeit bestand darin, weitere Intermediate bzw. DNA-Modifikationen dieses unbekannten Mechanismus in genomischer DNA ausfindig zu machen. Zu den vielversprechenden Kandidaten zählten die höher oxidierten Derivate 5fdC und 5cadC sowie 5hmdU als Desaminierungsprodukt von 5hmdC (**Schema 10**).[3] Da diese Modifikationen mit den bisher verwendeten massenspektrometrischen (MS) Methoden im Arbeitskreis Carell nicht detektierbar waren,[3] sollten geeignete Derivatisierungsreaktionen zur Anwendung kommen und eine empfindlichere MS-Technologie etabliert werden. Als "Goldstandard" für dieses spurenanalytische Projekt war die Etablierung einer Methode geplant, bei der die Ultrahochdruck-Flüssigkeitschromatographie mit einem Triple-Quadrupol-Massenspektro-meter (UHPLC-ESI-MS/MS) gekoppelt wird. Für die massenspektrometrische Quantifi-zierung der potenziellen DNA-Modifikationen sollten ferner isotopenmarkierte interne Standards auf organisch-chemischem Wege synthetisiert werden.

Als nächstes Projekt stand die Untersuchung eines möglichen DNA-Demethylierungs-mechanismus im Vordergrund, welcher über C−C-Bindungsbruchreaktionen von 5hmdC, 5fdC und 5cadC direkt unmodifiziertes dC wiederherstellen könnte. Hierfür sollten Isotopenverfolgungsexperimente im biologischen Milieu und chemische Modellsysteme entwickelt werden. Ferner sollte der Ursprung von genomischem 5hmdU quantitativ und durch metabolische Isotopenverfolgung untersucht werden. 5hmdU könnte sowohl per Desaminierung von 5hmdC als auch durch die Oxidation von dT entstehen.[3, 314] Da potentielle Demethylierungsmechanismen per Basenexzisionsreparatur (BER) ablaufen könnten,[105, 283, 315] war für den Nachweis von abasischen Stellen die Synthese eines Hydroxylamin-Reagenzes geplant. Dieses sollte unter milden Reaktionsbedingungen mit den abasischen Stellen ein stabiles Oxim bilden und hohe Detektionsempfindlichkeiten im Massenspektrometer erzielen.

3 Veröffentlichte Arbeiten

3.1 Die Entdeckung von 5-Formylcytosin in embryonaler Stammzell-DNA

Toni Pfaffeneder,[#] Benjamin Hackner,[#] Matthias Truß,[#] Martin Münzel, Markus Müller, Christian A. Deiml, Christian Hagemeier und Thomas Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012. *The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA*. ([#] geteilte Erstautorenschaft)

Prolog

Die Entdeckung von 5hmdC als Bestandteil des Genoms von Säugetieren war ein großer Meilenstein für die Erforschung der aktiven DNA-Demethylierung. Die Suche nach weiteren potentiellen Intermediaten blieb in einer früheren Studie ohne Erfolg.[3] Dieses Manuskript beschreibt nun die Entdeckung von 5fdC als natürlichem Bestandteil des Genoms von pluripotenten embryonalen Stammzellen der Maus. Diese Modifikation wird wahrscheinlich durch stufenweise Oxidation von 5mdC via 5hmdC gebildet und ihr Gehalt sinkt während der Zelldifferenzierung schneller als 5hmdC. 5fdC konnte für analytische Anwendungen mit einem Biotin-Hydroxylamin erfolgreich modifiziert werden. Die Entdeckung dieser Modifikation war richtungsweisend auf dem epigenetischen Gebiet. Kurz darauf wurde auch 5cadC durch zwei andere Gruppen entdeckt und dessen Bildung durch das Tet-Enzym demonstriert.[7-8] Ferner wurde die Derivatisierungsmethode von 5fdC mit Biotin-Hydroxylamin genutzt, um erste Informationen bezüglich dessen genomischer Lokalisation zu gewinnen.[316]
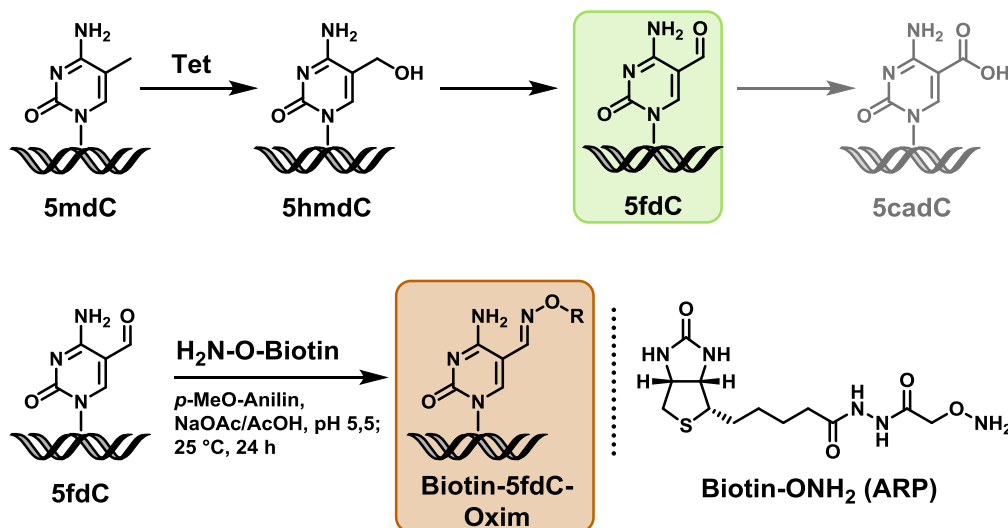
Autorenbeitrag

Die massenspektrometrische Identifizierung und Charakterisierung, sowie die chemische Derivatisierung von 5fdC in embryonaler Stammzell-DNA führte ich in enger Zusammenarbeit mit *Benjamin Hackner* durch. *Matthias Truss* (Charité Berlin) stellte die genomische DNA zur Verfügung. Die verwendete Referenzverbindung von 5fdC wurde von *Martin Münzel* im Zuge seiner Doktorarbeit hergestellt. Die verwendete LC-MS-Analytik beruhte im Wesentlichen auf Arbeiten von *Daniel Globisch*.[3-4]

Lizenz

Communications

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

Toni Pfaffeneder, Benjamin Hackner, Matthias Truß,* Martin Münzel, Markus Müller, Christian A. Deiml, Christian Hagemeier, 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).[1] The DNA methylome is significantly reprogrammed at various stages during early development,[2] during the development of primordial germ cells,[2c,3] or later in a locus-specific way at postdevelopmental stages.[4] 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.[5] The recent discovery that mC can be further oxidized to hydroxymethylcytosine (hmC) with the help of TET enzymes[6] has led to the idea that hmC is connected to epigenetic reprogramming,[7] maybe as an intermediate in an, as yet controversial, active demethylation process.[4,5,8] 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.[9] 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).[5a,10] Although such an oxidative active demethylation pathway with hmC as the starting point has been frequently postulated,[5a,10] none of the further oxidized bases (fC, caC) have so far been detected.[10a]

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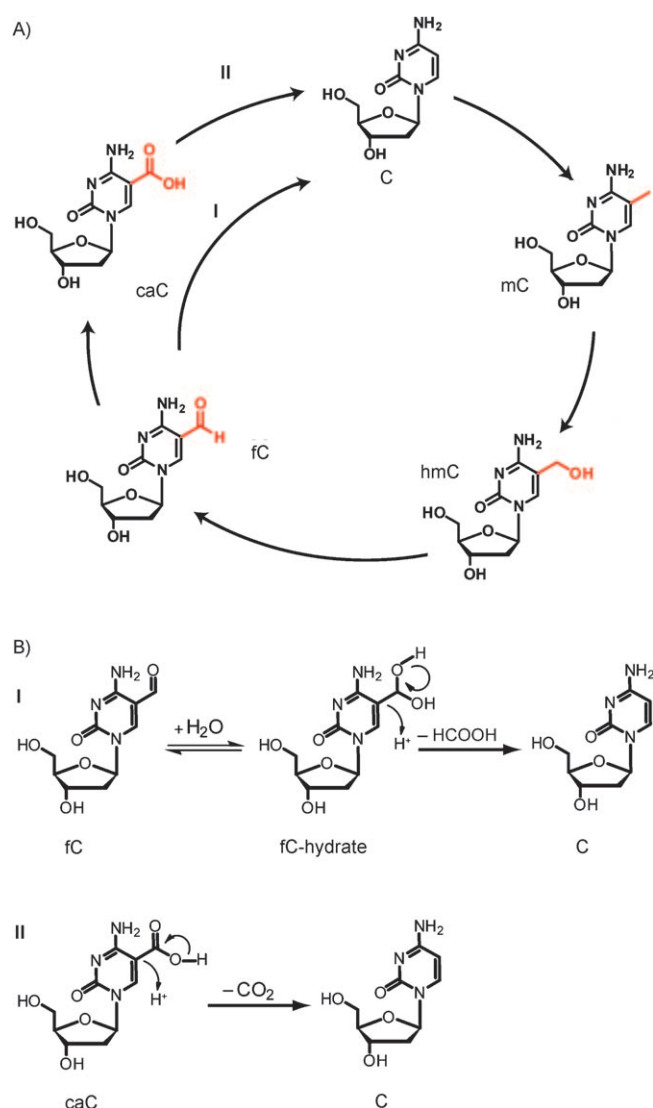
Scheme 1. A) Putative cycle of methylation and oxidative demethylation of cytidine derivatives. B) Details of the demethylation reaction via vinyl carbanions.[11] I: Deformylation of fC to C. II: decarboxylation of caC to C.

[*] M. Sc. T. Pfaffeneder,[+] M. Sc. B. Hackner,[+] 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,[+] Prof. Dr. C. Hagemeier
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.

© 2011 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Angew. Chem. Int. Ed. **2011**, *50*, 7008–7012

standards.[10a, 12] Specifically, we used DNA isolated from mouse embrionic stem cells and embyoid 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.[7d] 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,[12a] 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).[10a, 12a] 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^n studies. The unknown compound detected in the mES cell DNA had a high-resolution mass signal ($m/z_{found} = 256.0929$) which is in excellent agreement with the calculated exact mass for fC ($m/z_{calcd} = 256.0928$; 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,[10a] 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^n-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[13] and with the MS/MS data obtained from the authentic synthetic fC material (data not shown). In addition, the obtained MS^n data were identical with the unknown fragment ions detected in the mEB cell DNA. These data prove that the newly discovered compound
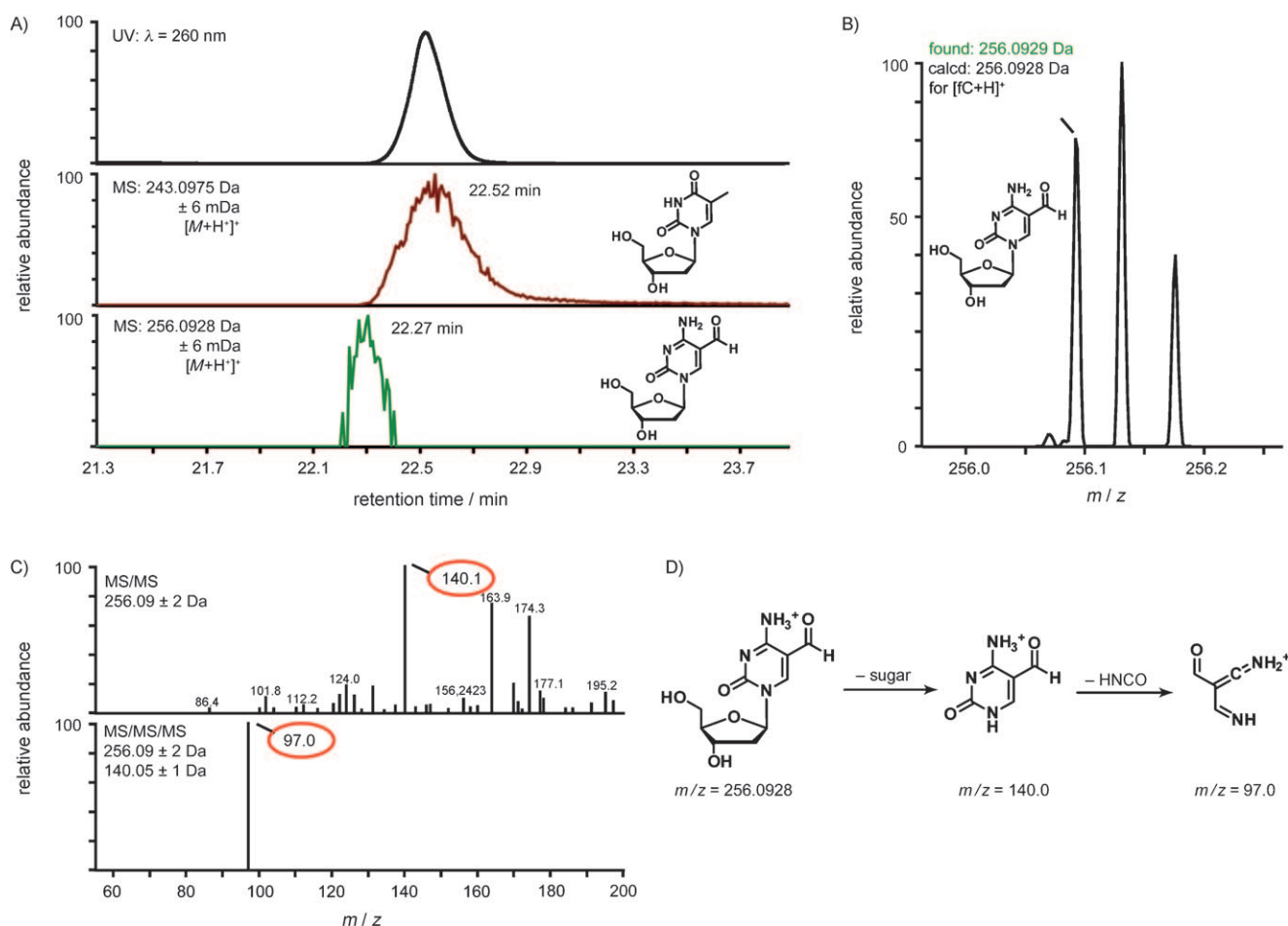
Figure 1. A) HPLC trace of digested mES cell DNA together with the MS signals from T and fC. The UV detection has a general delay of 0.2 min and was adjusted manually to the ion current. B) High-resolution mass data of fC. C) Fragment mass data from MS/MS and MS³ studies proving the structure of fC. D) Fragmentation pattern of fC in the MS² and MS³ 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
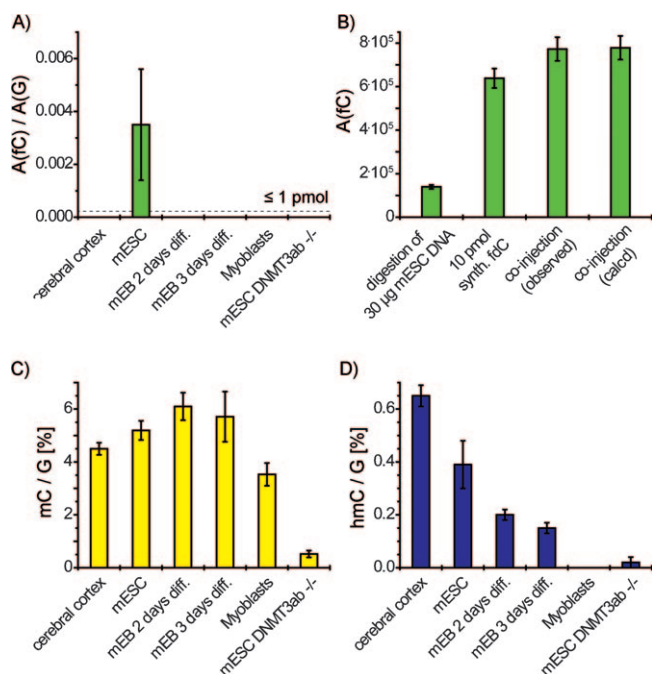
Figure 2. A) Correlation of the mass signal of fC and the UV signal of G in mES cell DNA, in DNA from nerve tissue, in DNA from mEBs after 2 and 3 days, in DNA from cultured myoblasts, and in DNA from mES lacking DNMT3a and -3b. The detection limit of fC was determined to be ≤ 1 pmol. B) Co-injection 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.[10a, 12a] 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.[10a, 12a] 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).[10a, 12a] 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^n 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.[14]), 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^n 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,[12a] 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.[15]

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[16] of fC-containing DNA fragments for sequencing[17] and more-sensitive detection of fC. Since formyl groups react selectively with hydroxylamines to give stable oxime derivatives,[18] we treated the mES DNA with the biotin-hydroxylamine reagent (Figure 3A).[19] After
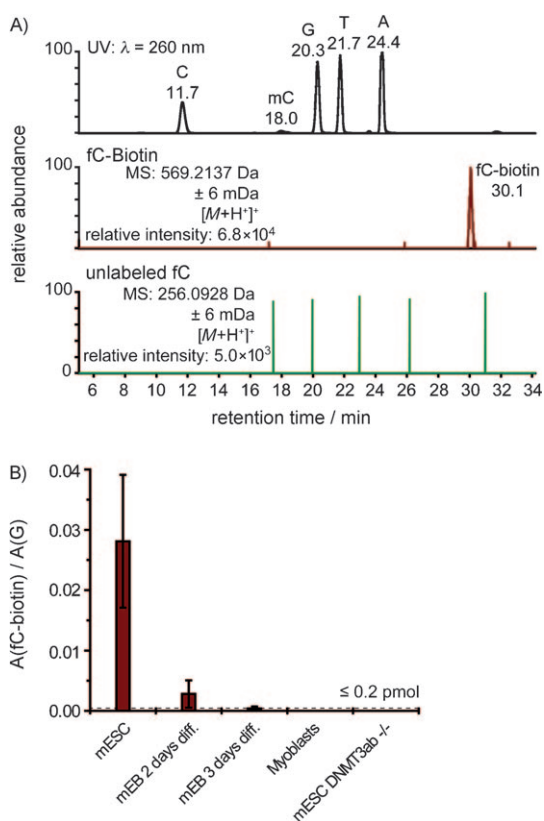
Figure 3. A) HPLC and MS signals of biotin-labeled fC obtained after treatment of embryonic stem cell DNA with the biotin hydroxylamine followed by DNA digest. The lowest trace shows only the background noise, specific signals for residual fC were not observed. B) Relative amount (mass area of fC-biotin/dG) 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 ≤ 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),[20] 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[21]). 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,[9] 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.

[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. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935; c) M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, DOI: 10.1002/anie.201101547.

[7] a) G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398–402; 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.-l. 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.

Communications

[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. Welzmiller, S. Michalakis, 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.

3.2 Verbesserte Synthese und Mutagenitätsstudien von 5hmdC-, 5fdC- und 5cadC-enthaltenden Oligodesoxynukleotiden

Martin Münzel, Ulrike Lischke, Dimitrios Stathis, <u>Toni Pfaffeneder</u>, Felix A. Gnerlich, Christian A. Deiml, Sandra C. Koch, Konstantin Karaghiosoff und Thomas Carell, *Chem. Eur. J.* **2011**, *17*, 13782–13788. *Improved Synthesis and Mutagenicity of Oligonucleotides Containing5-Hydroxy-methylcytosine, 5-Formylcytosine and 5-Carboxylcytosine.*

Prolog

Nachdem 5fdC und 5cadC als natürliche Bestandteile der DNA von Säugetieren gefunden wurden, bestand ein hoher Bedarf nach effizienten Synthesen von DNA-Strängen, die diese Modifikationen enthalten. Dieses Manuskript beschreibt die Synthese der entsprechenden Phosphoramidit-Bausteine via Pd(0)-katalysierter Funktionalisierung von 5-Iodo-2′-desoxy-cytidin. Mit Hilfe eines neu-entwickelten *Assays* konnte ferner gezeigt werden, dass 5fdC und 5cadC nur marginal mutagen sind, was eine wichtige Voraussetzung für ihre Beteiligung an epigenetischen Reprogrammierungsprozessen ist.

Autorenbeitrag

Die Hauptarbeit bei diesem Projekt lag bei *Martin Münzel*. Für dieses Manuskript führte ich die Synthese des 5fdC-Phosphoramidits und dessen Einbau in DNA durch.

Lizenz

Improved Synthesis and Mutagenicity of Oligonucleotides Containing 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxylcytosine

Martin Münzel, Ulrike Lischke, Dimitrios Stathis, Toni Pfaffeneder, Felix A. Gnerlich, Christian A. Deiml, Sandra C. Koch, Konstantin Karaghiosoff, and Thomas Carell*[a]

Abstract: 5-Formylcytosine (fC or [5-CHO]dC) and 5-carboxylcytosine (caC or [5-COOH]dC) have recently been identified as constituents of mammalian DNA. The nucleosides are formed from 5-methylcytosine (mC or [5-Me]dC) via 5-hydroxymethylcytosine (hmC or [5-HOMe]dC) and are possible intermediates of an active DNA demethylation process. Here we show efficient syntheses of phosphoramidites which enable the synthesis of DNA strands containing these cytosine modifications based on Pd[0]-catalyzed functionalization of 5-iododeoxycytidine. The first crystal structure of fC reveals the existence of an intramolecular H-bond between the exocyclic amine and the formyl group, which controls the conformation of the formyl substituent. Using a newly designed in vitro mutagenicity assay we show that fC and caC are only marginally mutagenic, which is a prerequisite for the bases to function as epigenetic control units.

Keywords: cytosine · DNA · DNA methylation · nucleosides · phosphoramidite

Introduction

Hydroxymethylcytosine (hmC) is a new DNA base that was detected in various tissues and in embryonic stem cells.[1] It is currently believed that this so-called sixth base of the genome is involved in epigenetic programming processes.[2] Hydroxymethylcytosine is generated from the epigenetic marker base 5-methylcytosine (mC) by hydroxylation of the C5-methyl group performed by ketoglutarate-dependent TET enzymes (TET1-3).[3] The presence of mC in promoter regions is considered to be a main signal that finally leads to the silencing of the corresponding gene.[4] Others and us speculated that the new base hmC is an intermediate in active demethylation pathways (Scheme 1) that allow cells to actively remove the C5-methyl groups of mC in the absence of replication.[1e,3,5] This removal would in principle enable cells to unblock silenced genes and therewith to dynamically modulate their genetic program.

Unblocking of genes via hmC can in principle occur in multiple ways (Scheme 1). First, hmC could be an early intermediate of an oxidative demethylation pathway in which hmC is further oxidized to 5-formylcytosine (fC) and 5-carboxylcytosine (caC). Demethylation is thereafter possible via either elimination of formic acid from fC or decarboxy-
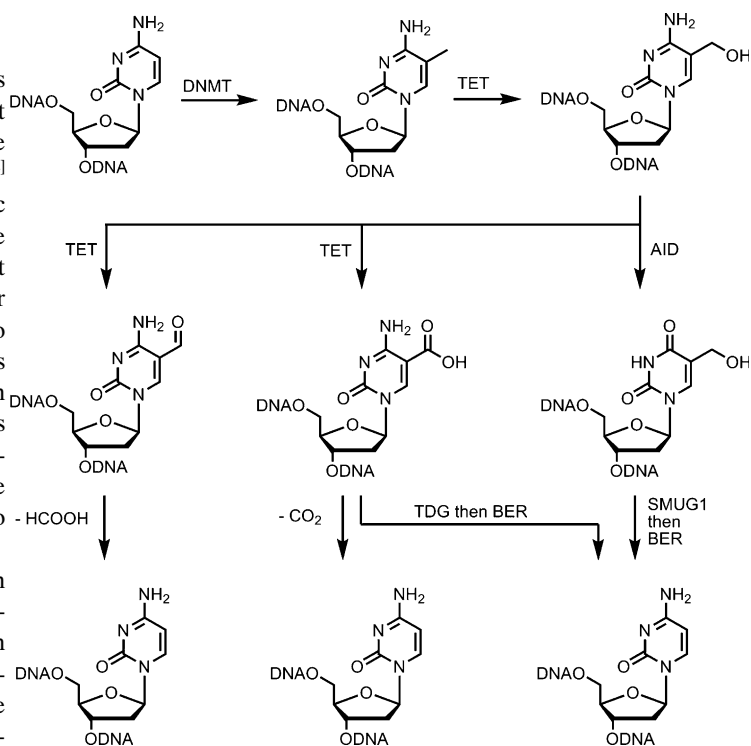
Scheme 1. Postreplicative modification of cytosine to hmC and putative active demethylation pathways. DNMT: DNA methyltransferase, TET: ten-eleven-translocation (methylcytosine dioxygenase), AID: activity induced cytosine deaminase, TDG: thymine DNA glycosylase, SMUG1: single-strand selective monofunctional uracil DNA glycosylase, BER: base excision repair.

[a] Dipl.-Chem. M. Münzel, Dipl.-Biol. U. Lischke, M. Sc. D. Stathis, M. Sc. T. Pfaffeneder, Dipl.-Chem. F. A. Gnerlich, Dipl.-Biochem. C. A. Deiml, M. Sc. S. C. Koch, Prof. Dr. K. Karaghiosoff, Prof. Dr. T. Carell Center for Integrated Protein Science (CiPS[M]) at the Department of Chemistry, Ludwig-Maximilians-University Butenandtstr. 5–13, 81377 Munich (Germany) E-mail: thomas.carell@cup.uni-muenchen.de

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.201102782.

lation of caC. Formic acid elimination is a known process that occurs during steroid biosynthesis,[6a] while decarboxylation at pyrimidine bases occurs during uridine biosynthesis

and in the pyrimidine salvage pathways of certain eukaryotes.[6a–d] Alternatively, it is possible that hmC or the further oxidized bases fC and caC are substrates of special DNA-repair glycosylases. These would generate abasic sites that are repaired by gap-filling with 2′-deoxycytidine (dC) using well established DNA repair pathways.[8a,9b]

Recently, others and us indeed discovered the oxidized nucleobases fC and caC in embryonic stem cells and in selected other tissues.[7–8] We could show that the fC levels decrease during stem cell development suggesting that fC is indeed involved in stem cell differentiation.[7] For fC, caC and hmC it was shown that these bases are substrates for repair enzymes. fC and caC are processed by the thymine-DNA glycosylase (TDG),[8a,9] while hmC was found to be converted to hydroxymethyluridine (hmU) by the enzyme cytosine deaminase (AID). hmU is removed from DNA by the glycosylase SMUG1 (single-strand selective monofunctional uracil DNA glycosylase).[10]

Currently the function of the new bases fC and caC is unknown and their role in a putative epigenetic active demethylation process is still unclear. Regarding their potential role as key epigenetic control intermediates during cell differentiation, the previously observed mutagenic potential of fC is problematic (Figure 1).[11] It was speculated that a putative strong intramolecular H-bond between the exocyclic N4 amino group and the carbonyl oxygen at C5 shifts the amino–imino equilibrium towards the mutagenic imino form, which would enable fC to form a wrong base pair with dA (Figure 1).[11–12] For caC mutagenicity data are so far not reported. The existence of a strong H-bond in caC is supported by the unusual low pK_a value of the C5-carboxylic acid of $pK_a < 1.5$.[13] In line with the electron-deficient character of the new bases fC and caC, the acidity of the corresponding protonated bases [HfC]$^+$ and [HcaC]$^+$ (protonation occurs at N3) is unusually high with $pK_{a[HfC]^+} = 2.4$ and $pK_{a[HcaC]^+} = 4.0$ (for comparison: $pK_{a[HC]^+} = 4.4$, $pK_{a[HhmC]^+} = 4.4$, and $pK_{a[HmC]^+} = 4.5$).[11b,13]
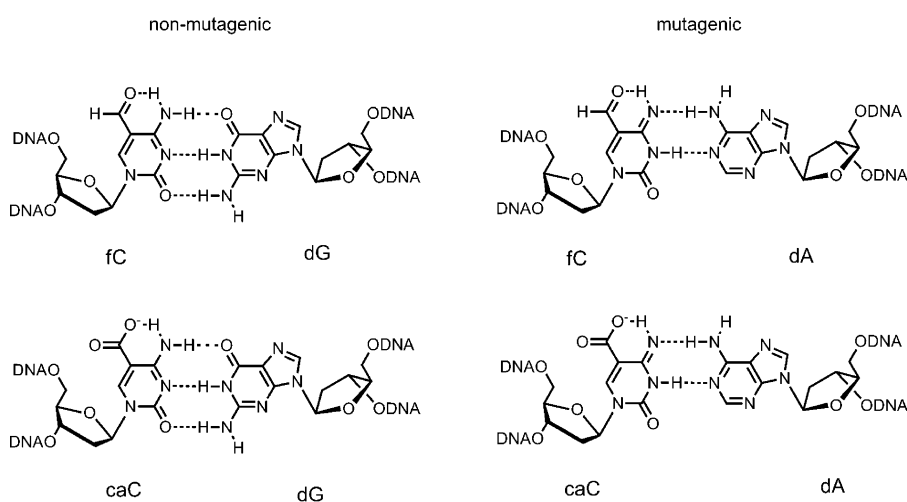
Results and Discussion

In order to investigate the mutagenic properties of caC and to re-examine the mutagenic potential of hmC and fC we prepared hmC, fC and caC oligonucleotides (Scheme 1).[14] Strands containing hmC were prepared as described by us before.[15] To access oligonucleotides containing fC in the required quantity and quality as needed for the current study we developed an improved fC phosphoramidite reagent using a carbonylative coupling strategy[1e,15a] with 5-iodo-2′-deoxycytidine (1) that was recently also employed by the He group to synthesize similar fC and caC phosphoramidites.[16] The improved fC phosphoramidite features a benzoyl protecting group at the N4-position (Scheme 2), which we found to be more stable than the acetyl protecting group. The protected fC building blocks give in our hands higher overall yields than the corresponding fC phosphoramidite without any protecting group at N4.[16] The direct incorporation of fC also avoids the oxidative diol cleavage that was developed by Karino et al. to generate fC in oligonucleotides from a diol-precursor phosphoramidite.[11b] The direct incorporation of an fC building block via phosphoramidite chemistry is also possible for RNA. Here, however, a DMF protecting group was chosen for N4.[17]

The caC phosphoramidite reagent was developed based on results of Sekine and co-workers.[13] We chose to protect the N4 amine groups with an acetyl group and we used a methyl ester to protect the C5 carboxylic acid (Scheme 2). Both protecting groups are cleaved under the standard basic conditions typically needed to cleave the oligonucleotide from the solid support.

Starting point for the synthesis of the fC phosphoramidite building block is TBS-protected 5-iodo-2′-deoxycytidine (1) which was synthesized as described previously.[15a] The formyl group was introduced by a carbonylative coupling, which we already employed for the synthesis of the free fC nucleoside and an hmC phosphoramidite.[1e,15a] The resulting TBS-protected 5-formyl-2′-deoxycytosine was benzoylated to give 2. Cleavage of the TBS groups was achieved with HF in pyridine.[18] When the latter reaction is carried out in EtOAc, the free nucleoside 3 precipitates and no further purification is needed. Finally, the 5′-hydroxyl group of compound 3 was protected with a DMT group and the 3′-hydroxyl group phosphitylated to give the final compound 4.

In the first step towards the caC phosphoramidite building block, 1 was converted to the methyl ester in a Pd^0-catalyzed reaction.[19] During the reaction the CO pressure has to be ad-

Figure 1. fC and caC exhibit an intramolecular hydrogen bond that was proposed to shift the amino/imino equilibrium towards the imino tautomeric form which would then form a mutagenic base pair with dA.

Scheme 2. Synthesis of the fC phosphoramidite **4** and of the caC phosphoramidite **7**. X = fC or caC.
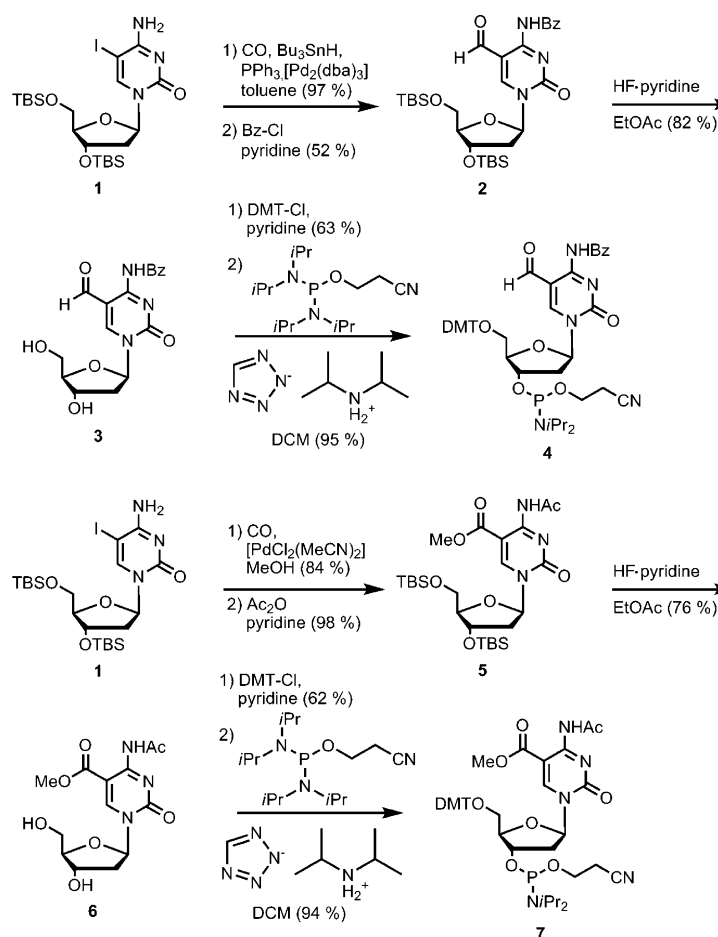
ODN1: 5' d(GTA ATG XGC TAG G) 3'
ODN2: 5' d(GTA GCC AGG TCG CAC GCG TGC TAX GAT GCG AGA CTG C) 3'

Figure 2. Analysis of ODN1 containing fC (Sequence of ODN1: 5′-GTAATGXGCTAG-3′; X = fC): A) Reversed-phase HPLC chromatogram of ODN1 (X=fC) directly after cleavage from the resin (0–50% buffer B in 45 min). B) Reversed-phase HPL chromatogram of ODN1 (X=fC) after cleavage of the DMT group and purification (0–20% buffer B in 45 min). C) MALDI-TOF spectrum of the purified strand ODN1 (X=fC). D) Digestion of the purified DNA strand ODN1 (X=fC) yields the canonical nucleosides plus fC, which co-elutes with dT.

justed carefully because otherwise a carbonylative Buchwald coupling occurs as a side reaction. Compound **5** was subsequently obtained after acetylation. In line with the reduced electron deficiency of the caC compound, the N4 acetyl protecting group is considerably more stable in compound **5** than in the fC derivative. For the cleavage of the silyl esters, we treated **5** with HF in pyridine/ethyl acetate which caused precipitation of pure **6**. The free nucleoside was finally DMT-protected and phosphitylated to give the caC phosphoramidite building block **7**.

For the incorporation of fC into oligonucleotides a standard solid-phase DNA synthesis was performed with phosphoramidite **4**. Coupling times for **4** were doubled to ensure good yields. As shown in Figure 2, the incorporation of fC proceeds without difficulty. After cleavage of the oligonucleotide from the resin and deprotection, the desired ODN1 with X=fC was obtained as the main product (Figure 2A, B). The same results were obtained for the 37mer ODN2 (Scheme 2). Deprotection was in both cases possible with either conc. NH_4OH or with 0.4M NaOH in MeOH. MALDI-TOF analysis of the oligonucleotides proved the

correct molecular weight of the synthesized strands (Figure 2C). After total digestion of the prepared oligonucleotides the four canonical bases were obtained with the fC base co-eluting with dT. Detection of the fC component was ensured by mass spectrometry (see Supporting Information).[7]

Incorporation of caC was possible with the phosphoramidite **7** using again rather standard solid-phase DNA synthesis conditions. The coupling times were extended to ensure good yields. In this case, however, standard deprotection conditions could not be employed, because treatment with ammonia or with K_2CO_3 in methanol would not yield the desired acid but either the amide or the methyl ester. Deprotection was therefore performed with NaOH in water/methanol.[15a] As shown in Figure 3A, this procedure allows to obtain the caC containing DNA strands as the only product. The caC containing DNA strands were subsequently purified by reversed phase HPLC (Figure 3B) and the correct composition was again proven by MALDI-TOF mass spectrometry (Figure 3C) and HPLC-MS analysis of the nucleotides obtained after total enzymatic digest. Under our conditions, the caC nucleoside eluted with a retention time of 5.3 min. Thus, fC- and caC-containing oligonucleotides can be efficiently prepared in high yields. With the improved phosphoramidite reagents the DNA strands are obtained under the described coupling and deprotection conditions in excellent purity.

In order to characterize and compare the base-pairing properties of hmC, fC and caC we measured melting points of sequence identical oligonucleotides under exactly the
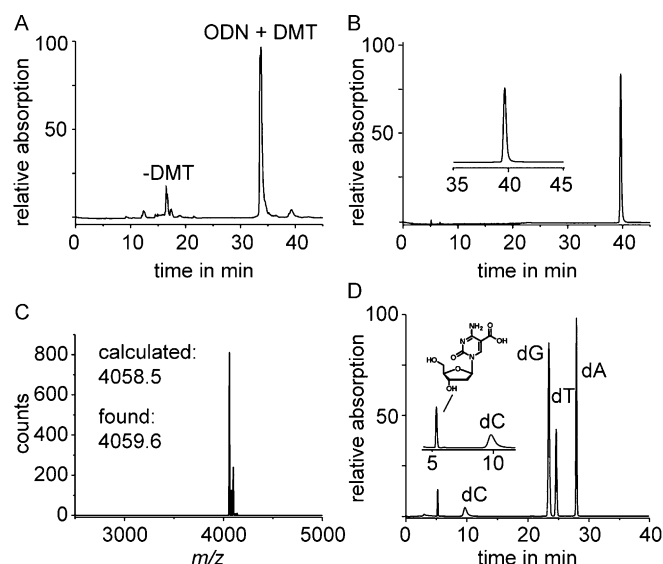
Figure 3. Analysis of ODN1 containing caC (Sequence of ODN1: 5'-GTAATGXGCTAG-3'; X = caC): A) Reversed-phase HPLC chromatogram of ODN1 (X = caC) directly after cleavage from the resin (0–50% buffer B in 45 min). B) Reversed-phase HPLC chromatogram of ODN1 (X = caC) after cleavage of the DMT group and purification (0–20% buffer B in 45 min). C) MALDI spectrum of the purified strand ODN1 (X = caC). D) HPLC chromatogram of purified ODN1 (X = caC) after enzymatic digestion.

Figure 4. Melting point analysis of ODNs, which contain different 5-substituted cytosines. 13 mer duplexes (ODN1: 5'-GTAATGXGCTAG-3') containing dC, hmC, fC, and caC as X was hybridized with different ODN3s (5'-CCTAGCYCATTAC-3') containing either dA, dC, dG, or dT in the Y-position opposite the cytosine bases. The duplexes (ODN1:ODN3) were repeatedly heated and cooled between 20 and 80 °C. Melting points of all individual base–counterbase combinations are shown.

same conditions (Figure 4). To this end, ODN1, with the respective modified dC-bases present at position X, was mixed with the ODN3 counter strand, which contained individual canonical nucleosides at position Y. The solutions were repeatedly heated and cooled between 80 and 20 °C. In all cases the cytosine derivatives paired best with dG (R^1 = H: 51.4 °C; R^1 = CH$_2$OH: 51.0 °C; R^1 = CHO: 52.7 °C; R^1 = COOH: 51.3 °C). Second best base pairing was in most cases observed with dA. For fC we noted a slightly increased stability of the fC:dT base pair. As already observed for hmC[20] the base pairing properties of fC and caC differ only marginally from dC. In addition, the stability differences between the nC:dG (n = hm, f, ca) base pairs and all mismatches are so high that a severe mutagenic effect of the modified dC bases can be excluded based on the melting point data.

To investigate the mutagenic potential further we developed a novel primer extension assay which we coupled directly to a pyrosequencing procedure. For the study we hybridized the nC containing oligonucleotides (ODN2) with a biotinylated primer strand (ODN4) and used various high and low fidelity polymerases to extend the primer past the modified cytosines to the full length product as shown in Figure 5. The resulting transcripts were isolated using streptavidine coated sepharose beads and subsequently used as new templates for pyrosequencing. As such they were hybridized to the reverse primer (ODN5) and subjected to standard pyrosequencing conditions using the Klenow polymerase.[21] The experiment has the advantage that it provides incorporation data in a situation in which all triphos-

Figure 5. Sequencing based mutagenicity assay with different cytosine derivatives. All values are the average of three independent measurements. A) ODN2, containing the modification, was elongated with different polymerases, the transcript was isolated, hybridized to a reverse primer and sequenced in a pyrosequencer. B) Nucleosides that are incorporated opposite the different cytosine derivatives. In contrast to the known mutagenic [4-NHOH]dC none of the other C5-modified cytosines are strongly mutagenic.

phates are present simultaneously and compete for incorporation.

We examined the mutagenic potential of hmC, fC and caC first with the high fidelity polymerase Klenow *exo*⁻.

However, due to the fact that modified bases are in vivo potentially replicated with a low fidelity polymerase, we also studied the common eukaryotic low fidelity polymerase eta (*Saccharomyces cerevisiae*) and polymerase kappa (*Homo sapiens*). For the study we used as control nucleosides dT as a non-mutagenic base and the mutagenic base [4-NHOH]dC, in which the exocyclic amino group is replaced by a hydroxylamine group.[22] This base is known to exhibit its mutagenicity due to the increased presence of the imino tautomeric form, which is also believed to be the basis for the mutagenicity of fC.[23] All data obtained are compiled in Figure 5. As expected, dC directs exclusively the incorporation of dG into the primer, while dT instructs the incorporation of dA. [4-NHOH]dC directs the incorporation of both dA and dG, which is in accordance with the proposal that it derives its mutagenic effect from the fact that it exists partially in the imino tautomeric form.[21b,23] To our surprise, we noted that fC in our coupled assay is, in contrast to previous reports, but in agreement with our melting point studies, only marginally mutagenic. All examined polymerases efficiently base paired fC with dG. In addition we indeed observed a small additional incorporation of dA. The incorporation is detected slightly above the error limit of the experiment, indicating that the formyl group promotes formation of the imino tautomeric form only to a very small extent. For caC and hmC we could not detect any other incorporation but dG. Our data clearly show that none of the new C5-modified nucleobases hmC, fC and caC are mutagenic. They all retain the base pairing behaviour of dC.

To investigate if the previously detected mutagenicity of fC may be caused by degradation products such as abasic sites, which may be generated by an accelerated proton catalyzed depyrimidination of fC, we performed stability studies. To this end fC and caC containing oligonucleotides were exposed to different pH conditions for 18 h and the decomposition was studied by HPLC. We noticed that neither the fC nor the caC containing ODN1 decomposed during this time in solutions buffered between pH 7 and 5 (37 °C). Under physiological conditions fC and caC are hence stable DNA modifications that show no increased lability. Only when we treated ODN1 (with X=fC and X=caC) under nonphysiological, mildly acidic conditions (0.01 % formic acid), we observed rapid decomposition of the strands. Under these conditions we observed an increased lability of fC and caC compared to dC because the dC containing control strands decomposed significantly slower.

We next recrystallized the fC base in order to obtain deeper insight into the intramolecular H-bond between the exocyclic amino group $NH_2(4)$ and the carbonyl–oxygen at C5, which was suggested to be the molecular reason for the increased mutagenicity. The structure of fC is depicted in Figure 6. The H-bond is indeed present in line with NMR studies carried out before.[12b,24] The low mutagenic potential of fC, however, suggests that this H-bond does not shift the amino/imino equilibrium substantially towards the imino-tautomeric form.
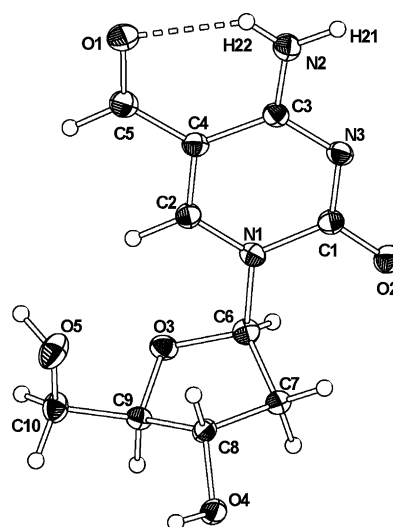
Figure 6. Molecular structure of fC in the crystal showing the intramolecular hydrogen bond between the aldehyde oxygen atom and the amino group. ORTEP representation, thermal ellipsoids are drawn at 50 % probability level.

Summary and Conclusion

In the last two years, it was discovered that the epigenetic base mC is further oxidized to hmC, fC and caC. Although the function of these new bases is still unclear at this point, they are present in substantial amounts in stem cells and neuronal cells. In principle oxidation of mC could occur in order to stimulate cellular mutagenic events. This is for example observed in the case of adenine, which is deaminated by special enzymes in immune cells to generate inosine, which base pairs with dC and not dT.[25] This mutagenic process is vital for a correct immune response. Alternatively, hmC, fC and caC could be new bases that have epigenetic functions. The previously reported mutagenicity of fC, which was found to be only slightly less mutagenic than [8-oxo]dG,[11a] argued against a pure epigenetic function. The strong H-bond between the $NH_2(4)$ group and the carbonyl oxygen at C5 was suggested to promote the imino tautomeric state, which was proposed to stimulate pairing possibilities with dA.[11a,12a] In this study we re-examined the mutagenicity of fC and we measured for the first time the mutagenic effect of caC. Both compounds were measured against [4-NHOH]dC, which is a known mutagenic dC modification due to an increased imine character. We also report a crystal structure of fC, which indeed shows the expected H-bond. Our data show that [4-NHOH]dC is mutagenic and that it pairs both with dG and dA in accordance with increased existence of the imino-tautomeric form. The new bases fC and caC, however, only pair with dG and do not show such dual base pairing characteristics. They were also found to be stable under physiological conditions, which eliminates the possibility that they exhibit a mutagenic potential due to rapid depyrimidination. We believe that the previously observed mutagenicity of fC in vivo is therefore not caused by

a polymerase acting on the fC base but by molecular structures that are formed due to the in vivo action of either base- or nucleotide excision repair acting on the modifications. Indeed, it was recently shown that fC and caC are selectively excised by the enzyme TDG, which would generate an abasic site.[8a, 9b] In summary, our data show that fC and caC are non-mutagenic and therefore can be involved in epigenetic programming of cells without disturbing the genetic sequence information.

Experimental Section

General remarks: Chemicals and solvents were purchased from ABCR, Alfa Aesar, Acros, Fluka, Sigma–Aldrich or TCI in the qualities puriss., p.a. or purum. Dry solvents (< 50 ppm H_2O) were obtained from Fluka and Acros. All reactions employing dry solvents were performed under inert atmosphere (N_2). Technical grade solvents were distilled prior to use for column chromatography and liquid–liquid extractions on a rotary evaporator (Heidolph Laborota 4000). Reaction products were dried at high vacuum (0.1 mbar). Aqueous solutions were concentrated on a SpeedVac plus CS110 A or SPD 111V from Savant or lyophilized (Christ ALPHA 2-4). Thin layer chromatography (TLC) was performed with aluminium plates (silica gel 60 F254, 10×5 cm). Substances were visualized by illumination with UV light ($\lambda = 254$ nm). ESI-MS was performed on a Finnigan LTQ FTICR. MALDI-TOF was performed on a Bruker Autoflex II. NMR spectra were recorded on the following spectrometers: Varian Oxford 200, Bruker AC 300, Varian XL 400 and Bruker AMX 600. The chemical shifts (δ) are given in ppm, the coupling constants (J) in Hz. Exonuclease-deficient DNA polymerase I from *E. coli* (Klenow exo^-) was obtained from NEB, Pol η was purified as described before[26] and Pol κ was cloned from human cDNA and expressed in Rosetta 2 (DE3) cells.

Oligonucleotide synthesis and deprotection: DNA synthesis was performed on an Expedite 8909 Nucleic Acid Synthesis System (PerSeptive Biosystems) or an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions. Phosphoramidites for dA, dC, dG, dT and CPG carriers were obtained from Amersham, Glen Research or PE Biosystems. The hmC and fC phosphoramidites were dissolved in dry MeCN, the caC phosphoramidite in dry toluene. The oligonucleotides were removed from the resin under concomitant cleavage of the standard nucleobase protecting groups by treatment with 750 μL conc. NH_3 and 250 μL ethanol at room temperature for 18 h (fC) or treatment with 0.4 M NaOH in water/methanol 1:4. Subsequently, the solution was decanted from the resin. When the deprotection was carried out with NaOH, a 1 M TEA/AcOH solution (600 μL) was added and the solution concentrated to a final volume of 750 μL. After deprotection with NH_4OH the solution was concentrated to dryness and subsequently taken up in water. Please note that deprotection with NaOH is not compatible with DMF protecting groups.

HPLC and cleavage of DMT groups: Purification and analysis of ODNs was performed on a Waters system (Alliance 2695 with PDA 2996; preparative HPLC: 1525EF with 2484 UV detector) with VP 250/10 Nucleosil 100-7 C18 and VP 250/4 Nucleosil 20-3 C18 columns from Macherey–Nagel using a gradient of 0.1 M triethylamine/acetic acid in water and 80% acetonitrile. The oligonucleotides still containing the trityl group were deprotected by addition of 100 μL of an 80% acetic acid solution. After incubation at RT for 20 min, 100 μL of water together with 60 μL of a 3 M solution of sodium acetate were added. Finally, the ODN was precipitated by the addition of 1600 μL ethanol at −20°C (30 min). The strands were stored in ddH₂O.

Enzymatic digestion: For the enzymatic digestion DNA mixtures (4 to 10 μg in a final volume of 100 μL H_2O) were heated to 100°C for 5 min to denature the DNA and rapidly cooled on ice. Buffer E (10 μL, 300 mM ammonium acetate, 100 mM $CaCl_2$, 1 mM $ZnSO_4$, pH 5.7) and nuclease S1 (80 units, *Aspergillus oryzae*) were added to the mixture and incubated

for 3 h at 37°C. Addition of buffer F (12 μL, 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 h at 37°C completed the digestion. The supernatant was removed, the volume reduced to 100 μL and measured with HPLC-ESI-MS.

HPLC-ESI-MS: The samples (100 μL injection volume) were analyzed by HPLC-ESI-MS on a Thermo Finnigan LTQ Orbitrap XL and were chromatographed by a Dionex Ultimate 3000 HPLC system using gradient of 2 mM ammonium format in water and 80% acetonitrile over an Uptisphere120–3HDO column from Interchim.

Melting points: Melting profiles were measured on a Jasco V-650 spectrophotometer using quartz glass cuvettes with 1 cm path length. In these were in a total volume of 1 mL: 1 μmol ODN, 1 μM counterstrand, 150 mM NaCl and 10 mM Tris pH: 7.4. First, the oligonucleotides were hybridized by slowly cooling the samples down from 80°C to 20°C. The melting profiles started with a denaturing run (20 to 80°C) with a slope of 0.5°C min⁻¹. At least two denaturing and two renaturing ramps were performed and averaged for evaluation of the melting point. ($T_M =$ zero-crossing of second derivative of the 400 nm-background corrected change in hyperchromicity at 260 nm.) For analysis of the data, the program Origin (Microcal) was used.

Pyrosequencing analysis: A 37 mer template DNA strand (ODN 2: 5′-d(GTA GCC AGG TCG CAC GCG TGC TAX GAT GCG AGA CTG C)-3′ was prepared containing either dC, mC, hmC, fC, caC, or ⁴⁻NHOHdC, or dT at position X. 10 pmol of the template were hybridized with a biotinylated primer (ODN 4: 5′-d(biotin-GCAGTCTCGCATC)-3′, Metabion). Subsequently, the primer extension experiments were performed with the exonuclease deficient DNA polymerase I from *E. coli* (Klenow exo^-), Pol η and Pol κ. The polymerases (1 U Klenow exo^-, 1 μM Pol η, or 0.5 μM Pol κ), 50 μM dNTPs and 0.5 μM dsDNA were incubated in a total volume of 20 μL 1× NEBuffer 2 for 30 min at different temperatures (30°C for Pol η, 37°C for Klenow exo^- and Pol κ). To this solution 2 μL streptavidin sepharose beads (GE Healthcare, Uppsala, Sweden), 40 μL binding buffer (Qiagen, Hilden, Germany) and 18 μL ddH₂O were added. After agitation at 1400 rpm for 15 min the beads were captured with a Vacuum Prep Tool (Qiagen, Hilden, Germany), washed with 70% EtOH, 0.1 M NaOH and Washing Buffer (Qiagen, Hilden, Germany). The beads were dissolved in 25 μL Annealing Buffer (Qiagen, Hilden, Germany) containing 5 pmol sequencing primer ODN5 (5′-d(GTAGCCAGGTCGCACGCGTGCTA)-3′, Metabion, Martinsried, Germany). Pyrosequencing was performed on a PyroMark Q24 Pyrosequencer using standard conditions (Qiagen, Hilden, Germany). The data was analyzed by the software provided by the manufacturer. Peak heights were exported to Microsoft Excel and the average of all blank sites for the individual nucleobases calculated. This resulted in one blank for dT, one blank for dC, one blank for dG and one blank for dA. This value was subtracted from the values for the incorporation of the individual triphosphates. With these data the relative incorporation at every variable position was calculated. The data are the average of three measurements.

Recrystallization: Single crystals, suitable for X-ray diffraction, were obtained by very slow evaporation of the solvent of a fC solution in methanol. The crystals were introduced into perfluorinated oil and a suitable single crystal was carefully mounted on the top of a thin glass wire. Data collection was performed with an Oxford Xcalibur 3 diffractometer equipped with a Spellman generator (50 kV, 40 mA) and a Kappa CCD detector, operating with Mo$_{Kα}$ radiation ($\lambda = 0.71071$ Å). Data collection was performed with the CrysAlis CCD software;[27] CrysAlis RED software[27] was used for data reduction. Absorption correction using the SCALE3 ABSPACK multiscan method[28] was applied. The structures were solved with SHELXS-97,[29] refined with SHELXL-97[30] and finally checked using PLATON.[31] All hydrogen atoms involved in hydrogen bonding were found in the differential Fourier map and refined. Details for data collection and structure refinement are summarized in Table S4.

CCDC 843055 contains the supplementary crystallographic data for this paper. The data can be obtained free of charge from The Cambridge Crystallographic Date Centre via www.ccdc.cam.ac.uk/data_request/cif.

Acknowledgements

[1] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935; c) M. Münzel, D. Globisch, T. Carell, *Angew. Chem.* **2011**, *123*, 6588–6596; d) M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem.* **2010**, *122*, 5503–5505; *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377; e) D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367; f) A. Szwagierczak, A. Brachmann, C. S. Schmidt, S. Bultmann, H. Leonhardt, F. Spada, *Nucleic Acids Res.* **2011**, *39*, e181.

[2] a) Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Mol. Cell* **2011**, *42*, 451–464; b) H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *Genes Dev.* **2011**, *25*, 679–684; c) H. Wu, A. C. D′Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. Eve Sun, Y. Zhang, *Nature* **2011**, *473*, 389–393; d) 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; e) K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. C. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343–348; f) J. Walter, *Cell Stem Cell* **2011**, *8*, 121–122; g) W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394–397; h) K. P. Koh, A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley, A. Rao, *Cell Stem Cell* **2011**, *8*, 200–213; i) K. Iqbal, S.-G. Jin, G. P. Pfeifer, P. E. Szabó, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3642–3647; j) G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398–402; k) 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; l) M. Ko, Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, A. Rao, *Nature* **2010**, *468*, 839–843; m) S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129–1133.

[3] C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580–583.

[4] R. Bonasio, S. J. Tu, D. Reinberg, *Science* **2010**, *330*, 612–616.

[5] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607–620.

[6] a) M. Akhtar, M. R. Calder, D. L. Corina, J. N. Wright, *Biochem. J.* **1982**, *201*, 569–580; b) P. M. Shaffer, C. A. Hsu, M. T. Abbott, *J. Bacteriol.* **1975**, *121*, 648–655; c) J. M. Simmons, T. A. Muller, R. P. Hausinger, *Dalton Trans.* **2008**, 5132–5142; d) J. A. Smiley, J. M. Angelot, R. C. Cannon, E. M. Marshall, D. K. Asch, *Anal. Biochem.* **1999**, *266*, 85–92.

[7] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146–7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.

[8] a) 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; b) 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] a) S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67–79; b) A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338.

[10] J. U. Guo, Y. Su, C. Zhong, G.-l. Ming, H. Song, *Cell* **2011**, *145*, 423–434.

[11] a) H. Kamiya, H. Tsuchiya, N. Karino, Y. Ueno, A. Matsuda, H. Harashima, *J. Biochem.* **2002**, *132*, 551–555; b) N. Karino, Y. Ueno, A. Matsuda, *Nucleic Acids Res.* **2001**, *29*, 2456–2463.

[12] a) C. J. La Francois, Y. H. Jang, T. Cagin, W. A. Goddard, L. C. Sowers, *Chem. Res. Toxicol.* **2000**, *13*, 462–470; b) A. Burdzy, K. T. Noyes, V. Valinluck, L. C. Sowers, *Nucleic Acids Res.* **2002**, *30*, 4068–4074; c) G. Kawai, T. Yokogawa, K. Nishikawa, T. Ueda, T. Hashizume, J. A. McCloskey, S. Yokoyama, K. Watanabe, *Nucleosides Nucleotides* **1994**, *13*, 1189–1199; d) Y. El Safadi, J. C. Paillart, G. Laumond, A. M. Aubertin, A. Burger, R. Marquet, V. Vivet-Boudou, *J. Med. Chem.* **2010**, *53*, 1534–1545.

[13] M. Sumino, A. Ohkubo, H. Taguchi, K. Seio, M. Sekine, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 274–277.

[14] T. Carell, M. Münzel, "Building Blocks and Methods for the Synthesis of 5-Hydroxymethylcytosine Containing Nucleic Acids", European Patent Application EP10191078.4, filed Nov. 12, 2010.

[15] a) M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671–5673; b) S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, *Nucleic Acids Res.* **1997**, *25*, 553–558; c) A. S. Hansen, A. Thalhammer, A. H. El-Sagheer, T. Brown, C. J. Schofield, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1181–1184; d) M. de Kort, P. C. de Visser, J. Kurzeck, N. J. Meeuwenoord, G. A. van der Marel, W. Ruger, J. H. van Boom, *Eur. J. Org. Chem.* **2001**, 2075–2082; e) Q. Dai, C.-X. Song, T. Pan, C. He, *J. Org. Chem.* **2011**, *76*, 4182–4188.

[16] Q. Dai, C. He, *Org. Lett.* **2011**, *13*, 3446–3449.

[17] a) H. Lusic, E. M. Gustilo, F. A. P. Vendeix, R. Kaiser, M. O. Delaney, W. D. Graham, V. A. Moye, W. A. Cantara, P. F. Agris, A. Deiters, *Nucleic Acids Res.* **2008**, *36*, 6548–6557; b) G. H. Clever, K. Polborn, T. Carell, *Angew. Chem.* **2005**, *117*, 7370–7374; *Angew. Chem. Int. Ed.* **2005**, *44*, 7204–7208.

[18] I. E. Szabo, T. C. Bruice, *Bioorg. Med. Chem.* **2004**, *12*, 4233–4244.

[19] Y. Nomura, N. Haginoya, Y. Ueno, A. Matsuda, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2811–2816.

[20] A. Thalhammer, A. S. Hansen, A. H. El-Sagheer, T. Brown, C. J. Schofield, *Chem. Commun.* **2011**, *47*, 5325–5327.

[21] a) M. Ronaghi, M. Uhlen, P. Nyren, *Science* **1998**, *281*, 363–365; b) M. Münzel, L. Lercher, M. Müller, T. Carell, *Nucleic Acids Res.* **2010**, *38*, e192.

[22] M. Münzel, C. Szeibert, A. F. Glas, D. Globisch, T. Carell, *J. Am. Chem. Soc.* **2011**, *133*, 5186–5189.

[23] D. M. Brown, M. J. E. Hewlins, P. Schell, *J. Chem. Soc. C* **1968**, 1925–1929.

[24] G. Kawai, T. Yokogawa, K. Nishikawa, T. Ueda, T. Hashizume, J. A. McCloskey, S. Yokoyama, K. Watanabe, *Nucleosides Nucleotides Nucleic Acids* **1994**, *13*, 1189–1199.

[25] E. Giblett, J. Anderson, F. Cohen, B. Pollara, H. Meuwissen, *Lancet* **1972**, *300*, 1067–1069.

[26] S. Schorr, S. Schneider, K. Lammens, K.-P. Hopfner, T. Carell, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20720–20725.

[27] CrysAlis, 1.171.27p5 beta (release 01-04-2005 CrysAlis171.NET) (compiled Apr 1, 2005, 17:53:34) ed., Oxford Diffraction Ltd., **2005**.

[28] SCALE3, 1.0.4, gui:1.0.3 ed., Oxford Diffraction Ltd., **2005**.

[29] G. M. Sheldrick, SHELXS-97: Program for Crystal Structure Solution, University of Göttingen, Göttingen (Germany), **1997**.

[30] G. M. Sheldrick, SHELXL-97: Program for the Refinement of Crystal Structures, University of Göttingen, Göttingen (Germany), **1997**.

[31] A. L. Spek, PLATON: A Multipurpose Crystallographic Tool, Utrecht University, Utrecht (The Netherlands), **1999**.

3.3 Mechanismus und Stammzellaktivität der Decarboxylierung von 5-Carboxycytosin, bestimmt mittels Isotopenverfolgung

Stefan Schiesser, Benjamin Hackner, <u>Toni Pfaffeneder</u>, Markus Müller, Christian Hagemeier, Matthias Truss und Thomas Carell, *Angew. Chem.* **2012**, *124*, 6622–6626; *Angew. Chem., Int. Ed.* **2012**, *51*, 6516–6520. *Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing.*

Prolog

Nach der Entdeckung der stufenweisen Oxidation von 5mdC zu 5hmdC, 5fdC und 5cadC in genomischer DNA durch die Tet-Enzyme wurden zwei Mechanismen zur aktiven Entfernung der beiden höher oxidierten Modifikationen vorgeschlagen und untersucht. Ein Mechanismus soll durch Deglykosylierung von 5fdC und 5cadC per Basenexzisionsreparatur (BER) erfolgen.[7, 9, 317-318] Beide Modifikationen werden hierbei durch die Thymin-DNA-Glykosylase (Tdg) erkannt und ausgeschnitten. Nach einem zweiten Mechanismus könnte 5cadC direkt decarboxyliert werden (siehe Abschnitt 1.6). Dieser Mechanismus ist eine attraktive Alternative, da so die intermediäre Bildung von abasischen Stellen bei der BER umgangen werden würde, was zu genomischer Instabilität führen könnte.[3, 76] In diesem Manuskript wird der erste Hinweis für eine solche Decarboxylierungsaktivität in embryonalen Stammzellen der Maus beschrieben. Experimentell wurde dieser Hinweis durch die Etablierung eines Massenspektrometrie-basierten Isotopenverfolgungsexperiments erhalten. Hierfür wurde als Reportermolekül ein [^{15}N$_2$]-markiertes 5cadC-enthaltendes und Biotin-modifiziertes Oligodesoxynukleotid synthetisiert und dieses mit Zellkernextrakten inkubiert. Die Re-Isolation wurde durch Biotin/Streptavidin-Anreicherung realisiert. Per hochauf-lösender Massenspektrometrie (HPLC-ESI-HRMS) konnte so decarboxyliertes [^{15}N$_2$]-dC nachgewiesen werden. In chemischen Modellstudien wurde der zugrundeliegende Mechanismus näher untersucht.
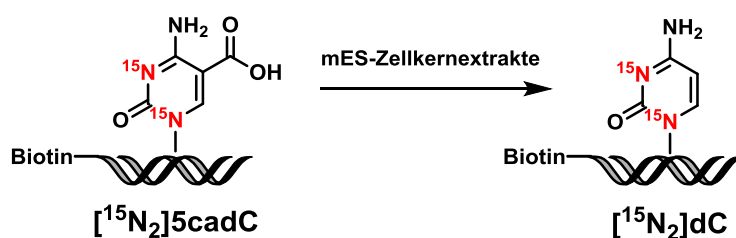
Autorenbeitrag

Für dieses Manuskript entwickelte ich die Synthese von 5-Iod-[^{15}N$_2$]-dC, welches *Stefan Schiesser* und *Benjamin Hackner* für die Synthese der [^{15}N$_2$]-5cadC-enthaltenden

Oligodesoxynukleotide nutzten. Ferner wirkte ich bei der Etablierung des *Assays* mit und entwickelte die HPLC-ESI-HRMS-Methode. *Benjamin Hackner* führte die Kernextrakt-Inkubationen und DNA-Isolationen durch.

Lizenz

Mechanismus und Stammzellaktivität der Decarboxylierung von 5-Carboxycytosin, bestimmt mittels Isotopenverfolgung**

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

5-Methylcytosin (mC) ist eine wichtige, gut untersuchte Nukleobase, die in vielen biologischen Prozessen, wie z.B. der Genexpression, der Genomprogrammierung, der Inaktivierung des X-Chromosoms und der Unterdrückung beweglicher Genelemente eine entscheidende Rolle spielt.[1–3] Kürzlich wurde herausgefunden, dass mC in Neuronen und embryonalen Stammzellen durch TET-Enzyme weiter modifiziert werden kann. Die TET-Enzyme nutzen 2-Ketoglutarat, um mC zunächst zum 5-Hydroxymethylcytosin (hmC)[4,5] und weiter zum 5-Formylcytosin (fC)[6] und 5-Carboxycytosin (caC)[7,8] zu oxidieren (Schema 1).[9,10] Es wird derzeit angenommen, dass diese oxidative Chemie an einem Prozess beteiligt ist, den man aktive Demethylierung nennt. Dieser Prozess erlaubt es Zellen, das Methylcytidin durch ein unmodifiziertes 2′-Desoxycytidin (dC) zu ersetzen. Hierdurch werden die biologischen Effekte, die durch Cytosinmethylierung eingeleitet

werden, revertiert.[11] Zwei mögliche aktive Demethylierungsmechanismen werden derzeit intensiv diskutiert (Schema 1). Mechanismus 1 (Schema 1) basiert auf speziellen Glykosylasen,[12–15] wie z.B. der Thymin-DNA-Glycosylase (TDG), bei der kürzlich gezeigt wurde, dass diese die glykosidische Bindung von fC und caC, aber auch von 5-Hydroxymethyluridin (hmU)[15,16] spalten kann. 5-Hydroxymethyluridin wird durch Deaminierung von hmC gebildet.[15] Diese Glykosylaseaktivität produziert zunächst abasische
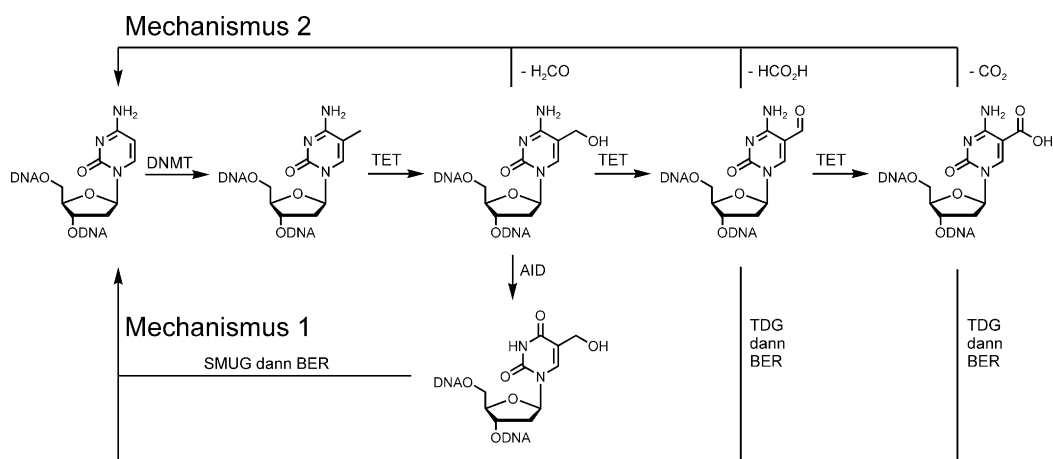
Schema 1. Aktive Demethylierungspfade, die z.B. Eliminierungsreaktionen beinhalten (Mechanismus 2). Formaldehyd kann von hmC und Ameisensäure von fC abgespalten werden, und es kann eine Decarboxylierung von caC erfolgen. Mechanismus 1 umfasst die Basenexzisionsreparatur (BER), durch die fC und caC durch dC ersetzt werden oder, nach vorhergehender Deaminierung von hmC zu hmU, hmC durch dC ersetzt wird.

[*] S. Schiesser,[+] B. Hackner,[+] T. Pfaffeneder, M. Müller,
 Prof. Dr. T. Carell
 Zentrum für integrative Proteinforschung, Department für Chemie,
 Ludwig-Maximilians-Universität München
 Butenandtstraße 5-13, 81377 München (Deutschland)
 E-Mail: matthias.truss@charite.de
  thomas.carell@lmu.de

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

 [+] Diese Autoren trugen zu gleichen Teilen zu der Arbeit bei.

Stellen, welche dann durch das Basenexzisionsreparatursystem (BER) repariert werden, indem eine dC-Base eingebaut wird (Schema 1 unterer Teil). Der ganze Prozess führt insgesamt zu einem Austausch von mC durch dC. Der zweite Mechanismus, so glaubt man, umfasst eine Retro-Aldolartige Reaktion, die mit hmC[17,18] oder fC beginnt. Alternativ wird eine Decarboxylierung von caC (Schema 1) als Möglichkeit formuliert. Diese Reaktionen erlauben es, mC durch dC auszutauschen, ohne dass intermediär abasische Stellen und Strangbrüche auftreten, bei denen es sich um gefährliche DNA-Schäden handelt. Darüber hinaus entsteht so kein potentiell gefährlicher Formaldehyd. Auch wenn die Decarboxylierung von caC bis heute nicht beobachtet werden konnte, sind ähnliche Transformationen von der Orotat- und Isoorotatdecarboxylase bekannt.[19]

Um tiefere Einblicke in die mögliche Decarboxylierung von caC in Stammzellen zu erhalten, haben wir eine sensitive

Isotopenverfolgungsmethode entwickelt. Für dieses Experiment wurde das doppelt ^{15}N-markierte caC-Phosphoramidit-reagens [[15]N$_2$]-caC entwickelt. Dieses wurde in ein 30-meres Oligonucleotid **D1** mit der Sequenz des Ecat1-Promoters eingebaut.[23] Bei diesem Promoter ist bekannt, dass aktive Demethylierung erfolgt. Das natürliche Vorkommen von doppelt ^{15}N-isotopologem dC ([[15]N$_2$]-dC) ist ausgesprochen gering, was es uns ermöglicht, die an der caC-Verbindung wirkende Chemie mit hochaufgelöster Massenspektrometrie zu verfolgen. Die Synthese des [[15]N$_2$]-caC-Phosphoramidits ist in Schema 2 dargestellt (siehe auch Hintergrundinformationen). Ausgangspunkt ist [[15]N$_2$]-Uracil (**1**), das aus [[15]N$_2$]-Harnstoff und Propinsäure hergestellt wurde.[20] Vorbrüggen-Nukleosidierung mit dem Bis(toluoyl)-geschützten Chlor-ribofuranosid[21] ergab das β-konfigurierte Nukleosid **2** in 60% Ausbeute. Elektrophile Iodierung zu **3**,[22] Umsetzung der C4-Ketofunktion zu einem Amin (**4**) und Schutzgruppenaustausch lieferten das Schlüsselintermediat **5**. Pd-katalysierte CO-Insertion und anschließende Methanolyse des Pd-CO-Intermediats mit Methanol ergaben in 65% Ausbeute das [[15]N$_2$]-caC-Vorläufermolekül **6**. Nach doppelter Benzoyl-schützung der C4-Aminogruppe (**7**) und Abspaltung der TBS-Gruppen (**8**) folgte die gezielte Schützung der 5′OH-Gruppe mit Dimethoxytritylchlorid (DMT-Cl) (**9**). Anschließend wurde die 3′OH-Gruppe mithilfe von Standardprozeduren in das finale Phosphoramidit **10** überführt. Das Phosphoramidit wurde zur Synthese der 30-meren Oligonukleotide **ODN1** und **ODN2** eingesetzt. Die zwei Stränge wurden anschließend zum Doppelstrang **D1** mit der Sequenz des Ecat1-Promoters hybridisiert, welcher nun zwei [[15]N$_2$]-markierte caC-Derivate enthielt.

In einen der DNA-Stränge (**ODN1**) wurde zusätzlich die Base 5-Octadienyl-dU (Schema 2, unten) eingebaut. Nachfolgend wurde milde Klickchemie mit dem dargestellten Biotinazid (Schema 2, unten und Hintergrundinformationen) durchgeführt, um den Biotinmarker mit dem Octadienyl-dU[23] zu verknüpfen. Alle DNA-Stränge, die für die nachfolgenden Experimente benötigt wurden, wurden sorgfältig per HPLC aufgereinigt, um das markierte DNA-Material in hoher Reinheit > 99.9% zu erhalten. Nach der Durchführung der Klickchemie wurde das biotinylierte Oligonukleotid daher erneut per HPLC aufgereinigt. Die Identität der Oligonukleotide wurde in allen Fällen durch MALDI-TOF-Massenspektrometrie bestätigt (Schema 2 und Hintergrundinformationen).

Für das Isotopenverfolgungsexperiment (Abbildung 1) wurde der DNA-Duplex **D1** (**ODN1** + **ODN2**) mit den zwei [[15]N$_2$]-markierten caC-Basen und dem Biotinmarker zu nuklearem Extrakt von embryonalen Stammzellen von Mäusen (mESC) gegeben. Nach 1 h Inkubation bei 37°C (für Details siehe Hintergrundinformationen) wurde die DNA mithilfe Streptavidin-beladener magnetischer Partikel re-isoliert. Die isolierte DNA wurde anschließend vollständig verdaut (Hintergrundinformationen) und der Verdau per HPLC-MS mithilfe eines hochauflösenden Massenspektrometers (Orbitrap XL, für Details siehe Hintergrundinformationen) analysiert (Abbildung 1 b–e). Wir untersuchten zuerst die Resultate aus Kontrollexperimenten. In diesen wurde **D1** mit nuklearem Extrakt von Nichtstammzellen oder einfach mit Puffer inku-
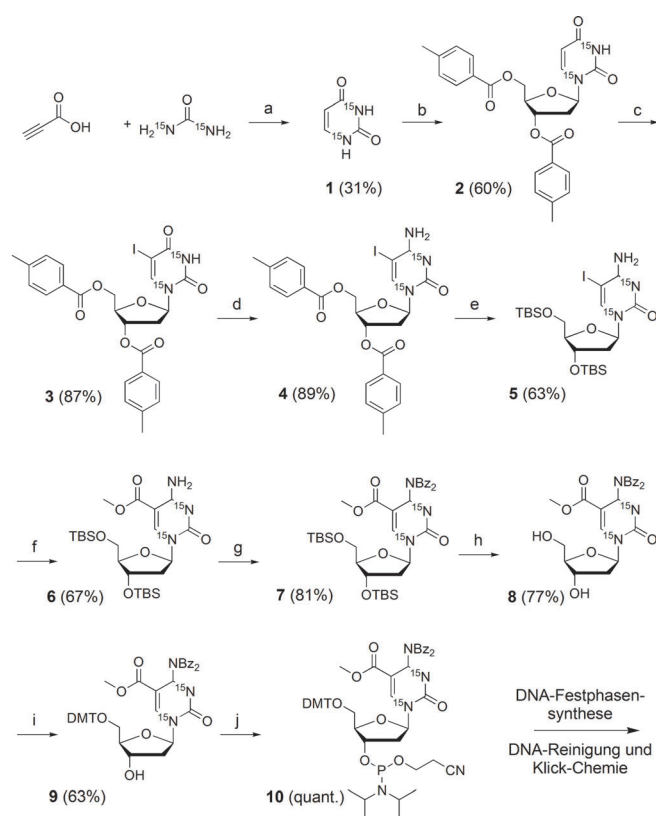
ODN 1: 5′-CCT TTC CGA AGG GAC GTT GAC **X**GG ATG C**Y**C-3′
ber. 9657 Da, gef. 9658 Da

ODN 2: 5′-GAG CAT CCG GTC AAC GTC CCT T**X**G GAA AGG-3′
ber. 9259 Da, gef. 9259 Da, **X**= [[15]N$_2$]-caC, **Y**= Biotin-markiertes dU
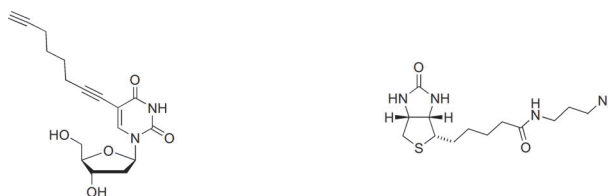
Schema 2. Synthese des [[15]N$_2$]-markierten caC-Phosphoramidits und der DNA-Stränge **ODN1** und **ODN2** mit eingebautem [[15]N$_2$]-markiertem caC. Der Biotinmarker wurde durch Anwendung von Klickchemie eingebaut. Reaktionen und Bedingungen: a) Polyphosphorsäure, 95°C, 19 h; b) 1. HMDS, TMS-Cl, 120°C, 1 h; 2. Hoffers Chlorzucker, CHCl$_3$, RT, 2.5 h; c) CAN, LiI, MeCN, 80°C, 2 h; d) 1. 1,2,4-Triazol, POCl$_3$, TEA, 30°C, über Nacht; 2. NH$_4$OH, 1,4-Dioxan, RT, 10 min; e) 1. K$_2$CO$_3$, MeOH, RT, 19 h; 2. TBS-Cl, Imidazol, DMF, RT, 4 d; f) [Pd-(MeCN)$_2$Cl$_2$], CO, MeOH, 60°C, 17 h; g) BzCl, Pyridin, 0°C nach RT, 19 h; h) HF-Pyridin, EtOAc, RT, 14 h; i) DMT-Cl, Pyridin, RT, 16.5 h; j) Diisopropylammoniumtetrazolid, 2-Cyanoethyl-*N,N,N′,N′*-tetraisopropylphosphordiamidit, DCM, RT, 17 h. HMDS = Hexamethyldisilazan, TMS = Trimethylsilyl, CAN = Cerammoniumnitrat, TEA = Triethylamin, TBS = *tert*-Butyldimethylsilyl, DMT = Dimethoxytrityl.

biert. Die massenspektrometrische Analyse dieser Kontrollexperimente zeigt, wie in Abbildung 1 d,e dargestellt ist, zwei zusätzliche Signale mit $m/z = 230.1014$ und 230.0976 im Massebereich des dC-Nukleosids (Abbildung 1 e). Diese m/z-Signale stammen von den natürlich vorkommenden Iso-
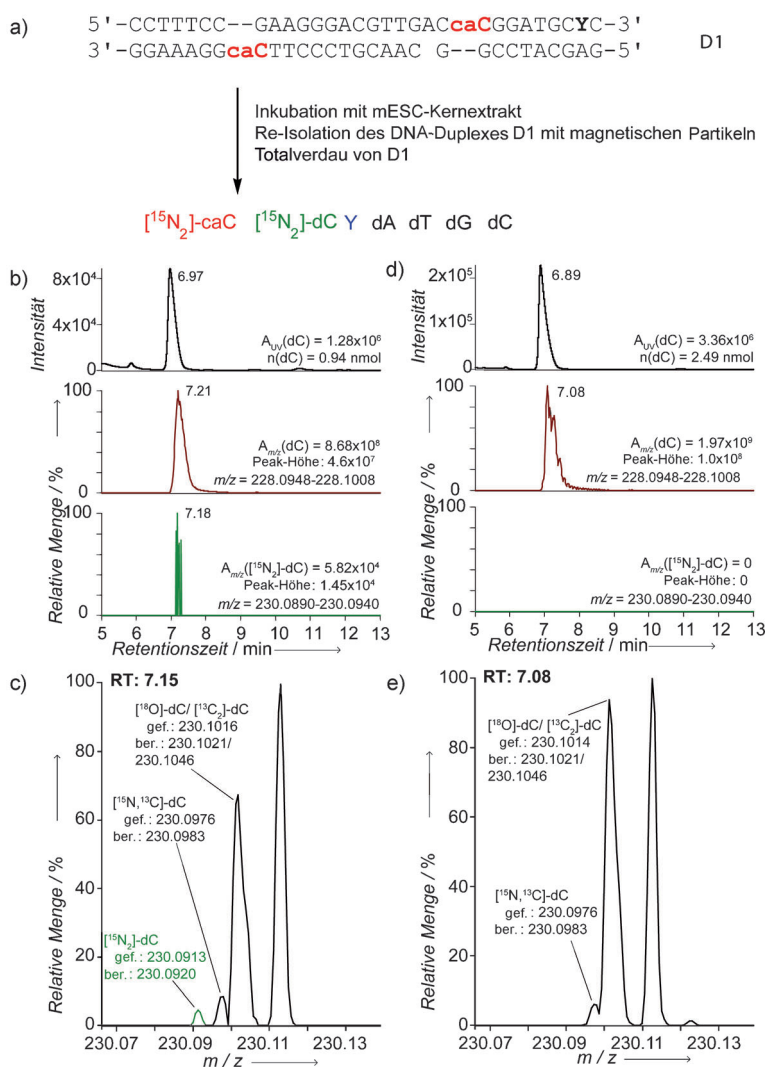
Abbildung 1. a) Darstellung des Nuklearextraktexperiments. b,d) HPLC-MS-Analyse von **D1** behandelt mit mESC-Nuklearextrakt (b) und Nicht-mESC-Nuklearextrakt (d). UV-Spuren der HPLC-Messungen bei der Retentionszeit von dC (oben) sowie Massespuren für dC (228.0948–228.1008, Mitte) und Massespuren für [^{15}N$_2$]-markiertes dC (230.0890–230.0940, unten). c,e) Hochaufgelöste Massenspektren (m/z=230.0700–230.1300) von **D1** nach Behandlung mit mESC-Nuklearextrakt (c) und Nicht-mESC-Extrakt (e).

pariert und war damit klar detektierbar. Darüber hinaus stimmten die hochaufgelösten Massedaten für die Verbindung [^{15}N$_2$]-dC mit dem berechneten Wert der Verbindung [^{15}N$_2$]-dC ($m/z_{\text{ber.}} =$ 230.0920) überein. Diese Daten zeigen zusammen, dass mESC-Nuklearextrakt die Fähigkeit hat, caC zu dC zu decarboxylieren. Ob diese Aktivität durch ein spezifisches Enzym oder durch andere Faktoren erzeugt wird, die in den Extrakten anwesend sind, bedarf der weiteren Untersuchung.

Um die Reaktionen zu untersuchen, die es Stammzellen ermöglichen würden, caC zu decarboxylieren, erhitzten wir zunächst das caC-Nukleosid und Oligonukleotide, welche die caC-Base beinhalteten (Abbildung 2). Unter keinen Umständen konnten wir jedoch Decarboxylierung beobachten. Auch wenn wir das caC-Nukleosid in Wasser mehrere Stunden zum Rückfluss erhitzten, zeigte die massenspektrometrische Analyse der Reaktionsmischung keine Hinweise auf das Stattfinden einer Decarboxylierung (Daten nicht gezeigt). Das ist nicht überraschend, denn durch die Decarboxylierung würde ein sehr energiereiches Cytosin-Carbanionintermediat entstehen. Demgegenüber sollte die Absättigung der C5-C6-Doppelbindung mit anschließender Tautomerisierung der C4-Aminogruppe ein Iminintermediat ergeben, das als Substruktur eine β-Iminocarbonsäure aufweist. Hier sollte eine effiziente Decarboxylierung (Abbildung 2) stattfinden können.[10,24] Um diese Idee näher zu untersuchen, reduzierten wir den 5-Carboxy-dC-methylester **11** mit NaBH$_4$. Die reduzierte Verbindung **12** ist stabil. Sie wurde isoliert und vollständig charakterisiert (Abbildung 2 a und Hintergrundinformationen). Nachfolgende Verseifung des Methylesters führte in der Tat zu einer recht schnellen Decarboxylierung zu Verbindung **13**. Behandlung des resultierenden Dihydro-dC-Derivats mit DDQ ergab die TBS-geschützte Verbindung dC. Dieses Resultat belegt, dass die Absättigung der C5-C6-Doppelbindung für die effiziente Decarboxylierung entscheidend ist, wie es kürzlich bereits vorgeschlagen wurde.[10] Da die Reaktion mit einer H$^-$-Spezies biologisch unwahrscheinlich ist, behandelten wir das TBS-geschützte caC nachfolgend mit einem Gemisch aus einem Thiol und Imidazol, in der Hoffnung, eine temporäre 1,4 Addition zu erreichen, welcher sich nach der Decarboxylierung eine Eliminierungsreaktion anschließen könnte. Diese chemische Kaskade wird in der Tat beobachtet (Daten nicht gezeigt). Die Behandlung des caC-Nukleosids mit Octanthiol in Anwesenheit von Imidazol ergab einen neuen Fleck auf dem Dünnschichtchromatogramm. Dieser Fleck wurde nicht beobachtet in Abwesenheit von Imidazol, was zeigt, dass Protonenkatalyse für die Reaktion entscheidend ist. Die Isolierung der neuen Verbindung und ihre vollständige Charakterisierung (Hintergrundinformationen) zeigte, dass es sich bei der neuen Verbindung in der Tat um das TBS-geschützte dC handelte. Diese Verbindung wird durch De-

topologen [^{13}C$_2$]- bzw. [^{18}O]-dC und [^{15}N,^{13}C]-dC. Die erwartete exakte Masse für das doppelt ^{15}N-markierte [^{15}N$_2$]-dC war in allen diesen Kontrollexperimenten nicht detektierbar, auch wenn höhere Mengen an DNA für die Experimente verwendet und nachfolgend verdaut wurden (2.5 nmol dC versus 0.9 nmol dC, Abbildung 1 d,e). Wir analysierten nachfolgend den Verdau, der von **D1** nach Inkubation mit nuklearem Stammzellextrakt erhalten wurde. Überraschenderweise beobachteten wir schon nach 1 h der Inkubation das Auftreten von [^{15}N$_2$]-markiertem dC. Diese Verbindung eluierte mit einer Retentionszeit von 7.18 min. Das erhaltene Massenspektrumsignal zeigt die exakte Masse $m/z_{\text{gef.}} =$ 230.0913 (Abbildung 1 c). Das Signal ist von den Signalen der anderen natürlich vorkommenden Isotopologen [^{15}N,^{13}C]-dC ($m/z = 230.0976$) und [^{18}O/^{13}C$_2$]-dC ($m/z = 230.1016$) gut se-

Abbildung 2. a) Reaktion von caC mit NaBH₄ ergibt Decarboxylierung: 1) NaBH₄, MeOH, RT, 20 min; 2) LiOH, MeCN/H₂O, RT, 2 h; 3) DDQ (2,3-Dichlor-5,6-dicyano-1,4-benzochinon), 1,4-Dioxan, RT, 25 h. b) Behandlung von caC enthaltender DNA mit Cystein alleine oder in Kombination mit Histidin oder Arginin führt zur Decarboxylierung. Die Analyse wurde nach Totalverdau der behandelten DNA mithilfe hochaufgelöster HPLC-MS durchgeführt. Illustration des Mechanismus der Decarboxylierung, die durch einen nukleophilen Angriff auf die C6-Position ausgelöst wird.

sucht. Abbildung 2b (links) zeigt die Daten, die für die Cystein/Histidin-Kombination erhalten wurden. Die Daten, die in Abbildung 2b (rechts) dargestellt sind, zeigen die Resultate aus dem Cystein/Arginin-Experiment. In beiden Fällen wird das Signal für die Verbindung [¹⁵N₂]-dC mit einem

m/z-Wert von 230.0922 klar nachgewiesen. Die Menge an decarboxyliertem Produkt nimmt mit der Menge an zugegebenen Thiol zu und betrug in unseren Fällen ca. 5 % (halbgesättigte Lösung von Aminosäuren, 50 °C, über Nacht).

Zusammenfassend zeigen wir mithilfe von Isotopenverfolgungsexperimenten mit DNA-Strängen, in welche wir doppelt ¹⁵N-markiertes caC eingebaut haben, dass Stammzellextrakte die Fähigkeit haben, caC zu decarboxylieren. Chemische Modellstudien belegen, dass diese Decarboxylierung eine Absättigung der C5-C6-Doppelbindung benötigt. Eine solche Absättigung ist z.B. durch einen nukleophilen Angriff auf die elektrophile C6-Position von caC z.B. durch ein Thiol möglich. Das so an C6 umgesetzte caC wird schnell decarboxylieren und nachfolgend rearomatisieren. Wir wollen erwähnen, dass auch die Bildung von mC durch Methyltransferasen (DNMTs) einem ähnlichen Mechanismus folgt. Hier erfolgt nach dem Angriff an der C6-Position durch ein Cystein die Reaktion des Produkts mit einem CH₃⁺-Äquivalent. Darüber hinaus muss Erwähnung finden, dass DNMTs bereits mit Demethylierungen in Verbindung gebracht worden sind; so können sie aus hmC Formaldehyd eliminieren.[17,25] Eine Decarboxylierung mithilfe von Methyltransferasen konnte bisher nicht nachgewiesen werden. Eine weitere enzymatische Reaktion, bei der die Absättigung der C5-C6-Doppelbindung zu einer Decarboxylierung führt, ist im Fall der Synthese von Uracil aus Isoorotat zu beobachten. Es ist bekannt, dass diese Decarboxylierung durch das Enzym Isoorotat-Decarboxylase (IDCase) durchgeführt wird. In diesem Enzym wird der nukleophile Angriff durch ein OH⁻ ausgeführt, das durch ein Zn²⁺-aktiviertes Wassermolekül bereitgestellt wird.[26,27] Der hier vorgestellte chemische Mechanismus ist daher in Einklang mit anderen Enzymreaktionen. Wir glauben, dass die Decarboxylierung von caC wichtig für die aktive Demethylierung in Stammzellen ist.

carboxylierung und anschließende Eliminierung des zuvor addierten Thiols gebildet. Um zu untersuchen, ob diese Chemie kompatibel mit der Decarboxylierung von caC in Oligonukleotiden ist (Abbildung 2b,c), wendeten wir erneut die Isotopenverfolgunsmethode an. Für das Experiment inkubierten wir den [¹⁵N₂]-caC enthaltenden DNA-Strang **ODN2** mit einer Mischung verschiedener Aminosäuren. Es wurde beobachtet, dass bei Zugabe der Aminosäuren Cystein und Histidin/Arginin zu einer wässrigen Lösung von **ODN2** Decarboxylierung stattfand (Abbildung 2b). Im Fall der Oligonukleotidexperimente reichte auch Cystein alleine aus, vermutlich weil die ungeschützte Aminogruppe die Rolle des Imidazols übernehmen kann. Um die Decarboxylierung zu beweisen, wurde die DNA über Nacht behandelt und nachfolgend bis auf das Nukleosid-Niveau verdaut. Diese Mischung wurde erneut mit hochaufgelöster HPLC-MS unter-

[1] H. Cedar, Y. Bergman, *Nat. Rev. Genet.* **2009**, *10*, 295–304.
[2] M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **2008**, *9*, 465–476.

[3] H. Sasaki, Y. Matsui, *Nat. Rev. Genet.* **2008**, *9*, 129 – 140.

[4] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929 – 930.

[5] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930 – 935.

[6] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146 – 7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008 – 7012.

[7] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303 – 1307.

[8] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300 – 1303.

[9] K. I. Ladwein, M. Jung, *Angew. Chem.* **2011**, *123*, 12347 – 12349; *Angew. Chem. Int. Ed.* **2011**, *50*, 12143 – 12145.

[10] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *ChemBioChem* **2011**, *12*, 206 – 222.

[11] S. K. Ooi, T. H. Bestor, *Cell* **2008**, *133*, 1145 – 1148; M. R. Branco, G. Ficz, W. Reik, *Nat. Rev. Genet.* **2012**, *13*, 7 – 13.

[12] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101 – 1105.

[13] M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Scholer, J. Walter, *EMBO J.* **2010**, *29*, 1877 – 1888.

[14] P. Hajkova, S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, M. A. Surani, *Science* **2010**, *329*, 78 – 82.

[15] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67 – 79.

[16] K. A. Haushalter, M. W. T. Stukenberg, M. W. Kirschner, G. L. Verdine, *Curr. Biol.* **1999**, *9*, 174 – 185.

[17] Z. Liutkevičiutė, G. Lukinavicius, V. Masevičius, D. Daujotytė, S. Klimašauskas, *Nat. Chem. Biol.* **2009**, *5*, 400 – 402.

[18] Z. Liutkeviciute, E. Kriukienė, I. Grigaitytė, V. Masevičius, S. Klimašauskas, *Angew. Chem.* **2011**, *123*, 2138 – 2141; *Angew. Chem. Int. Ed.* **2011**, *50*, 2090 – 2093.

[19] J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes Sr., A. A. Axhemi, *Biochim. Biophys. Acta Gen. Subj.* **2005**, *1723*, 256 – 264.

[20] C. J. LaFrancois, J. Fujimoto, L. C. Sowers, *Chem. Res. Toxicol.* **1998**, *11*, 75 – 83.

[21] G. H. Clever, C. Kaul, T. Carell, *Angew. Chem.* **2007**, *119*, 6340 – 6350; *Angew. Chem. Int. Ed.* **2007**, *46*, 6226 – 6236.

[22] J. Asakura, M. J. Robins, *J. Org. Chem.* **1990**, *55*, 4928 – 4933.

[23] G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *474*, 398 – 402; P. M. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew. Chem.* **2008**, *120*, 8478 – 8487; *Angew. Chem. Int. Ed.* **2008**, *47*, 8350 – 8358.

[24] N. K. Thalji, W. E. Crowe, G. L. Waldrop, *J. Org. Chem.* **2009**, *74*, 144 – 152.

[25] R. Métivier, R. Gallais, C. Tiffoche, C. Le Peron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, *Nature* **2008**, *452*, 45 – 50.

[26] A. Liu, H. Zhang, *Biochemistry* **2006**, *45*, 10407 – 10411.

[27] R. D. Palmatier, R. P. McCroskey, M. T. Abbott, *J. Biol. Chem.* **1970**, *245*, 6706 – 6710.

3.4 Desaminierungs-, Oxidations-, und C–C-Bindungsbruch-Reaktivitäten von 5hmdC, 5fdC und 5cadC

Stefan Schiesser,[#] Toni Pfaffeneder,[#] Keyarash Sadeghian, Benjamin Hackner, Barbara Steigenberger, Arne S. Schröder, Jessica Steinbacher, Gengo Kashiwazaki, Georg Höfner, Klaus T. Wanner, Christian Ochsenfeld und Thomas Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593−14599. *Deamination, Oxidation and C–C bond cleavage reactivity of 5-hydroxy-methylcytosine, 5-formylcytosine and 5-carboxycytosine.* ([#] geteilte Erstautorenschaft)

Prolog

Viele Mechanismen wurden für die aktive DNA-Demethylierung von 5mdC und seinen Oxidationsprodukten vorgeschlagen und untersucht (siehe Abschnitt 1.6). Darunter fallen Desaminierungs-, Oxidations- und C−C-Bindungsbruch-Reaktionen von 5hmdC, 5fdC und 5cadC. Dieses Manuskript befasst sich erstens mit der Sensitivität dieser Verbindungen bzgl. Desaminierung und Oxidation unter physiologischen Bedingungen. Da in der vorangegangen Studie in Abschnitt 3.3 eine Decarboxylierungsaktivität von 5cadC in mES-Zellen festgestellt werden konnte, werden in diesem Manuskript die mechanistischen Grundlagen sowie die Effizienz der C−C-Bindungsbruchreaktionen von 5hmdC, 5fdC und 5cadC im direkten Vergleich untersucht. Hierfür wurden enzymatische Bedingungen mit Imidazol als Histidin-Ersatz und β-Mercaptoethanol als Cystein-Ersatz simuliert. Die Untersuchungen gelangen mit Hilfe der Etablierung einer umfassenden Massenspektrometrie-basierten Methode (UHPLC gekoppelt mit einem ESI-Tandemmassenspektrometer), welche die Quantifizierung aller Reaktionsprodukte in nur einem analytischen Lauf ermöglichte.
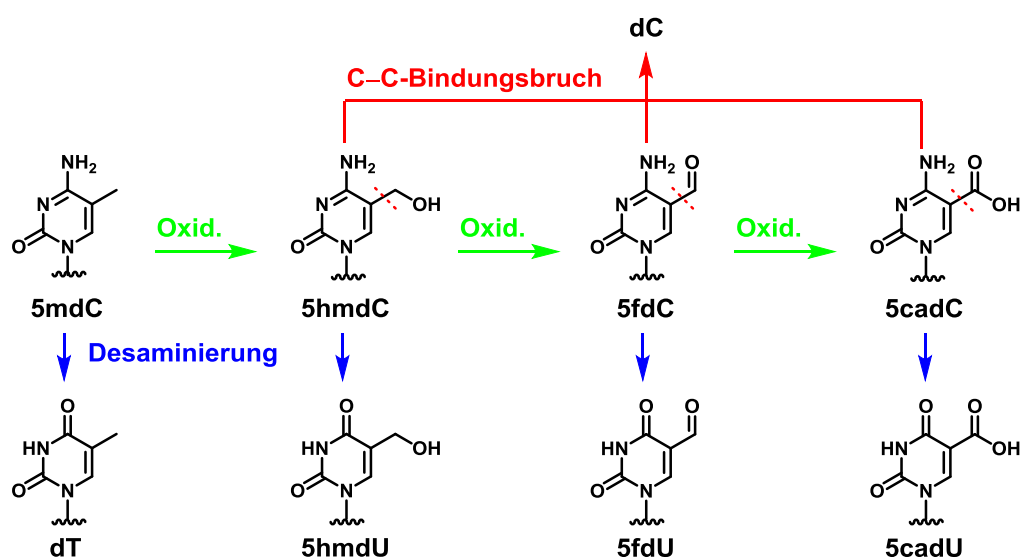
Autorenbeitrag

Für dieses Manuskript etablierte ich ein neues MS-System in unserem Arbeitskreis (UHPLC-UV-ESI-MS/MS). Ich synthetisierte die internen MS-Standards [^{15}N$_2$]-5fdC, [^{15}N$_2$]-5cadC, [^{15}N$_2$]-dU, [D$_3$]-dT, [D$_2$]-5hmdU sowie [^{15}N$_2$]-5fdU, entwickelte alle Probenpräparations- und MS-Methoden aller Reaktionsprodukte, plante die Experimente mit *Stefan Schiesser*, führte alle Messungen und Auswertungen durch. *Stefan Schießer* synthetisierte die Oligodesoxynukleotide für die *Assays* unter Mitarbeit von *Barbara Steigenberger* und *Arne S. Schröder*. Er führte ferner die Reaktionskinetiken von Oligodesoxynukleotiden und Nukleosiden durch und synthetisierte die MS-Standards [D$_2$,^{15}N$_2$]-5hmdC und 5fdU. *Benjamin Hackner* synthetisierte den MS-Standard [^{15}N$_2$]-dC. *Jessica Steinbacher* synthetisierte [^{15}N$_2$]-5fdC im Zuge ihrer Masterarbeit. *Keyarash Sadeghian* führte die quantenchemischen Berechnungen durch. *Georg Höfner* unterstützte die Arbeit als MS-Experte in analytischen Fragen und half bei der Etablierung initialer MS-Methoden.

Lizenz

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

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

[†]Center for Integrated Protein Science (CiPS^M) at the Department of Chemistry, [‡]Chair for Theoretical Chemistry at the Department of Chemistry, [§]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 modifiying 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.
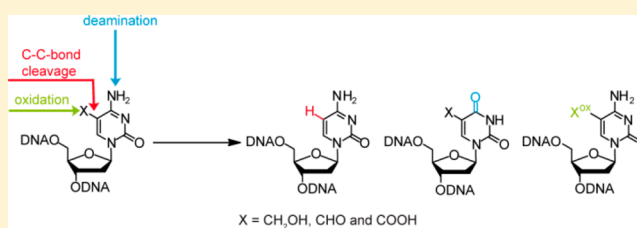
INTRODUCTION

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 α-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).[1−5] 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.[6,7] However, newer *in vitro* data suggest that deamination of hmdC is unlikely to occur enzymatically,[8] raising the question of whether the occurrence of hmdU may result from nonenzymatic spontaneous deamination of hmdC. Furthermore, since hmdC and fdC are known oxidative lesions of mdC[9−11] and because the levels of fdC and in particular cadC[3−5,12] are in the range of those reported for 8-oxo-dG,[13] 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
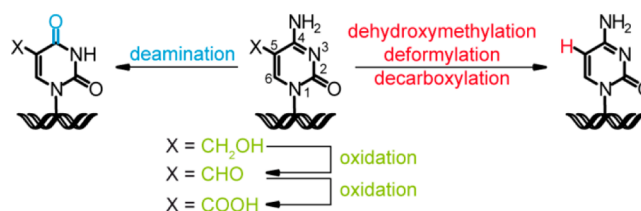
during DNA isolation and analysis.[13,14] Finally, while there is a possibility that enzymatic dehydroxymethylation of hmdC, deformylation of fdC, and decarboxylation of cadC might occur,[15−19] it cannot be ruled out that these processes occur

already to a significant extent without the help of an enzyme.[18,20]

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).[21−25] 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[a]

ODN1: X=hmdC   ODN4: X=hmdC
ODN2: X=fdC   ODN5: X=fdC
ODN3: X=cadC   ODN6: X=cadC
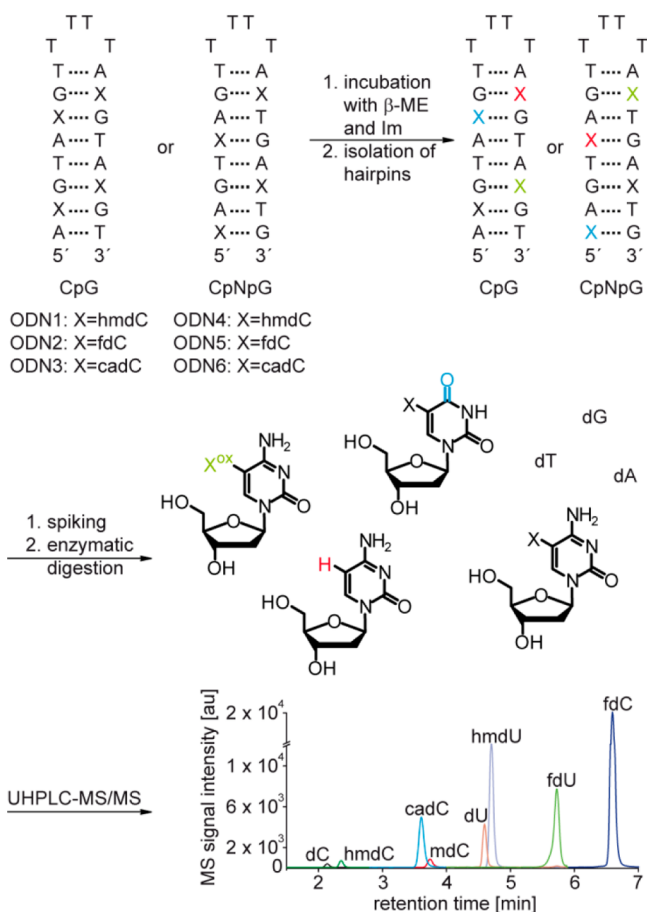
[a]After incubation of the xdC-containing hairpins in β-mercaptoethanol (β-ME)/imidazole, the DNA was isolated. The corresponding isotopologues were added (omitted for clarity in the chromatogram), and the DNA was digested. Quantification of the reaction products was performed by UHPLC-MS/MS. Im = imidazole.

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).[18] Examples are the DNA methyltransferases (DNMT), which convert dC to mdC.[19]

We first studied the reactivity of the different xdC-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 β-mercaptoethanol (β-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.[17,18] 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.[12,13,26,27] For the synthesis of the used isotopologues $[^{15}N_2]$-dC, $[D_2,^{15}N_2]$-hmdC, $[^{15}N_2]$-fdC, $[^{15}N_2]$-cadC, $[D_3]$-dT, $[^{15}N_2]$-dU, $[D_2]$-hmdU, and $[^{15}N_2]$-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).
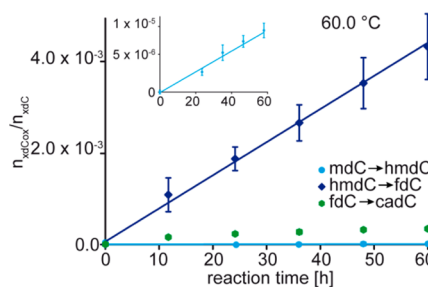
Figure 1. Oxidation kinetics of mdC to hmdC (cyan), hmdC to fdC (blue), and fdC to cadC (green) at 60.0 °C at pH 7.4. While hmdC is efficiently oxidized to fdC, fdC gives only a little cadC. Depicted are the means of triplicate experiments; error bars reflect the standard deviations. For details of the 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[a]

	$k_{37\,°C}$ [s^{-1}]	$k_{60\,°C}$ [s^{-1}]	E_a [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 ± 2.2

[a]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 β-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 μm 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 β-ME in ODN1 (hmdC). The fdC compound reached surprisingly high levels of up to 20% at 12 M β-ME (80% v/v β-ME; Figure 2, blue curve), showing that the C5–C6
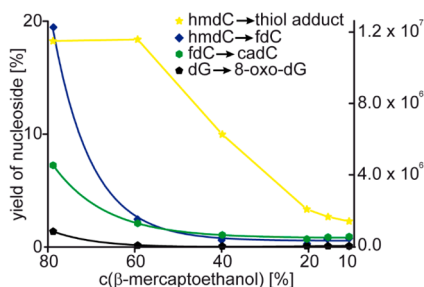
Figure 2. Investigation of the oxidation reactions of either hmdC or fdC in a hairpin-oligonucleotide (CpG-ODN 1, 2) with increasing concentration of β-mercaptoethanol/imidazole (pH 5.0, 50 °C, 48 h). Reaction yields (normalized to dT [%]) are plotted against the concentration of β-mercaptoethanol [% 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.[28]

Treatment of ODN2 (fdC) at 80% v/v β-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 = β-Mercaptoethanol R = CH$_2$CH$_2$OH
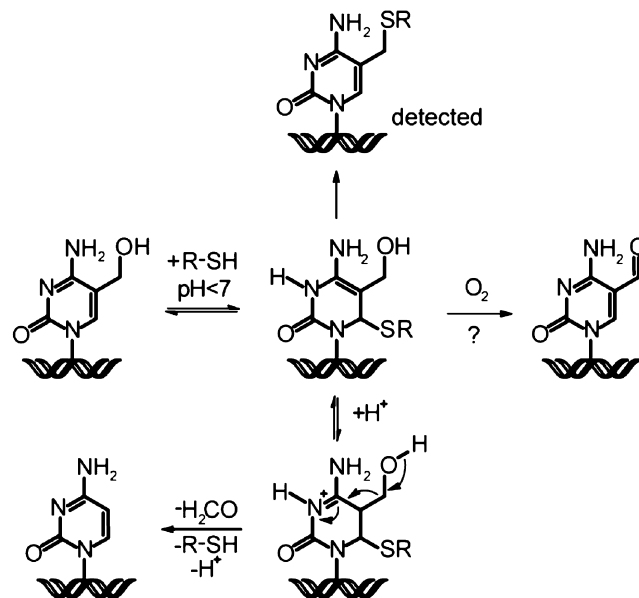
comparatively slow (Figure 2, green curve). The formation of a dithioacetal of fdC was not detected (see SI). It should be noted that concomitant deformylation of fdC or decarboxylation of the oxidation product cadC to form dC in the presensce of β-ME (see last section) may cause a slight underestimation 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
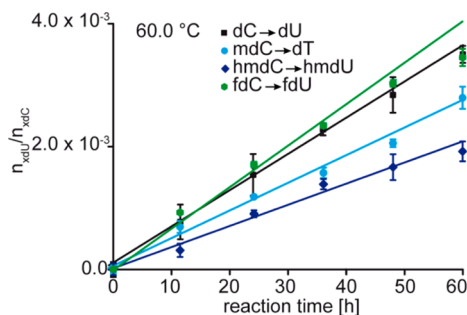
Figure 3. Deamination kinetics of dC to dU (black), mdC to dT (cyan), hmdC to hmdU (blue), and fdC to fdU (green) at 60.0 °C, pH 7.4. Depicted are the means of triplicate experiments; error bars reflect the standard deviations. For details of the kinetic measurements at 67.5, 75.0, and 82.5 °C and linear 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

	$k_{37\,°C}$ $[s^{-1}]$	E_a $[kJ\,mol^{-1}]$
dC→dU	$9.4 \pm 0.5 \times 10^{-10}$	108.7 ± 1.9
mdC→dT	$7.8 \pm 0.3 \times 10^{-10}$	105.0 ± 2.5
hmdC→hmdU	$5.8 \pm 0.8 \times 10^{-10}$	104.8 ± 3.9
fdC→fdU	$1.2 \pm 0.2 \times 10^{-9}$	102.2 ± 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.[29] Bearing in mind that the rates are more than 2 orders of magnitude lower in double-stranded DNA,[29,30] 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 ($k_{37\,°C} = 1.3 \pm 0.1 \times 10^{-8}$ 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.[18] To investigate the conditions of the thiol-mediated decarboxylation reaction in more detail, we treated the cadC-containing ODN3 with increasing concentrations of β-ME for 48 h at different pH-values and different temperatures in the presence of imidazole.[18] 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 β-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
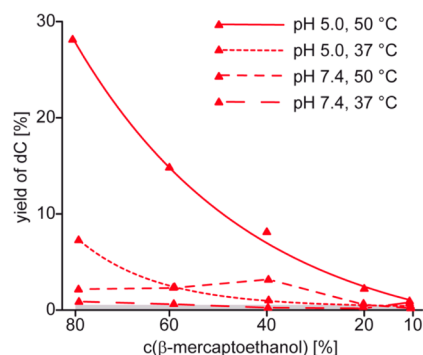
Figure 4. Investigation of the C−C bond cleavage reaction of cadC to dC in a hairpin-oligonucleotide (CpG-ODN 3) in a β-mercapto-ethanol/imidazole mixture. Depicted are the reaction yields (normalized to dT [%]) depending on the concentration of β-mercaptoethanol [% v/v] at pH = 5.0 or 7.4 and at 37 or 50 °C after 48 h. The gray area reflects the limit of quantification.
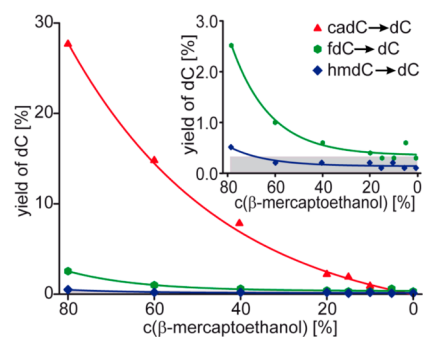
Figure 5. Investigation of the C−C bond cleavage reactions of either hmdC, fdC, or cadC in a hairpin-oligonucleotide (CpG-ODN 1, 2, 3) in a β-mercaptoethanol/imidazole mixture (pH 5.0, 50 °C, 48 h). Depicted are the reaction yields (normalized to dT [%]) of hmdC (blue), fdC (green), and cadC (red) to dC depending on the concentration of β-mercaptoethanol [% v/v]. The gray area in the inset shows the limit of quantification.

group, who used photohydration conditions.[16,17,20] 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
$k_{80\,°C}$ $[s^{-1}]$	$1.3 \pm 0.2 \times 10^{-9}$	$7.3 \pm 1.1 \times 10^{-9}$	$6.9 \pm 0.7 \times 10^{-9}$
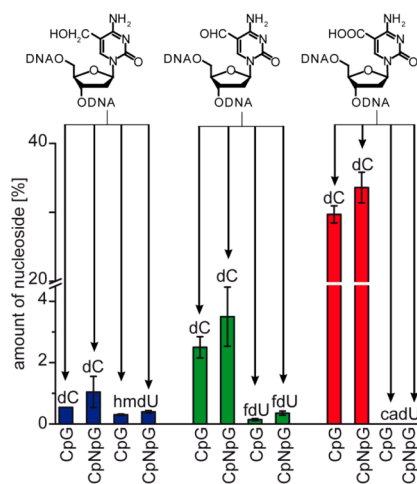
Figure 6. The C−C-bond cleavage of hmdC, fdC, and cadC to dC is almost independent of the sequence context. Investigation of the C−C bond cleavage and deamination reactions of either hmdC, fdC, or cadC in a hairpin-oligonucleotide (CpG-ODN 1, 2, 3; CpNpG-ODN 4, 5, 6) in a 80% (v/v) β-mercaptoethanol/imidazole 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[31] (data shown in SI). Although the energetics are clearly affected, the overall trend remains the same. Triple-zeta basis sets[32] were used throughout the calculations. RI-MP2[32,33] reaction energies, obtained using the DFT/B3LYP-D3[34−36] 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).[37,38] 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.[39] We obtained by our calculations a rather high energy of +34 kcal mol^{-1} 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^{-1} significantly higher. This may in part explain the observed weak activity[39] of IDCase to decarboxylate 5caC 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^{-1} for the decarboxylation of the thiol-reacted anionic intermediate. Overall, the reaction of the C6 position with the thiol reduces
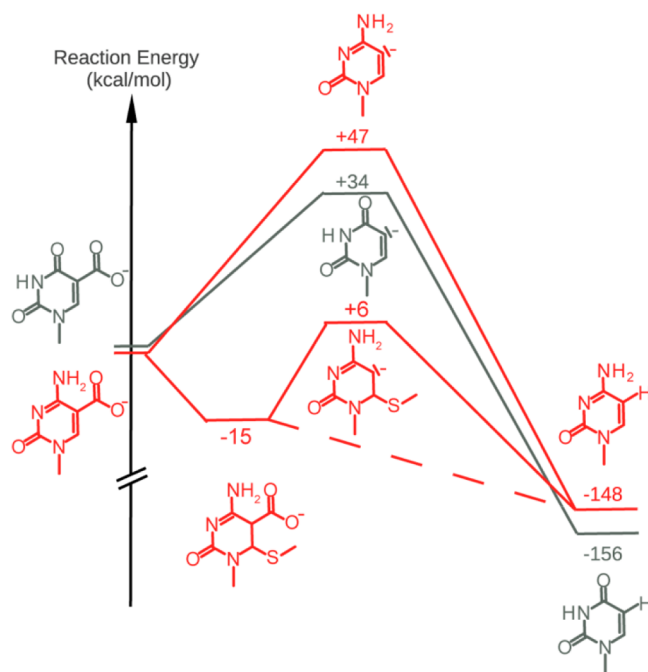
Figure 7. Schematic representation of different decarboxylation pathways. Depicted are the reaction energies obtained from quantum chemical calculations.

the energy of the corresponding anionic intermediate by more than 40 kcal mol^{-1} (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_2, 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[a]
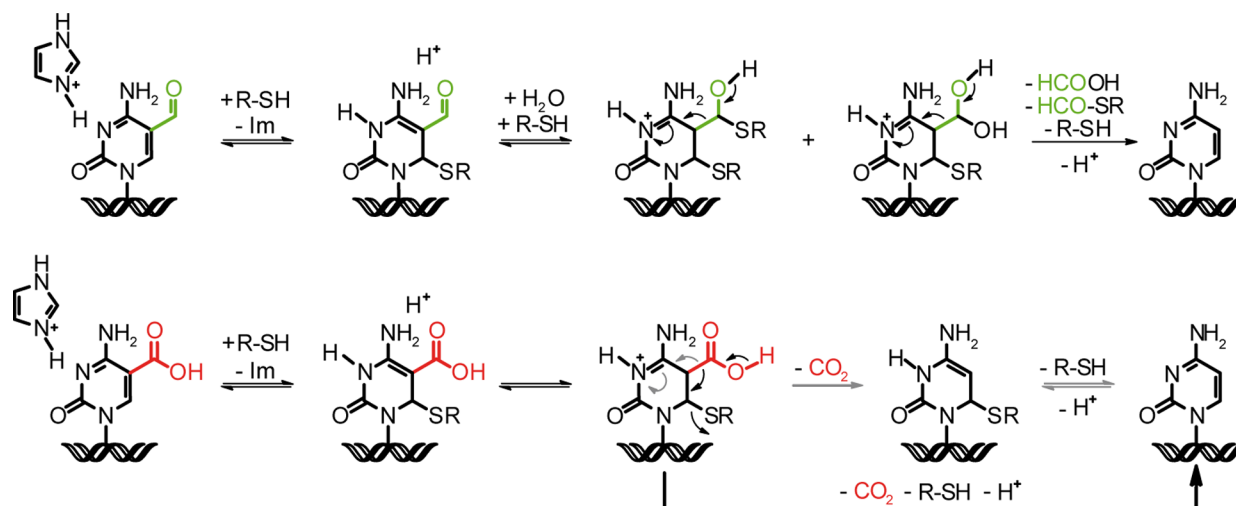
[a]The reactions are thought to proceed via a covalent enamine intermediate. The deformylation reaction of fdC further requires the addition of a nucleophile (water or thiol) to the aldehyde before the release of formic acid (or the thiol ester) can proceed. In principal, the release of formic acid from 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 = β-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

[⊥]S.S. and T.P. contributed equally.

Notes

The authors declare no competing financial interest.

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.; Hagemeier, 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.; Michalakis, 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.; Hagemeier, 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.; Welzmiller, V.; Michalakis, 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.

3.5 Dynamisch bindende Proteine für 5hmdC und seine oxidierten Derivate

Cornelia G. Spruijt, Felix Gnerlich, Arne H. Smits, Toni Pfaffeneder, Pascal W. T. C. Jansen, Christina Bauer, Martin Münzel, Mirko Wagner, Markus Müller, Fariha Khan, H. Christian Eberl, Anneloes Mensinga, Arie B. Brinkman, Konstantin Lephikov, Udo Müller, Jörn Walter, Rolf Boelens, Hugo van Ingen, Heinrich Leonhardt, Thomas Carell und Michiel Vermeulen, *Cell* **2013**, *152*, 1146−1159. *Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives.*

Prolog

Nach der Entdeckung der 5mdC-oxidierenden Funktion der Tet-Enzyme[5, 7-8] stellte sich die zentrale Frage, ob es sich bei den Oxidationsprodukten 5hmdC, 5fdC und 5cadC lediglich um Intermediate der DNA-Demethylierung handelt, oder ob mit diesen zusätzliche regulatorische Funktionen einhergehen. Um sich einer Antwort dieser Frage zu nähern, beschreibt dieses Manuskript die Identifikation von Proteinen, die mit den oxidierten Nukleobasen spezifisch und dynamisch wechselwirken, mittels Massenspektrometrie-basierter Proteomics.
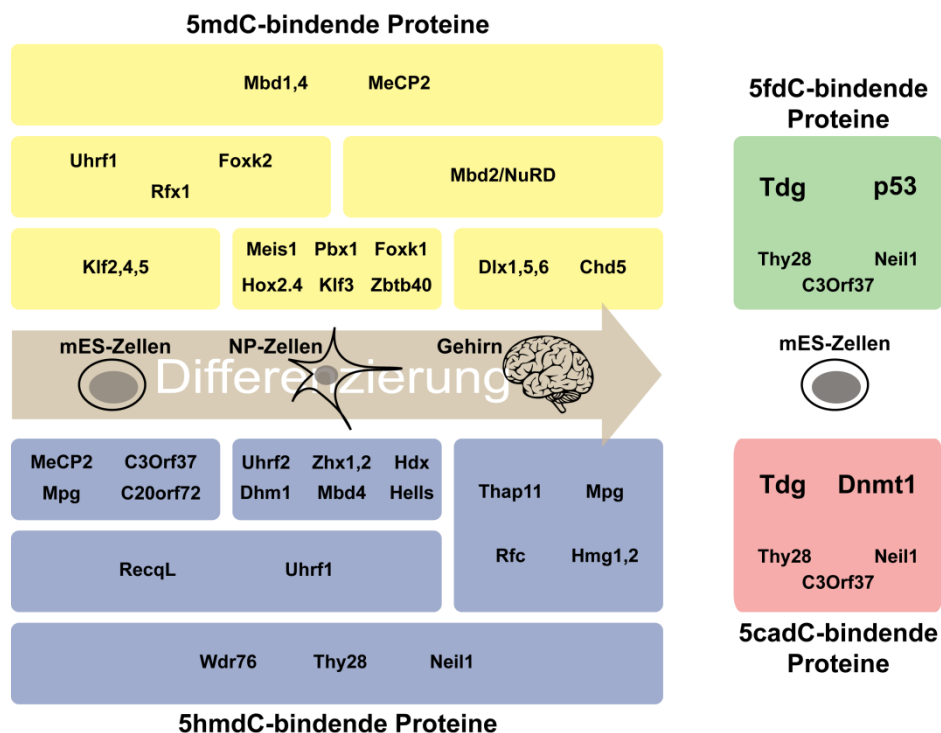
Autorenbeitrag

Ich entwickelte die quantitativen UHPLC-UV-ESI-MS/MS-Methoden für die Gehaltsbestimmung von 5mdC, 5hmdC, 5fdC und 5cadC in synthetischer und genomischer DNA. Hiermit verifizierte ich die Stabilität der eingesetzten DNA-Sonden in den Proteomics-Studien und untersuchte den Einfluss der identifizierten Binderproteine auf die genomischen Level der modifizierten Bausteine.

Lizenz

Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives

Cornelia G. Spruijt,[1,9] Felix Gnerlich,[2,9] Arne H. Smits,[1] Toni Pfaffeneder,[2] Pascal W.T.C. Jansen,[1] Christina Bauer,[3] Martin Münzel,[2] Mirko Wagner,[2] Markus Müller,[2] Fariha Khan,[4,5] H. Christian Eberl,[6] Anneloes Mensinga,[1] Arie B. Brinkman,[7] Konstantin Lephikov,[8] Udo Müller,[3] Jörn Walter,[8] Rolf Boelens,[5] Hugo van Ingen,[5] Heinrich Leonhardt,[3] Thomas Carell,[2,*] and Michiel Vermeulen[1,*]

[1]Department of Molecular Cancer Research, Proteomics and Chromatin Biology, UMC Utrecht, 3584 CG Utrecht, the Netherlands
[2]Center for Integrated Protein Science at the Fakultät für Chemie und Pharmazie, Ludwig-Maximilians-Universität München, 81377 Munich, Germany
[3]Center for Integrated Protein Science at the Fakultät für Biologie, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany
[4]University Institute of Biochemistry and Biotechnology, Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Rawalpindi, Pakistan
[5]NMR Spectroscopy Research Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands
[6]Proteomics and Signal Transduction, Max-Planck-Institut für Biochemie, 82152 Martinsried, Germany
[7]Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands
[8]Genetik/Epigenetik, Universität des Saarlandes, 66123 Saarbrücken, Germany
[9]These authors contributed equally to this work
*Correspondence: thomas.carell@lmu.de (T.C.), m.vermeulen-3@umcutrecht.nl (M.V.)
http://dx.doi.org/10.1016/j.cell.2013.02.004

SUMMARY

Tet proteins oxidize 5-methylcytosine (mC) to generate 5-hydroxymethyl (hmC), 5-formyl (fC), and 5-carboxylcytosine (caC). The exact function of these oxidative cytosine bases remains elusive. We applied quantitative mass-spectrometry-based proteomics to identify readers for mC and hmC in mouse embryonic stem cells (mESC), neuronal progenitor cells (NPC), and adult mouse brain tissue. Readers for these modifications are only partially overlapping, and some readers, such as Rfx proteins, display strong specificity. Interactions are dynamic during differentiation, as for example evidenced by the mESC-specific binding of Klf4 to mC and the NPC-specific binding of Uhrf2 to hmC, suggesting specific biological roles for mC and hmC. Oxidized derivatives of mC recruit distinct transcription regulators as well as a large number of DNA repair proteins in mouse ES cells, implicating the DNA damage response as a major player in active DNA demethylation.

INTRODUCTION

Methylation of cytosine residues at carbon atom 5 of the base (mC) represents a major mechanism via which cells can silence genes. Cytosine methylation mostly occurs in a CpG dinucleo-tide context. However, CpG islands (CGIs), which are characterized by a very high CpG density and are often found in promoter regions of genes, are typically hypomethylated. Methylation of these CGIs results in transcriptional silencing. The molecular mechanisms underlying the association between DNA methylation and repression of transcription have proven difficult to decipher. The classic view is that methylation of DNA results in the recruitment of methyl-CpG-binding proteins (MBPs) that possess transcriptionally repressive enzymatic activities (Defossez and Stancheva, 2011). However, in vivo validation for this model on a genome-wide level is still lacking. In contrast, recent in vivo data have revealed that CXXC-domain-containing proteins specifically bind to nonmethylated cytosines. In this case, hypomethylated CGIs serve as a recruitment signal for CXXC-domain-containing activators that establish a transcriptionally active chromatin state (Thomson et al., 2010).

It was discovered 4 years ago that Tet enzymes convert mC to 5-hydroxymethylcytosine (hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). This modification is particularly abundant in the brain and in embryonic stem cells but is detectable in all tissues tested (Globisch et al., 2010; Szwagierczak et al., 2010). Tet enzymes can catalyze further oxidation of hmC to 5-formylcytosine (fC) and 5-carboxylcytosine (caC) (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011). fC and caC can subsequently serve as substrates for thymine-DNA glycosylase (Tdg), which eventually results in the generation of a nonmethylated cytosine (He et al., 2011; Maiti and Drohat, 2011). Therefore, this Tet-Tdg pathway represents an active DNA demethylation pathway. It is not clear whether hmC, fC, and caC have additional DNA-demethylation-independent
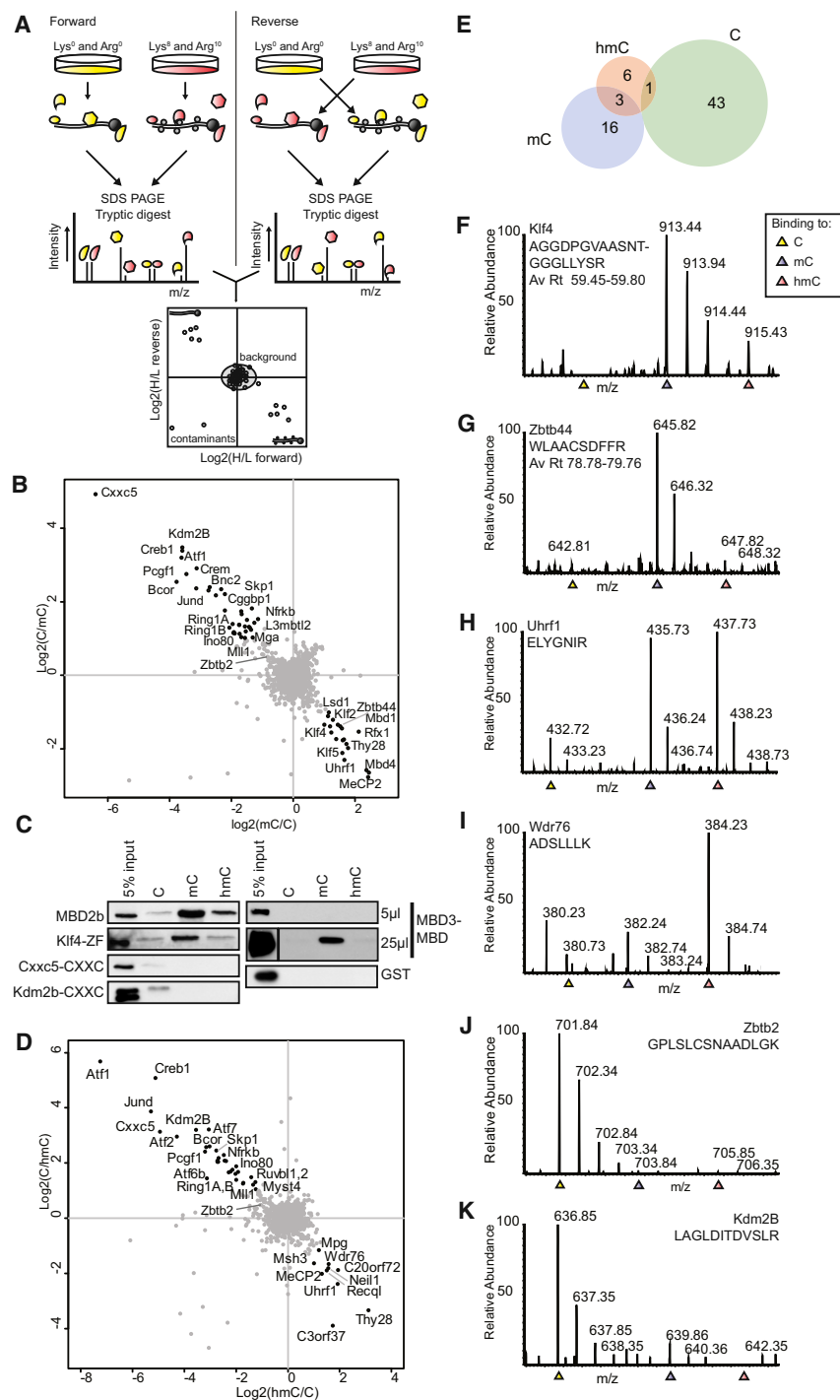
Figure 1. Identification of mC- and hmC-Specific Readers in Mouse Embryonic Stem Cells

(A) Schematic overview of the workflow.

(B) Scatterplot of a SILAC-based mC DNA pull-down in mESC nuclear extracts.

(C) Validation of the mC-specific binding of Klf4 and nonmethyl-C-specific binding of Cxxc5 and Kdm2b. DNA pull-downs were performed with recombinant GST-fusion proteins followed by western blotting. For MBD3_25, an empty lane was cut out.

(D) Scatterplot of a SILAC-based hmC DNA pull-down in mESC nuclear extract.

(E) Venn diagram showing overlap of readers for C, mC, and hmC.

(F–L) Representative mass spectra obtained in the triple-SILAC DNA pull-down in mESCs. Each spectrum shows the relative affinity of the indicated peptides and proteins for nonmethylated (yellow), methylated (blue), and hydroxymethylated (red) DNA.

See also Figure S1 and Table S1.

data reveal that each cytosine modification recruits a distinct and dynamic set of proteins. The known biology of these interacting proteins suggests a role for hmC, fC, and caC in active DNA demethylation pathways via base excision repair (BER), as well as an epigenetic recruitment function in certain cell types.

RESULTS

Identification of mC and hmC Readers in mESCs

To identify readers for methylcytosine and its oxidized derivatives, we made use of a DNA pull-down approach combined with quantitative MS. In brief, nuclear extracts from mESCs grown in "light" or "heavy" SILAC medium were incubated with a nonmodified or modified double-stranded DNA sequence (5′-AAG.ATG.ATG.AXG.AXG.AXG.AXG.ATG.ATG-3′, with X representing C, mC, or hmC; "forward" pull-down; Figure 1A). As a control, a label-swap, or "reverse," experiment was performed. Following incubation and washes, beads were combined and bound proteins were in-gel digested with trypsin and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Raw MS data were analyzed using MaxQuant (Cox and Mann, 2008). Specific interactors are distinguishable from background proteins by their H/L ratio. Proteins binding selectively to the modified DNA have a high ratio in the forward pull-down and a low ratio in the reverse pull-down, whereas readers for the nonmodified DNA show opposite

functions, as very few specific binders, or "readers," for these oxidized versions of mC have been described thus far.

We applied quantitative mass spectrometry (MS)-based proteomics to identify a large number of readers for mC and its oxidized derivatives in mouse embryonic stem cells (mESCs). Furthermore, we also identified readers for mC and hmC in neuronal progenitor cells (NPCs) and in adult mouse brain. Our

binding (low forward ratio, high reverse ratio). Background proteins will have a ~1:1 ratio in both pull-downs (Figure 1A).

As shown in Figure 1B and Table S1 available online, we identified 19 proteins enriched for mC compared to C in mESC nuclear extracts (p < 0.05 and ratio >2 in both pull-downs). Among these are the methyl-CpG-binding proteins MeCP2, Mbd1, Mbd4, and Uhrf1 (Defossez and Stancheva, 2011). Other interactors include Rfx1 and Zfhx3, which were previously identified as mC readers (Bartke et al., 2010; Sengupta et al., 1999). Interestingly, three Klf proteins were identified as mC readers: Klf2, -4, and -5. These proteins carry three Krüppel-like zinc fingers, just like the Kaiso family of mC-binding proteins. Klf4 is one of the four Yamanaka reprogramming factors and has not been previously identified as a mC-binding protein in HeLa or U937 cells (Bartels et al., 2011; Bartke et al., 2010). This may be due to the low expression of Klf4 in differentiated cells relative to mESCs. We confirmed the direct binding of the Klf4 Krüppel-like zinc fingers to mC using recombinant protein and two different DNA sequences (Figure 1C and S1A). A motif bearing similarities to a recently published consensus binding site for Klf4, as determined by ChIP-seq (GGGXGTG) (Chen et al., 2008), revealed that Klf4 binds this motif with the highest affinity when "X" is mC (Figure S1A). These results establish Klf4 as a sequence-specific mC binding protein.

Mining published bisulfite sequencing data of mESCs and NPCs (Stadler et al., 2012) and overlapping this data with the Klf4 ChIP-seq profile in mESCs (Chen et al., 2008) revealed a substantial number of methylated Klf4-binding sites in this cell type (Figure S1B), which are mainly intronic and intergenic (Figure S1C). Out of the 7,321 Klf4-binding sites in mESCs that were covered in the bisulfite sequencing data set, 1,356 show high levels of DNA methylation in mESCs (18.5%). Many of these Klf4-binding sites contain a methylated Klf4-binding motif, such as GGCGTG (Figures S1D and S1E). Interestingly, many Klf4-binding sites that are nonmethylated in ES cells become hyper-methylated in NPC cells (Stadler et al., 2012) (Figures S1B and S1D). This finding may be highly relevant in the context of Klf4-mediated cellular reprogramming. During reprogramming, Klf4 may be able to bind these methylated loci in differentiated cells to initiate stem-cell-specific gene expression patterns. Enrichment analyses for functional domains among the mC interactors revealed DNA-binding zinc fingers to be significantly enriched (Benj.Hoch.FDR = $10^{-2.45}$; Figure S3A). These zinc fingers may also interact with the methylated DNA in a sequence-specific manner.

In addition to the cluster of mC-binding proteins, a large number of proteins displayed preferential binding to nonmethylated DNA (Figure 1B, upper-left quadrant). Consistent with previous observations, this cluster of proteins contains a number of CXXC-domain-containing proteins that are known to preferentially bind to nonmethylated CpGs (Blackledge et al., 2010; Thomson et al., 2010). Examples include Cxxc5, Kdm2b, and Mll1 (see Figure 1C). We also identified other subunits of the Mll1 and PRC1.1 (Bcor) complexes, which most likely bind to the nonmethylated DNA indirectly via Mll1 and Kdm2b, respectively. Other interactors include the Ino80 chromatin-remodeling complex and zinc-finger-containing transcription factors such as Zbtb2, as well as basic leucine zipper-containing proteins

(enriched Benj.Hoch.FDR = $10^{-5.57}$; Figure S3A) such as JunD, Creb1, and Atf7, for which sequence-specific DNA binding is most likely abolished by DNA methylation.

Readers for hmC showed partial overlap with proteins observed to interact with mC (Figure 1D, lower-right quadrant, and Figure 1E), as only three proteins interacted with both modified baits: MeCP2, Uhrf1, and Thy28. Uhrf1 and MeCP2 are known to bind both mC and hmC, although MeCP2 clearly binds with a higher affinity to mC compared to hmC (Frauer et al., 2011; Hashimoto et al., 2012; Mellén et al., 2012). Thy28 is an uncharacterized protein that is associated with apoptosis (Toyota et al., 2012) and contains an EVE domain, which is possibly involved in (ds)RNA binding (Bertonati et al., 2009). Interestingly, two DNA glycosylases (Mpg and Neil3) and a helicase (Recql) were identified as hmC readers in mESCs. These proteins might be involved in active DNA demethylation pathways to convert hmC back to cytosine via base excision repair mechanisms, as has been suggested previously (Hajkova et al., 2010; Wossidlo et al., 2010). In addition, a number of previously uncharacterized proteins, Wdr76 and C3orf37, preferentially bound to hmC compared to C. We purified WDR76 as a GFP fusion protein from HeLa cells and found interactions with OCR, HELLS, and GAN (Figure S1F). The mouse protein Hells, or Lsh, is a DNA helicase that has previously been implicated in regulating DNA methylation levels in cells (Dennis et al., 2001). Interestingly, OCR, or Spindlin-1, is a protein that is known to bind trimethylated H3 lysine 4 (H3K4me3) (Bartke et al., 2010). A large number of proteins preferentially bound to the nonmodified DNA, as was observed for the mC pull-down (Figure S1G). We validated some of these findings using western blotting for endogenous proteins (Figure S1H).

To further investigate the relative affinity of proteins for C versus mC versus hmC in a single experiment, we made use of a triple pull-down approach (Vermeulen et al., 2010), in which mESCs are grown in three different SILAC media. "Light," "medium," and "heavy" nuclear extracts derived from these cells are incubated with C-, mC-, and hmC-containing DNA, respectively (Table S1). Quantitative MS is used to visualize the relative abundance of a protein in each of the three different pull-downs. This experiment confirmed most of the observations made in Figures 1B and 1D, although for some proteins, the ratios in the triple pull-down are lower. As shown in Figures 1F and 1G, Klf4 and Zbtb44 preferentially bind to the methylated DNA. Other proteins bind to both modified baits, such as Uhrf1 (Figure 1H). Kdm2b preferentially binds to the nonmodified DNA (Figure 1K). Contrary to a previous report (Yildirim et al., 2011), we did not observe a specific interaction between MBD3 and hmC (forward ratio, 0.448; reverse ratio, 1.823). We validated these observations using recombinant protein (Figure 1C). At higher concentrations of recombinant MBD3 protein, we observed a specific interaction with mC (Figure 1C), which is in agreement with a recent study that revealed that MBD3 has the highest affinity for mC compared to hmC and C (Hashimoto et al., 2012).

Taken together, these experiments reveal that mC and hmC both recruit distinct proteins in mESCs with little overlap. Furthermore, a large number of proteins preferentially bind to nonmodified DNA. The number of observed interactions with

hmC is moderate, and some of these suggest that hmC acts as an intermediate in active DNA demethylation pathways in mESCs.

fC and caC Recruit a Large Number of Proteins in Mouse Embryonic Stem Cells, Including DNA Glycosylases and Transcription Regulators

We also applied our SILAC-based DNA pull-down approach to identify readers for fC and caC in mESCs. Colloidal blue analysis revealed that the total amount of protein binding to each bait is similar (Figure S2A). Ratios of the forward and reverse pull-downs with hmC, fC, or caC were individually averaged, and these average ratios were then plotted against each other in two-dimensional graphs (Figures 2A–2C and Table S1). From these plots, it is clear that both fC (blue, purple, and green) and caC (yellow and green) recruit many more proteins than hmC does (red and purple). Strikingly, there is only limited overlap between fC and caC binders (green) (Figure 2D). One of the proteins that binds to fC and caC, but not to hmC, is Tdg, which is consistent with its reported substrate specificity (Maiti and Drohat, 2011). We validated this binding behavior using recombinant protein in electromobility shift assays (EMSA) (Figures 2E and 2F). We also purified GFP-Tdg from ES cells to identify Tdg interaction partners (Figure S2B and Table S1). None of the Tdg interactors were identified as specific readers in the fC and caC pull-down, indicating that these fC and caC interactions are Tdg independent. Another fC-specific reader is the p53 protein, which plays an important role in DNA damage response (Kastan et al., 1991). Interestingly, Dnmt1 specifically interacted with caC. This interaction was confirmed by EMSA as well as western blotting using an antibody against endogenous protein (Figures 2F and S2C). We also identified subunits of the Swi/Snf chromatin-remodeling complex, such as BAF170, as readers for caC. Three proteins bind to all oxidized derivatives of mC: Thy28, C3orf37, and Neil1. GO term enrichment for biological processes shows that fC significantly enriches for proteins that are related to DNA repair (Benj.Hoch.FDR = $10^{-2.71}$) (Figure S3A), whereas caC interactors are not enriched for any biological process. RNA-binding proteins, mitochondrial proteins, and other proteins that are less likely to be associated with regulation of gene expression or DNA repair binding were identified as binders for fC and caC (Table S3). Some of these may have a basic affinity for the formyl and carboxyl groups on the DNA strands, which are more reactive than methyl or hydroxymethyl. To exclude the possibility that many fC and caC interactors are binding to damaged or abasic DNA, we validated the homogeneity of the DNA strands using HPLC (Figure S2D). Furthermore, we analyzed the DNA before (blue) and after incubation (red) with mESC nuclear extract by MALDI-TOF-MS (Figure S2E). Quantification of the modified residues by LC-MS/MS shows that there is no significant loss of the modified bases after incubation with nuclear extract (Figure S2F). Figures 2A–2C also show that the group of proteins that bind preferentially to nonmodified cytosine (black, lower-left quadrant) shows a large overlap between the three pull-downs and contains the PRC1.1, Mll1, and Ino80 complexes. To compare the relative affinity of proteins for these three modifications in a single experiment, we performed a triple pull-down. Analyses of the triple pull-down ratios

for the identified fC and caC readers show similar trends, although some of the observed ratios are less prominent. As shown in Figures 2G–2L (and Table S1), the representative spectra of the indicated peptides of Tdg, Neil3, Mpg, Dnmt1, MeCP2, and Uhrf1 show relative ratios that are in agreement with ratios obtained in the independent experiments shown in Figures 2A–2C.

In summary, our data suggest that oxidized cytosine bases may induce a DNA damage response and trigger base excision repair pathways, which may finally result in DNA demethylation. In addition, each of these modifications recruits transcription regulators and other proteins that are not likely to be related to active DNA demethylation.

NPCs Contain a Distinct Set of mC and hmC Readers, Including Uhrf2, which Has the Highest Affinity for hmC

To investigate whether interactions with mC and hmC are dynamic during differentiation, we differentiated mESCs to NPCs. Nuclear extracts were generated from these cells followed by DNA pull-downs. Because no SILAC-compatible neurobasal medium is available, these experiments were performed using label-free quantification (LFQ) (Eberl et al., 2013; Hubner and Mann, 2011). Each DNA pull-down is analyzed separately and in triplicate. For all of the identified proteins (Table S1), we used ANOVA statistics (p = 0.025 and S_0 = 2) to compare the relative enrichment of proteins for each of the three baits. All significant outliers (192) were hierarchically clustered based on correlation after normalization by row mean subtraction (Figure 3A). Protein enrichment is indicated in red, whereas lack of enrichment is shown in blue. A large number of proteins bind to C or mC, whereas fewer proteins are specifically enriched in the pull-downs with hmC. Three smaller groups of proteins bind specifically to two of the baits (C/hmC, C/mC, or mC/hmC). As was observed in the DNA pull-downs from mESC nuclear extracts, CXXC-domain-containing proteins (Kdm2b and Mll, indicated in black) and their associated factors Bcor/Ring1a/b (blue) and Rbbp5/Ash2l (black) are enriched in the DNA pull-downs with nonmodified DNA relative to mC- and hmC-containing DNA. We identified Mbd2 and associated Mi-2/NuRD complex subunits as mC readers (indicated in yellow). Other identified MBD proteins include Mbd4, MeCP2, and Mbd1. Furthermore, a number of winged-helix (WH)-domain-containing proteins bound specifically to mC, including Rfx5 and its associated factors Rfxap and Rfxank (orange), which have previously been identified as methyl CpG interactors (Bartke et al., 2010).

Strikingly, these proteins bind more strongly to C compared to hmC. We further substantiated these observations by using recombinant protein (Figure 3B). This result indicates that, for some readers, oxidation of mC not only weakens the interaction, but also repels the mC interactor. The homeobox domain is significantly enriched in the cluster of mC-specific readers (Benj.Hoch.FDR = $10^{-1.8}$, Figure S3A), which is consistent with a previous study (Bartke et al., 2010). In addition, several known mC readers, such as Kaiso, Uhrf1, and Mbd4, bind both modified forms of cytosine. A number of DNA glycosylases bind specifically to hmC (Neil1, Neil3), as well as some helicases (Hells, Harp, Recql, and its homolog Bloom), which again suggests a
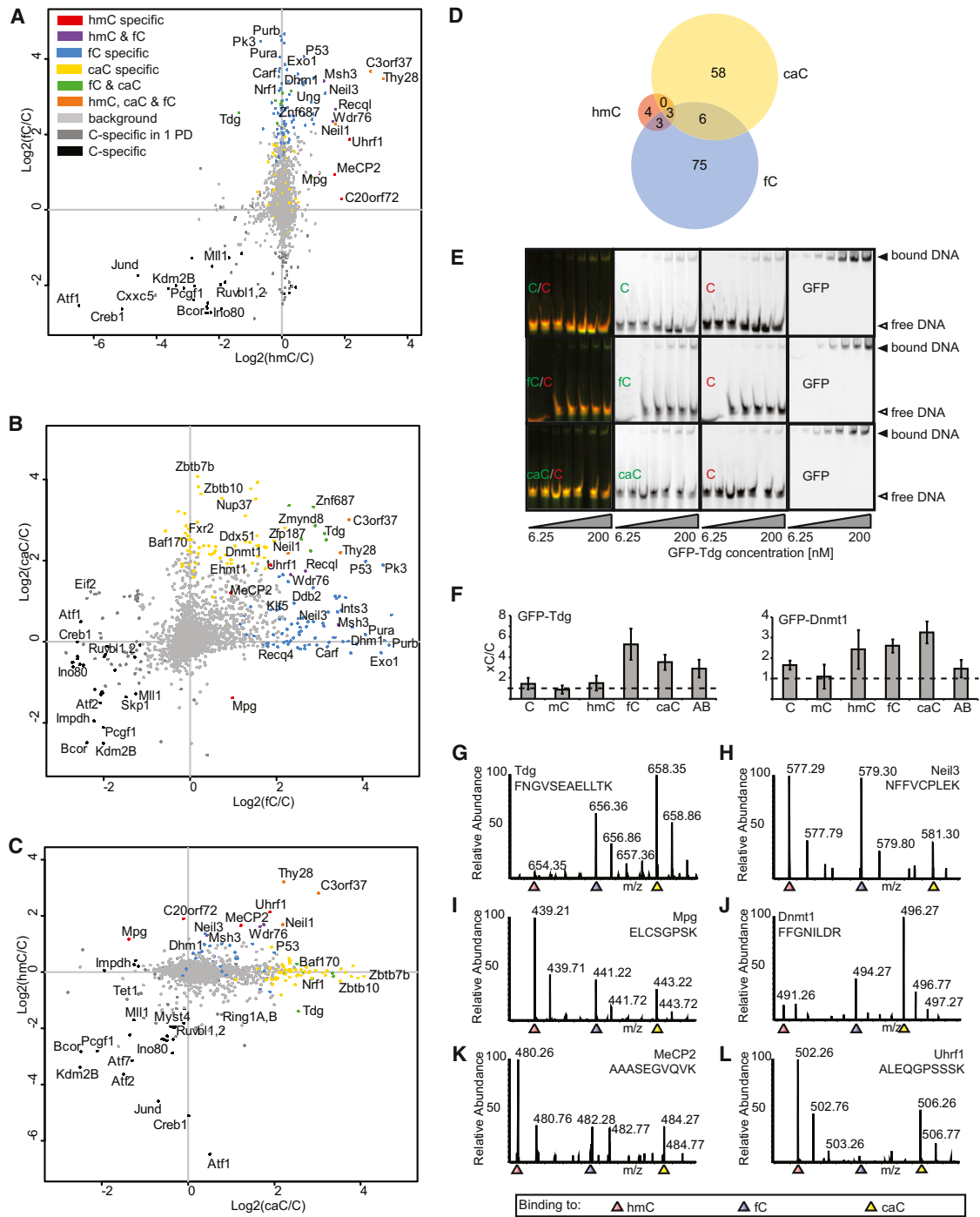
Figure 2. fC and caC Recruit a Large Number of Nonoverlapping Proteins in Mouse Embryonic Stem Cells

(A–C) Scatterplots of SILAC-based hmC, fC, and caC DNA pull-downs in mESC nuclear extract. The average ratio of all the identified and quantified proteins in the forward and reverse experiment for each of the three modifications is plotted on the X, Y, and Z axes of a three-dimensional cube. Shown in (A–C) are different side views of the cube. Colors indicate in which of the three pull-downs a protein was significantly enriched.

(D) Venn diagram showing the number of significantly enriched proteins for each of the baits.

(E) EMSA with GFP-Tdg at increasing protein concentrations (6.25–200 nM) incubated with dsDNA (250 nM of differentially labeled xC- and C-containing oligonucleotide, each).

(F) Electrophoretic mobility shift assays as shown in (E) performed with GFP-Tdg and GFP-Dnmt1 for all six residue variants (C, mC, hmC, fC, caC, and abasic site [AB]) in direct comparison to unmodified DNA. The binding preference was determined as the ratio of fluorescence signals of the different DNA substrates in the shifted bands. Shown are the means of three experiments; error bars represent SD.

(legend continued on next page)

DNA-repair-involved DNA demethylation pathway (GO DNA repair; Benj.Hoch.FDR = $10^{-3.91}$; Figure S3A). Although homeobox proteins are known to bind specifically to mC, a number of homeobox proteins show preferential binding to hmC in NPC extracts (examples include Zhx1 and -2). Finally, Uhrf2 was identified as a specific hmC-binding protein in NPCs, which we confirmed using recombinant protein (Figure 3B). Uhrf2 is not expressed in mESCs, and its levels increase upon differentiation (Pichler et al., 2011). This explains why Uhrf2 was not identified as an hmC-specific reader in mESC DNA pull-downs.

Taken together, these experiments reveal that interactions with mC and hmC are highly dynamic during differentiation. Furthermore, the observations made in NPCs strengthen our hypothesis that oxidation of mC serves as a trigger for active DNA demethylation. Nevertheless, some hmC-specific readers in NPCs do not appear to be linked to DNA repair mechanisms, indicating that, in these cells, hmC may also serve a role as a "classical" epigenetic mark that recruits transcriptional regulators.

NMR-Based Analysis of the Rfx5 WH Domain Bound to mC DNA

The specific interaction between the Rfx5 WH domain and mC DNA was studied in detail using solution nuclear magnetic resonance (NMR) spectroscopy in order to derive binding affinity and identify the mC-binding site. Addition of a singly methylated 18bp DNA fragment to the Rfx5-WH domain results in large changes in the ^{1}H-^{15}N HSQC "fingerprint" spectrum (Figure 3C). After addition of a slight molar excess of DNA, the spectrum does not show any further changes, indicating that Rfx5-WH strongly binds mC DNA and preferentially at only one of the two mC sites (Figure 3C). The affinity of Rfx5 for mC DNA was derived from the observed peak displacement for residues in the fast exchange regime, such as T104 and E102, assuming that the two mC are independent and equivalent, which resulted in an apparent dissociation constant $K_{D,app}$ of ~3 µM (with 95% probability limits 10 nM < K_D < 16 µM) (Figure 3D and Supplemental Information). Based on DNA pull-downs done with recombinant protein, which revealed a quantitative depletion of the WH domain from the lysate, we anticipate the K_D to be in the nM range (Figure 3B) To identify the residues that are responsible for specific mC binding, we used the DNA-bound Rfx1 WH domain crystal structure (PDB ID 1DP7; sequence identity 35%; Avvakumov et al., 2008; Gajiwala et al., 2000) to construct a homology model structure of Rfx5-WH and validated it against the experimental chemical shifts (data not shown). The homology model contains a hydrophobic pocket that includes residues with the largest chemical shift changes and is well aligned with an extended basic surface that is responsible for DNA binding in Rfx1. This binding pocket, formed by the side chains of K110, V113, Y114, T132, F135, L139, and Y169, is appropriately shaped to capture the mC base via a flip-out

mechanism, as seen in the case of UHRF1 (Figure 3E). Steric clashes introduced by the presence of an additional hydroxyl group could cause the observed specificity for mC. Given the apparent high affinity and DNA-sequence-independent binding to mC, we propose that the WH domain that is present in Rfx proteins is a bona fide mCpG-binding domain.

Brain-Specific Readers for mC and hmC Include Dlx Proteins

The adult brain is the organ with the highest levels of hmC (Globisch et al., 2010). Tet enzymes and hmC have been shown to play a role in active DNA demethylation of certain genes in this organ (Guo et al., 2011). To identify readers for C, mC, and hmC in the adult brain, nuclear extracts were prepared from this tissue, and these extracts were used for DNA pull-downs. LFQ was used to determine differential binders (Table S1). In brain extracts, we identified fewer specific readers compared to NPCs (108, p = 0.025 and S_0 = 0; Figure 4), most likely due to the presence of highly abundant structural proteins derived from connective tissue and extracellular matrix in these nuclear extracts. Interestingly, more proteins specifically bind to hmC compared to mC in brain extracts. This is in contrast to NPCs and mESCs, in which more interactions with mC relative to hmC are observed, which may imply a specific role for hmC in brain tissue.

The nonmodified DNA pull-down enriched for the same factors as those observed in mESCs and NPCs, including Cxxc5, Kdm2b, and Bcor (CXXC-domains indicated in black, PRC1 complex in blue, and Ino80 in red). In this case, mC DNA was bound by the Mbd2/NuRD complex, which contains the brain-specific ATPase Chd5 (Eberl et al., 2013; Potts et al., 2011) (indicated in yellow). Interestingly, we identified three distal-less homeobox proteins (Dlx1, -5, and -6) as specific mC interactors. Dlx proteins play a role in the development of the brain and are also expressed in specific regions of the adult brain (Jones et al., 2011; Wang et al., 2011). Wdr76 and Thy28 are hmC specific, as was also observed in NPCs. Thap11 (or Ronin) is identified as a brain-specific hmC reader. Interestingly, this protein is highly expressed in certain regions of the brain, including Purkinje cells (Dejosez et al., 2008). Finally, we identified all four subunits of replication factor C (Rfc2-5) and the associated factor Rfc1 as hmC-specific readers (indicated in green).

Altogether, these experiments further emphasize the dynamic nature of the mC and hmC interactomes during development.

Global Absolute Quantification of Protein Levels in mESCs, NPCs, and Adult Mouse Brain Extracts Reveals Expression-Level-Dependent and -Independent Interaction Dynamics

Our screening for mC- and hmC-specific readers in mESCs, NPCs, and adult mouse brain revealed a large number of cell-type- or organ-specific interactors (Figure S3B). The most

(G–L) Representative spectra of the indicated peptides obtained in the triple-labeled DNA pull-down in mESCs. Each spectrum shows the relative affinity of the indicated peptides and proteins for hmC-containing (red), fC-containing (blue), and caC-containing (yellow) DNA. Spectra are shown for Tdg (G), Neil3 (H), Mpg (I), Dnmt1 (J), MeCP2 (K), and Uhrf1 (L).

See also Figure S2 and Table S1.
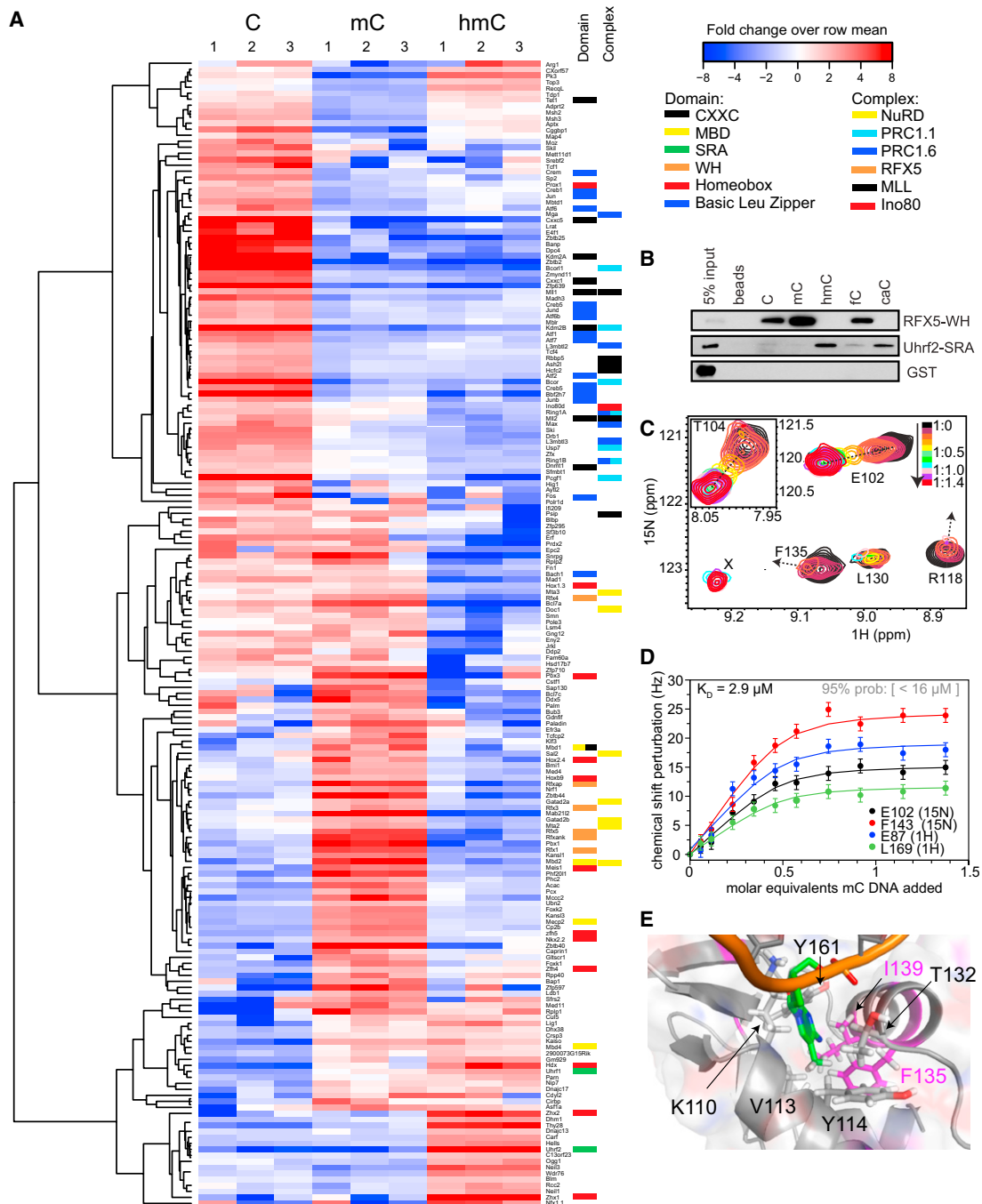
Figure 3. Hierarchical Clustering of NPC-Specific C, mC, and hmC Readers

(A) Correlation-based clustering of the LFQ intensities after log2 transformation and normalization by row mean subtraction. Included in the clustering are proteins that are significantly binding to at least one of the baits as determined by an ANOVA test. Blue indicates lack of enrichment, whereas enrichment is indicated in red. Domain and Complex columns indicate the DNA-binding domain(s) that may be responsible for direct binding to the bait and the complexes that readers are part of, respectively.

(B) Biochemical validation experiments using DNA pull-downs with recombinant DNA-binding domains.

(C) Overlay of Rfx5-WH HSQC spectra with increasing amounts of mC DNA added and color-coded on the indicated scale listing the WH domain:DNA ratio. Some residues, such as F135 and R118, cannot be unambiguously tracked to their bound states because their chemical shift changes are very large. Peaks corresponding to their bound state, such as "X," appear only after addition of a full molar equivalent of DNA.

(D) Selected binding curves and fits for resonances that are in the fast exchange regime throughout the titration. Error bars (SD) for the peak positions are set to 1.2 Hz.

(E) Close-up of the putative mC-binding pocket in the RFX5 WH domain. The methylated cytosine is indicated in green.

See also Table S1.
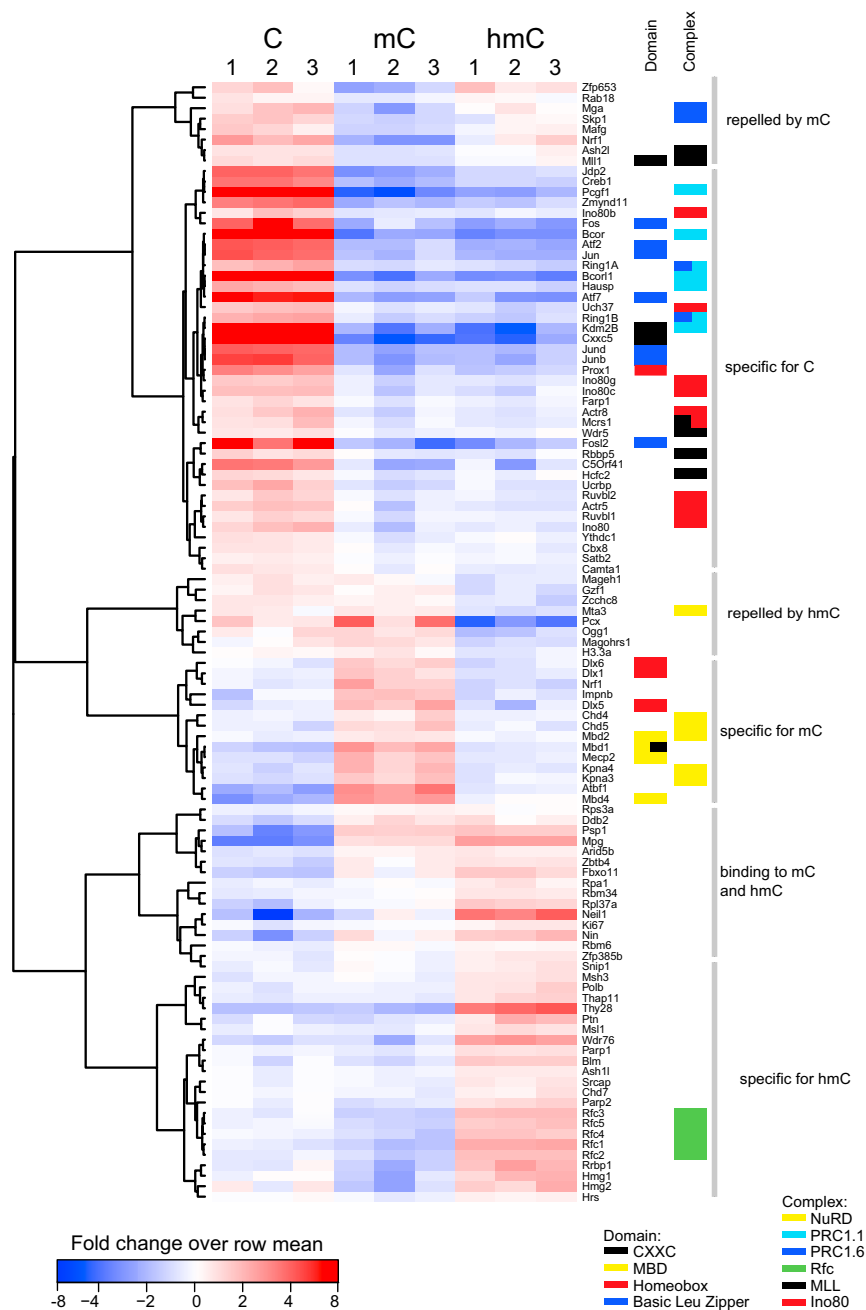
Figure 4. Hierarchical Clustering of Brain-Specific C, mC, and hmC Readers

Correlation-based clustering of the row-mean-subtracted LFQ intensities of proteins in C, mC, and hmC DNA pull-downs in adult mouse brain nuclear extracts.

See also Table S1.

The cluster of mESC-specific proteins is enriched for anchoring junction (Benj.Hoch.FDR = $10^{-2.96}$) and cell adhesion (Benj.Hoch.FDR = $10^{-2.14}$), whereas proteins in brain enriched GO terms such as synaptic transmission (Benj.Hoch.FDR = $10^{-3.77}$) and cognition (Benj.Hoch.FDR = $10^{-2.75}$), as expected (Figure S4C). The molar concentrations of proteins that are significantly enriched in one of the DNA pull-downs are spread over several orders of magnitude, indicating that our screening is not biased toward high-abundant proteins (Figure 5A). Of the 259 proteins that showed dynamic interactions through development (Table S3), 20 proteins were not quantified in the iBAQ measurements. The 74 proteins (~31%) that do show a correlation between interaction pattern and protein abundance in the different extracts can be divided into six clusters (Figure 5B). A correlation was defined as gaining or losing an interaction accompanied by at least a 2-fold change in protein abundance. An example of a protein that was identified as a specific (mC) reader only in mESCs was Klf4. As shown in Figure 5B, this protein is highly expressed in mESCs but is less abundant in NPCs or in the adult mouse brain. Another example is represented by the Dlx5 and Dlx6 proteins, which are highly abundant in brain nuclear extract and exclusively bind to mC in pull-downs from these extracts. For about 185 proteins, no correlation is observed between expression levels (at least 2-fold change) and binding behavior. For these proteins, the cause of differential binding may be explained through PTMs that affect the interaction between a reader and DNA or a differentially expressed cofactor. A good example of the latter is the Mi-2/NuRD complex. Although most of its subunits display equal expression levels in mESCs, NPCs, and brain, mC-specific interactions are not observed in mESCs. This can be explained by the fact that Mbd2, which is the direct reader of mC within the NuRD complex, is low in abundance in mESCs and is upregulated during differentiation (Figure 5B). Thereby, it controls the mC-specific binding of the entire complex. In mESCs, the majority of the Mi-2/NuRD complex contains Mbd3, which is

obvious explanation for these observed differential interactions is regulation of reader abundance at the protein level. Alternatively, the interaction between a reader and (modified) DNA may be affected by posttranslational modifications (PTMs). To investigate global absolute protein levels in the different nuclear extracts that were used for the pull-downs, we made use of a method called intensity-based absolute quantification (iBAQ) (Schwanhäusser et al., 2011). Approximately 8,000 proteins were quantified in at least one of the extracts (Table S2). All proteins with at least a 10-fold change in concentration were clustered based on their expression pattern (Figure S4B).

A — mESC, NPC, Brain scatter plots: Amount (fmoles/100ug (log10)) vs All / Significant

B — Clustered heatmaps:

Cluster I (n=8): Klf5, Klf4, Myst4, Bnc2, Nalp4f, Rmi1, Klf2, C20orf72

Cluster II (n=18): Prdx2, Rplp2, Pk3, Zasc1, Cggbp1, Ubn2, Polr1d, Pbx3, Hox2.4, Rreb1, Rcor2, Erf, Uhrf1, Efr3a, Bcl7c, Pole3, Acac, Ddp2

Cluster III (n=5): Blbp, Rfx4, Bmi1, Hox1.3, Hoxb9

Cluster IV (n=7): Ldb1, Sal2, Meis1, Pbx1, Ptn, Cbx8, Jun

Cluster V (n=8): Satb2, Zmynd11, RP23478C2.1, Dlx5, Zfp385b, Zbtb4, Chd5, Dlx6

Cluster VI (n=28): Ogg1, Mbtd1, Nfe2l1, Msl1, Fbxo11, Ash1l, Thap11, Mbd2, Camta1, Srcap, Mageh1, Hcfc2, Kpna4, Atbf1, Rbm6, Snip1, Ino80, Ino80g, Psp1, Ino80c, Hmga1, Ythdc1, Hrs, Kpna3, Zcchc8, Actr5, Ucrbp, Actr8

Fold change (log10) compared to mean abundance
−2 1 0 1 2 3

(legend on next page)

the MBD-containing protein that has lost its high-affinity mC binding ability. Furthermore, technical reasons for not identifying an interactor could be the presence of highly abundant structural proteins in the brain lysate or binding competition among different readers in the extracts. Altogether, the absolute quantification of protein abundance in the different nuclear extracts revealed large differences in protein levels between mESC, NPCs, and adult mouse brain. This data set serves as a rich resource on its own but also enables us to explain many of the differential interactions that we identified using quantitative MS-based interactomics.

Uhrf2 Stimulates the Sequential Activity of the Tet1 Enzyme

The first protein that was identified as an hmC binder was Uhrf1 (Frauer et al., 2011), a protein that is involved in maintenance of DNA methylation (Bostick et al., 2007). Our data revealed that Uhrf1 binds with a similar affinity to mC and hmC, which is consistent with previously published data (Frauer et al., 2011). This is in contrast to Uhrf2, which we identified as a high-affinity hmC-binding protein in NPC cells that shows a lower affinity for mC. The function of Uhrf2 is not well understood. It is clear, however, that Uhrf2 cannot rescue the phenotype of Uhrf1 knockout cells, which lose DNA methylation (Pichler et al., 2011; Zhang et al., 2011). Uhrf1 is highly expressed in mESCs, whereas Uhrf2 levels increase during differentiation (Table S3 and Pichler et al., 2011). Altogether, this prompted us to investigate whether Uhrf2 expression affects the levels of mC and its oxidized derivatives. The Tet1-catalytic domain was transfected into HEK293T cells with and without coexpression of Uhrf2. Total genomic DNA modification levels were determined using LC-MS/MS (Figure 6 and Supplementary Information). As shown in Figure 6D, Uhrf2 overexpression increases the level of hmC. More striking is the increase of fC and caC levels upon Uhrf2 coexpression together with the Tet1 catalytic domain. Because fC and caC serve as substrates for Tdg and BER, the detected increase in the levels of fC and caC following Uhrf2 expression may be an underestimation of the actual production of these bases. It therefore seems that Uhrf2 promotes repetitive oxidation of mC by the Tet proteins. We hypothesize that flipping the modified cytosine base out of the DNA double helix, as has been described for Uhrf1 binding to methylated and hydroxymethylated DNA (Arita et al., 2008; Frauer et al., 2011), may enhance accessibility of the hydroxymethylated base to the Tet enzymes, thereby promoting further oxidation.

DISCUSSION

In this study, we have used quantitative MS-based proteomics to identify readers for mC and its oxidized derivatives in mESCs,

as well as readers for mC and hmC in NPCs and adult mouse brain. Readers for individual modifications were found to be highly dynamic throughout the three cell types and tissues that we investigated (Figure 7). This is in contrast to interactions with histone modifications, such as trimethylated lysines on histone H3. For these modifications, the majority of interactors are constant between different cell types or developmental stages (Eberl et al., 2013 and M.V., unpublished data). Readers for distinct cytosine modifications show limited overlap. This indicates that, at least from a biochemical perspective, mC, hmC, fC, and caC behave quite differently. Although little overlap was observed with regard to proteins that interact with each of the epigenetic marks, they all repelled a common set of proteins, such as several CXXC-domain-containing proteins and their interactors. It remains to be determined which of the consequences of DNA (hydroxy)methylation is functionally most relevant: recruitment of transcriptionally repressive complexes or preventing the binding of certain (activating) proteins to unmodified DNA. A detailed biochemical characterization of the interactions and their dissociation constants will be important to answer this question.

Our experiments revealed a number of DNA glycosylases and DNA repair proteins that bind to hmC, fC, and caC, whereas we identified few such proteins binding to mC. The enriched binding of DNA-repair-associated proteins was most pronounced for fC. From this observation, one can conclude that the conversion of hmC to fC is a signal that is likely to result in repair-associated removal of the modified base by proteins that are rather ubiquitously expressed. It is therefore surprising that, in different cell types and tissues, rather constant levels of hmC, fC, and caC are found. The maintenance of such constant levels of these bases in mESCs may indicate a high turnover of DNA methylation, probably involving a constant "correction" by de novo methylation. Regardless, it will be important to investigate which mechanisms control Tet enzyme conversion of mC to hmC and further oxidation to fC and caC. Our data reveal that coexpression of Uhrf2 with the catalytic domain of Tet1 results in a (transient) upregulation of hmC, fC, and caC, indicating that Uhrf2 promotes the sequential oxidation of mC by Tet1. One of the other factors influencing the catalytic activity of the Tet enzymes is the concentration of cellular metabolites. It has been shown that oncometabolites such as 2-hydroxyglutarate can competitively inhibit the activity of 2-oxo-glutarate-dependent enzymes, such as the Tet proteins (Chowdhury et al., 2011; Xu et al., 2011). Furthermore, mutations in IDH1 and -2, which generate 2-oxo-glutarate, are phenocopied by mutations in the TET enzymes and result in cancer (Figueroa et al., 2010). Mutations in the IDH2 and TET2 genes were also linked to lower genomic hmC levels and altered gene expression patterns in myeloid cancers (Ko et al.,

---

Figure 5. Global Absolute Protein Quantification in mESCs, NPCs, and Adult Mouse Brain

(A) Graphs indicating the concentration of all proteins identified in the nuclear extract (all) and the identified readers (significant) in each of the cell types. The gray area indicates the concentration at which protein quantification is inaccurate.

(B) Readers for which protein expression levels correlate with DNA binding patterns were clustered into six groups based on their expression in the three different nuclear extracts. The color indicates protein levels (white, low; red, high), whereas binding preference is indicated by C, mC, hmC, or combinations thereof.

See also Figure S4 and Table S2 and Table S3.
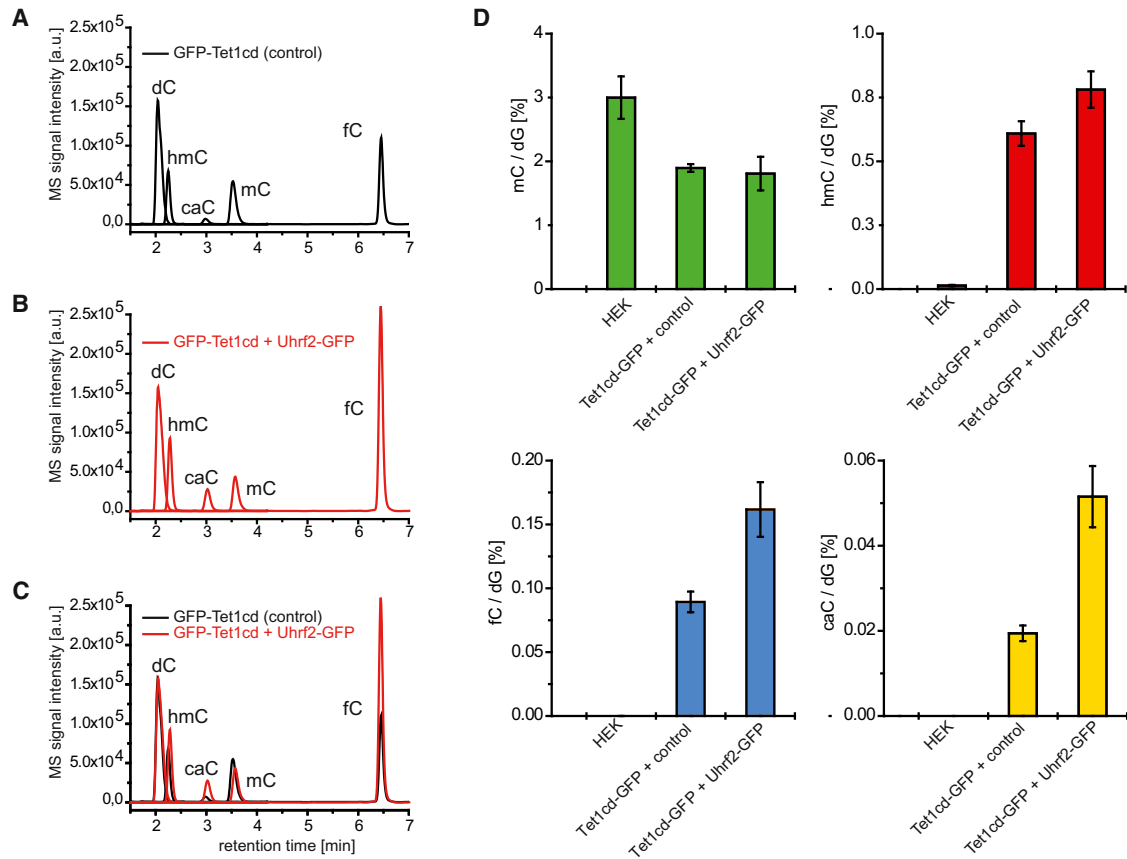
Figure 6. Stable Isotope-Dilution-Based LC-ESI-MS/MS Quantification of Cytosine Derivatives in HEK293T Cells

(A) Nonquantitative LC-MS/MS chromatogram of digested genomic DNA from HEK293T cells cotransfected with Tet1-catalytic domain-GFP (GFP-Tet1cd) and an unrelated expression construct (control). Depicted are the overlaid ion chromatograms of the MS/MS transitions for dC and the cytosine derivatives (black curves). dC, mC, and hmC were measured by a factor of $\sim10^2$–10^3 less sensitive in comparison to caC and fC.

(B) Same as (A) except that Uhrf2-GFP was coexpressed together with GFP-Tet1cd. The MS signal intensities were normalized to the dC content of (A).

(C) Superposition of (A) and (B).

(D) Levels of cytosine derivatives relative to the total cytosine content (dG) as determined by quantitative LC-MS/MS mass spectrometry.

Shown are the means of technical triplicates; error bars reflect SD.

2010; Konstandin et al., 2011). In support of these observations, which clearly link hmC to cancer, we noticed that many hmC, fC, and caC readers are implicated in cancer, including UHRF2, CARF, p53, and HELLS (Lee et al., 2000). Interestingly, mutations in the Hells helicase, which we identified as an hmC reader in NPCs, result in a decrease of DNA methylation levels in cells (Myant et al., 2011). It seems clear that regulating the levels of mC and its oxidized derivatives is essential for normal cell homeostasis and that deregulation of the readers, writers, and erasers of these marks results in a disturbance of the balance between cell proliferation and differentiation during development.

EXPERIMENTAL PROCEDURES

Cell Culture

IB10 mESCs were cultured in light (R^0K^0) or heavy ($R^{10}K^8$) SILAC medium in the presence of 2i compounds. For triple labeling, a third type of medium was used containing medium-labeled L-lysine (K^4) and L-arginine (R^6). mESCs were differentiated to NPCs in N2B27 medium and cultured in NSA medium, consisting of NSA MEM, 1% glutamine, 1× N2 supplement, 10 ng/ml bFGF, and 10 ng/ml EGF.

DNA Pull-Downs

Nuclear extracts were generated as described previously (Eberl et al., 2013; Vermeulen et al., 2010). DNA (see Table S4) immobilized on Dynabeads My-One C1 was incubated with nuclear extract in 50 mM Tris-HCl (pH 8), 150 mM NaCl, 1 mM DTT, 0.25% NP40, and complete protease inhibitors (Roche, EDTA-free) in the presence of poly-dAdT. After extensive washes (using incubation buffer without poly-dAdT), bound proteins were in-gel digested using trypsin. After sample preparation, peptides were desalted on Stage-tips (Rappsilber et al., 2003).

Mass Spectrometry

Peptides were separated using an EASY-nLC (Proxeon) connected online to an LTQ-Orbitrap Velos mass spectrometer (Thermo) as described (Smits et al., 2013). Raw data were analyzed using MaxQuant version 1.2.2.5 and searched against protein database ipi.MOUSE.v3.68.fasta. Using Perseus, data were filtered and scatter plots were made using R. The raw mass spectrometry data have been deposited to the ProteomeXchange Consortium

mESC

Klf2
Klf4
Klf5

Zbtb40
Foxk1
Hox2.4
Mab21l2
Meis1
Pbx1
Phf20l1
Klf3

NPC

NH₂
Uhrf1 Rfx
Foxk2

CH₃

N

Mbd1, 4,
MeCP2

O

N

Mbd2/NuRD

Dlx1
Dlx6
Dlx5
Chd5

Brain

mESC

Mecp2
C3orf37
C20orf72

Uhrf2
Hdx
Hells
Dhm1
Zhx1, 2
Mbd4

NPC

NH₂
RecqL Uhrf1

OH
CH₂

N

Wdr76
Thy28
Neil1

O

Neil1, Mpg

N

Thap11
Rfc
Hmg1
Hmg2
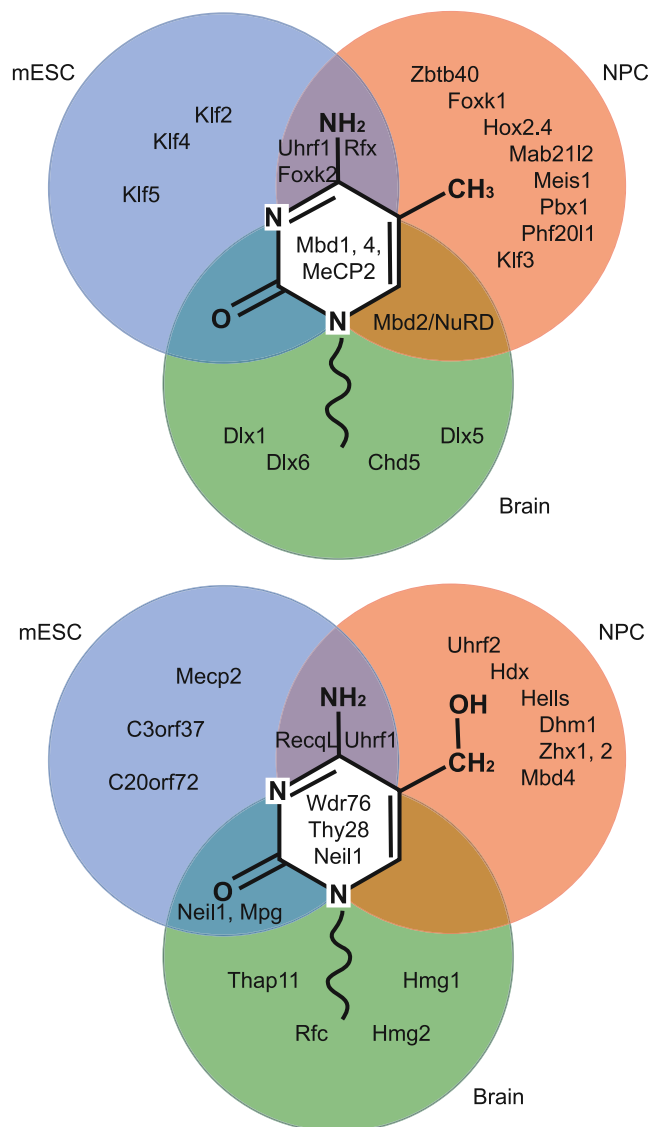
Brain

Figure 7. Visualization of Cell-Type-Specific and General mC and hmC Readers

Venn diagram showing examples of mC (A) and hmC (B) readers that were identified in mESCs (blue), NPCs (orange), and adult mouse brain (green). See also Figure S3 and Table S3.

(http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaíno et al., 2013) with the dataset identifier PXD000143.

Recombinant Protein Expression and DNA Pull-Downs

DNA-binding domains were cloned into the GST-containing PRP256NB vector. The Uhrf2(aa416–626) GST-fusion construct was kindly provided by Dr. Jiemin Wong. Protein expression was performed in *E. coli* BL21 codon⁺ cells. Bacterial lysate was cleared by ultracentrifugation. DNA pull-downs were performed as described above with the addition of 10 μM ZnCl₂ to the incubation buffer.

iBAQ

iBAQ was performed essentially as described in Schwanhäusser et al. (2011). A more detailed description can be found in the Extended Experimental Procedures.

LC-MS/MS Analysis of Genomic DNA

Cotransfections were performed in HEK293T cells, and genomic DNA was purified according to Münzel et al. (2010). Quantification of DNA nucleosides from genomic DNA is based on a further development of our isotope dilution method (Pfaffeneder et al., 2011 and data not shown). LC-MS/MS analysis was performed on an Agilent 6490 triple quadrupole mass spectrometer coupled to an Agilent 1290 UHPLC system. For general source- and compound-dependent parameters, see the Extended Experimental Procedures and Tables S5 and S6. The transitions of the nucleosides were analyzed in the positive-ion-selected reaction monitoring mode (SRM) operating MS1 and MS2 under unit mass resolution conditions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.02.004.

REFERENCES

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455, 818–821.

Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y.J., Duan, S.L., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature 455, 822–825.

Bartels, S.J.J., Spruijt, C.G., Brinkman, A.B., Jansen, P.W.T.C., Vermeulen, M., and Stunnenberg, H.G. (2011). A SILAC-based screen for Methyl-CpG binding proteins identifies RBP-J as a DNA methylation and sequence-specific binding protein. PLoS ONE 6, e25884.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzar-ides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell 143, 470–484.

Bertonati, C., Punta, M., Fischer, M., Yachdav, G., Forouhar, F., Zhou, W.H., Kuzin, A.P., Seetharaman, J., Abashidze, M., Ramelot, T.A., et al. (2009). Structural genomics reveals EVE as a new ASCH/PUA-related domain. Proteins 75, 760–773.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. Mol. Cell *38*, 179–190.

Bostick, M., Kim, J.K., Estève, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760–1764.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W.W., Jiang, J.M., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Chowdhury, R., Yeoh, K.K., Tian, Y.M., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K.H., Li, X.S., Woon, E.C.Y., Yang, M., et al. (2011). The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. EMBO Rep. *12*, 463–469.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Defossez, P.A., and Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. Prog. Mol. Biol. Transl. Sci. *101*, 377–398.

Dejosez, M., Krumenacker, J.S., Zitur, L.J., Passeri, M., Chu, L.F., Zhou, S.Y., Thomson, J.A., and Zwaka, T.P. (2008). Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. Cell *133*, 1162–1174.

Dennis, K., Fan, T., Geiman, T., Yan, Q.S., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev. *15*, 2940–2944.

Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. Mol. Cell *49*, 368–378, Published onlne November 29, 2012. http://dx.doi.org/10.1016/j.molcel.2012.10.026.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell *18*, 553–567.

Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., and Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PLoS ONE *6*, e21306.

Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature *403*, 916–921.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS ONE *5*, e15367.

Guo, J.U., Su, Y.J., Zhong, C., Ming, G.L., and Song, H.J. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423–434.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science *329*, 78–82.

Hashimoto, H., Liu, Y.W., Upadhyay, A.K., Chang, Y.Q., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X.D. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q.Y., Ding, J.P., Jia, Y.Y., Chen, Z.C., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Hubner, N.C., and Mann, M. (2011). Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). Methods *53*, 453–459.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Jones, D.L., Howard, M.A., Stanco, A., Rubenstein, J.L.R., and Baraban, S.C. (2011). Deletion of Dlx1 results in reduced glutamatergic input to hippocampal interneurons. J. Neurophysiol. *105*, 1984–1991.

Kastan, M.B., Onyekwere, O., Sidransky, D., Vogelstein, B., and Craig, R.W. (1991). Participation of p53 protein in the cellular response to DNA damage. Cancer Res. *51*, 6304–6311.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839–843.

Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P.M., Schneider, S., et al. (2011). Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. Leukemia *25*, 1649–1652.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930.

Lee, D.W., Zhang, K.J., Ning, Z.Q., Raabe, E.H., Tintner, S., Wieland, R., Wilkins, B.J., Kim, J.M., Blough, R.I., and Arceci, R.J. (2000). Proliferation-associated SNF2-like gene (PASG): a SNF2 family member altered in leukemia. Cancer Res. *60*, 3612–3622.

Maiti, A., and Drohat, A.C. (2011). 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.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. Cell *151*, 1417–1430.

Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmiller, V., Michalakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew. Chem. Int. Ed. Engl. *49*, 5375–5377.

Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J.I., and Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Res. *21*, 83–94.

Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeier, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed. Engl. *50*, 7008–7012.

Pichler, G., Wolf, P., Schmidt, C.S., Meilinger, D., Schneider, K., Frauer, C., Fellinger, K., Rottach, A., and Leonhardt, H. (2011). Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. J. Cell. Biochem. *112*, 2585–2593.

Potts, R.C., Zhang, P.S., Wurster, A.L., Precht, P., Mughal, M.R., Wood, W.H., 3rd, Zhang, Y.Q., Becker, K.G., Mattson, M.P., and Pazin, M.J. (2011). CHD5, a brain-specific paralog of Mi2 chromatin remodeling enzymes, regulates expression of neuronal genes. PLoS ONE *6*, e24515.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. *75*, 663–670.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Sengupta, P.K., Ehrlich, M., and Smith, B.D. (1999). A methylation-responsive MDBP/RFX site is in the first exon of the collagen alpha2(I) promoter. J. Biol. Chem. *274*, 36649–36655.

Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res. *41*, e28.

Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., and Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. *38*, e181.

Tahiliani, M., Koh, K.P., Shen, Y.H., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930–935.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464, 1082–1086.

Toyota, H., Jiang, X.Z., Asakura, H., and Mizuguchi, J. (2012). Thy28 partially prevents apoptosis induction following engagement of membrane immunoglobulin in WEHI-231 B lymphoma cells. Cell. Mol. Biol. Lett. 17, 36–48.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 142, 967–980.

Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., et al. (2013). The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 41(Database issue), D1063–D1069.

Wang, B., Lufkin, T., and Rubenstein, J.L.R. (2011). Dlx6 regulates molecular properties of the striatum and central nucleus of the amygdala. J. Comp. Neurol. 519, 2320–2334.

Wossidlo, M., Arand, J., Sebastiano, V., Lepikhov, K., Boiani, M., Reinhardt, R., Schöler, H., and Walter, J. (2010). Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. EMBO J. 29, 1877–1888.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. Cancer Cell 19, 17–30.

Yildirim, O., Li, R.W., Hung, J.H., Chen, P.B., Dong, X.J., Ee, L.S., Weng, Z.P., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell 147, 1498–1510.

Zhang, J.Q., Gao, Q.Q., Li, P.S., Liu, X.L., Jia, Y.H., Wu, W.C., Li, J.W., Dong, S., Koseki, H., and Wong, J.M. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. Cell Res. 21, 1723–1739.

3.6 Tet-Enzyme oxidieren Thymin zu 5-Hydroxymethyluracil in murinen embryonalen Stammzellen

Toni Pfaffeneder[#], Fabio Spada[#], Mirko Wagner[#], Caterina Brandmayr, Silvia K. Laube, David Eisen, Matthias Truss, Jessica Steinbacher, Benjamin Hackner, Olga Kotljarova, David Schuermann, Stylianos Michalakis, Olesea Kosmatchev, Stefan Schiesser, Barbara Steigenberger, Nada Raddaoui, Gengo Kashiwazaki, Udo Müller, Cornelia G. Spruijt, Michiel Vermeulen, Heinrich Leonhardt, Primo Schär, Markus Müller, Thomas Carell, *Nat. Chem. Biol.* **2014**, *10,* 574–581. *Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.* ([#] geteilte Erstautorenschaft)

Prolog

Die aktive DNA-Demethylierung könnte, wie in Abschnitt 1.6 erläutert, per Desaminierung von 5hmdC zu 5hmdU und anschließender Basenexzisionsreparatur ablaufen.[313] 5hmdU ist in genomischer DNA aber auch ein oxidativer Schaden von dT (siehe Abschnitt 1.5), der durch reaktive Sauerstoffspezies (ROS) entsteht.[249] In diesem Manuskript werden die Entstehungswege von 5hmdU in (differenzierenden) mES-Zellen per quantitativer Massenspektrometrie und Isotopenverfolgung aufgeklärt. Dabei wurde aufgedeckt, dass der Großteil von 5hmdU in mES-Zellen aus der Tet-vermittelten Oxidation von Thymin stammt.
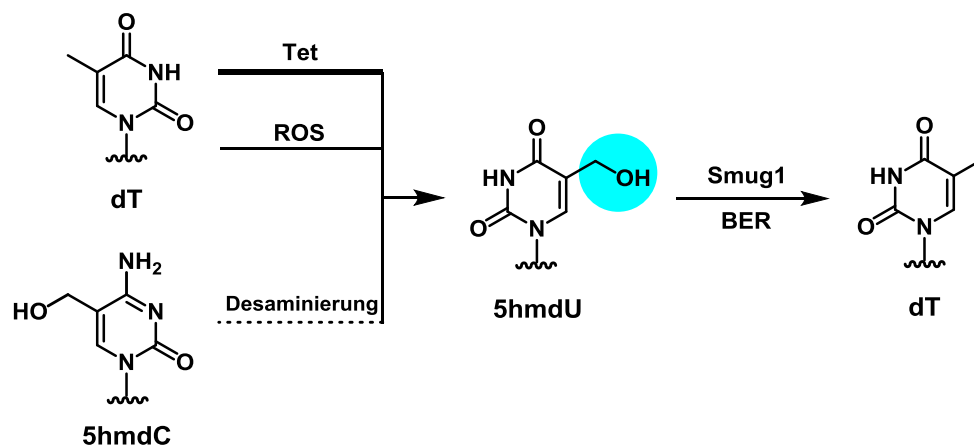
Autorenbeitrag

Für dieses Manuskript synthetisierte ich MS-Standards, entwickelte die Methoden der Probenvorbereitung und der UHPLC-UV-ESI-MS/MS-Analyse von allen DNA-Modifikationen (Charakterisierung, Validierung, Gehaltsbestimmung, Isotopenverfolgungs-experimente) und führte die Messungen der Stammzellexperimente sowie der Uhrf1/2-überexprimierten HEK-293T-Zellen durch. Ferner interpretierte ich alle Daten mit Hilfe mathematischer und statistischer Methoden. *Fabio Spada* führte die mES-Zelldifferenzier-ungen und Fütterungsexperimente mit Isotopen-markierten Verbindungen durch. *Mirko Wagner* und *Caterina Brandmayr* führten die HEK-293T-Experimente und die entsprechenden LC-MS/MS-Messungen durch. Für eine weitere Auflistung des Beitrags der übrigen Autoren siehe Manuskript.

Lizenz

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

Toni Pfaffeneder[1,8], Fabio Spada[1,8], Mirko Wagner[1,8], Caterina Brandmayr[1], Silvia K Laube[1], David Eisen[1], Matthias Truss[2], Jessica Steinbacher[1], Benjamin Hackner[1], Olga Kotljarova[1], David Schuermann[3], Stylianos Michalakis[4], Olesea Kosmatchev[1], Stefan Schiesser[1], Barbara Steigenberger[1], Nada Raddaoui[1], Gengo Kashiwazaki[1], Udo Müller[5], Cornelia G Spruijt[6], Michiel Vermeulen[6,7], Heinrich Leonhardt[5], Primo Schär[3], Markus Müller[1]* & Thomas Carell[1]*

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 cytosine-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 thymine 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.

Methylcytosine is an epigenetically important nucleobase associated with the control of transcriptional activity, genomic imprinting, X-chromosome inactivation and suppression of transposable elements[1]. Controlled formation and removal of mC at specific genomic loci is critical for correct genome programming or reprogramming during cellular differentiation[2]. Recently, it was discovered that Tet proteins (Tet1–3) oxidize mC to give the oxidized C-derived nucleobases hmC[3], 5-formylcytosine (fC)[4,5] and 5-carboxycytosine (caC)[5,6], whose biological functions are still yet unclear (**Fig. 1a**)[7]. As fC and caC are both removed by thymine DNA glycosylase (Tdg)[6,8], 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)[9,10]. It was recently suggested that hmU might also be produced by deamination of hmC, a hypothesis that remains controversial[11–14]. 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 Nthl1 (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[17,18]. 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 $[D_2]$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

[1]Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, München, Germany. [2]Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Berlin, Germany. [3]Department of Biomedicine, University of Basel, Basel, Switzerland. [4]Center for Integrated Protein Science at the Department of Pharmacy—Center for Drug Research, Ludwig-Maximilians-Universität München, München, Germany. [5]Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. [6]Department of Molecular Cancer Research, Cancer Genomics Netherlands, Utrecht, The Netherlands. [7]Present address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands. [8]These authors contributed equally to this work. *e-mail: markus.mueller@cup.uni-muenchen.de or thomas.carell@cup.uni-muenchen.de
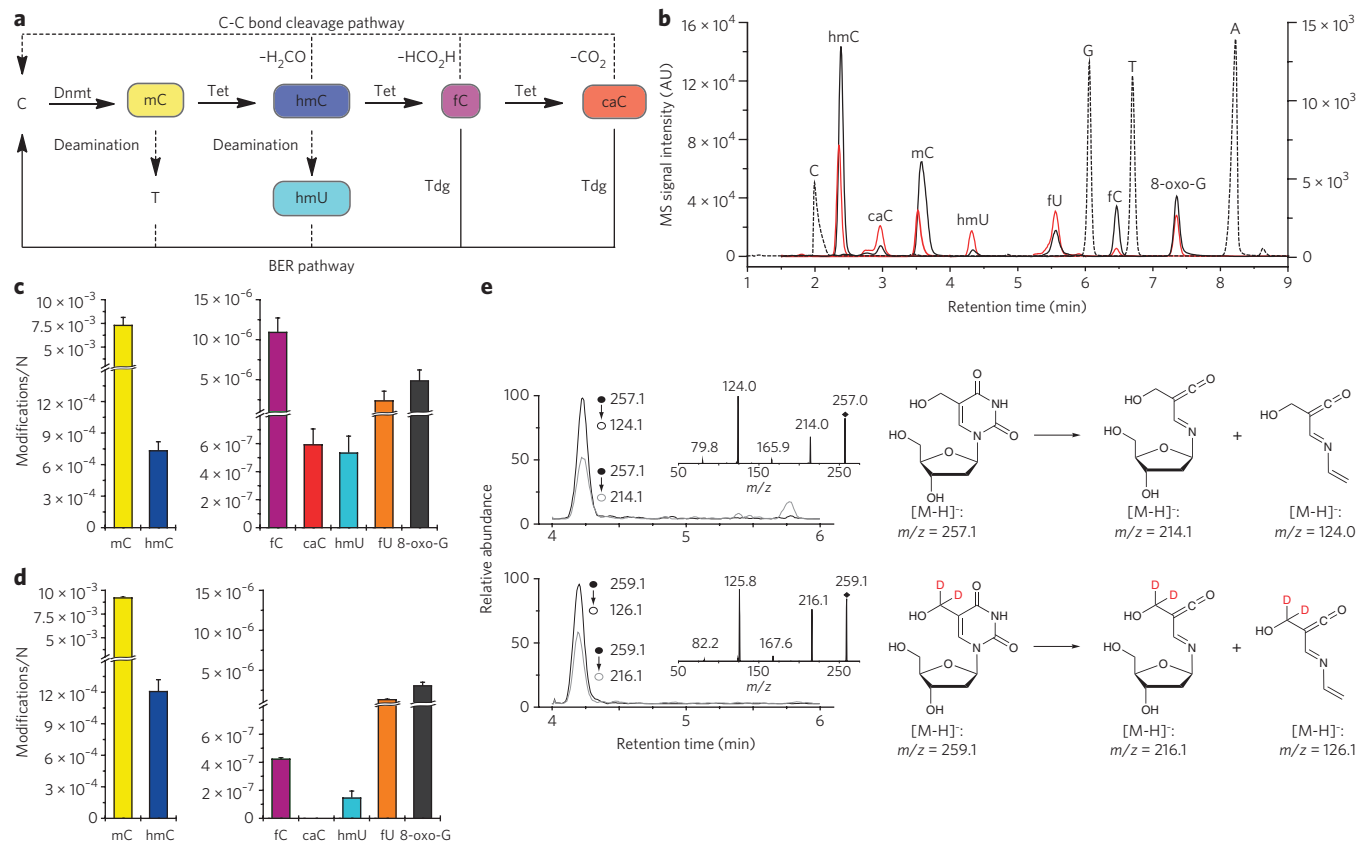
Figure 1 | Metabolism of cytosine derivatives, their detection by LC-UV-ESI-MS/MS and levels in mESCs and mouse cortex. (a) Potential active demethylation pathways. (**b**) Overlaid LC/UV and LC/MS/MS chromatograms of a representative DNA sample from mESCs. The dotted LC/UV chromatogram of C, G, T and 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 ± 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 [D$_2$]hmU internal standard (bottom trace) derived from a mESC DNA sample. The insets show the MS/MS full-scan spectra of synthetic hmU and [D$_2$]hmU matching the proposed fragmentation pathway.

Supplementary Fig. 2), the hmC levels are very high (13% relative to mC), as previously reported[19–21]. In contrast, fC was detected only in very small amounts (0.03% relative to hmC), and caC was not observed at all[22]. 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 [*methyl*-^{13}CD$_3$]L-methionine ([^{13}CD$_3$]L-Met) 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 [^{13}CD$_3$]L-Met for 5 d (2 passages) in the medium at a concentration of 0.2 mM furnished 89% labeled [^{13}CD$_3$]mC, 88% labeled [*hydroxymethyl*-^{13}CD$_2$]hmC and 93% labeled [*formyl*-^{13}CD]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}C and D into hmU and fU was observed, showing that hmC is not the precursor of hmU. We next added isotope-labeled [^{13}C,^{15}N$_2$]thymidine ([^{13}C,^{15}N$_2$]T) to the growth medium and observed ~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 (T → hmU; **Fig. 3c**). Consequently, all of the detected hmU

575

a

b
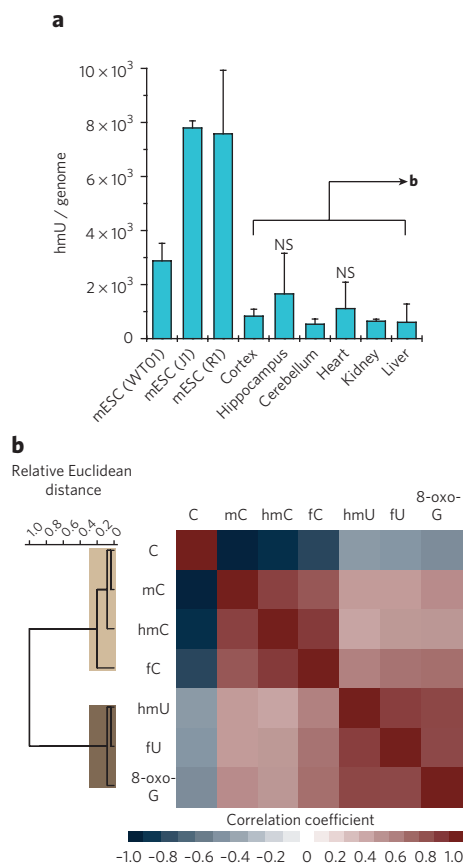
Figure 2 | hmU is present at elevated levels in mESCs compared to tissue.
(**a**) hmU levels per genome in mESCs (n_{WT01} = 7 replicates, n_{J1} = 2, n_{R1} = 3) and mouse tissue (3-month-old individuals, n = 3). Levels per genome were obtained considering a mouse genome size of 2.7×10^9 base pairs. Depicted are mean values ± s.d. The differences between mESCs and mouse tissues are significant ($P = 2.9 \times 10^{-5}$ to 3.6×10^{-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**. NS, not significant. (**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.

(and also fU) resides in an A base pair context (hmU:A). In this base pair, hmU is repaired by Smug1 but not Tdg[12]. 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[12]. To assess this possibility, we performed isotope tracing experiments using [$^{13}CD_3$]ʟ-Met in $Tdg^{-/-}$ 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-$^{13}CD_2$] hmU was not detected. However, in the cell line complemented with inactive Tdg, some labeled [hydroxymethyl-$^{13}CD_2$]hmU was indeed observed, albeit only in small amounts (~7% of total hmU and ~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[9,10,17,18], 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[11,12], we again replaced natural ʟ-Met with [$^{13}CD_3$]ʟ-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 premethylated 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[23–25]. 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[26,27]. 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[23–25]. 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 [$^{13}CD_3$]ʟ-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
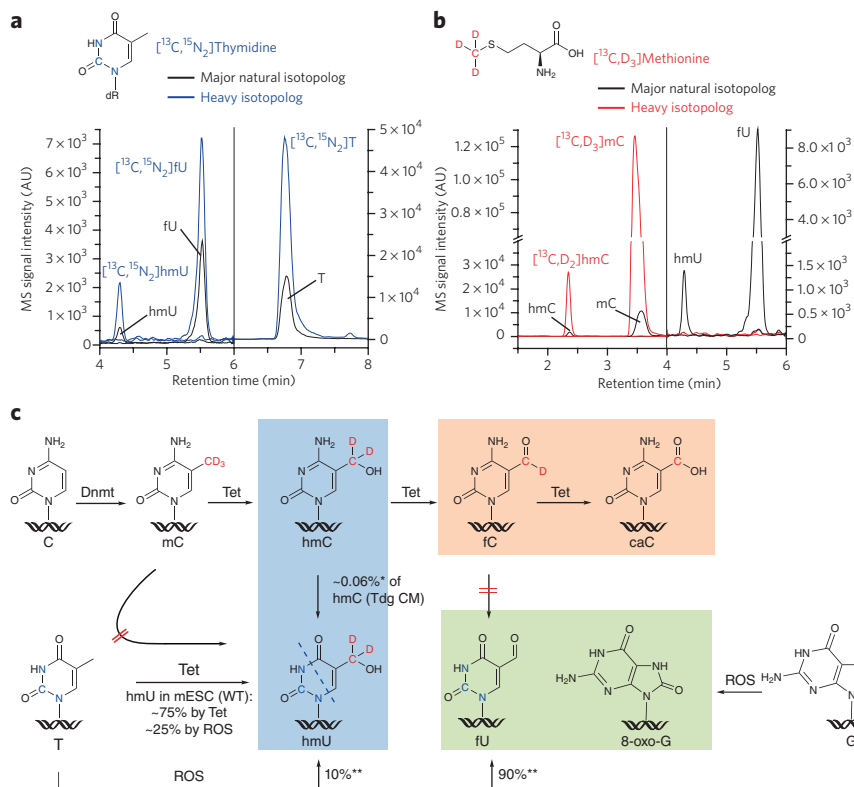
Figure 3 | hmU and fU are thymine oxidation products in WT mESCs with no detectable contribution from hmC or fC deamination. (**a**) Overlaid LC/MS/MS chromatograms of heavy ([$^{13}C,^{15}N_2$]hmU, [$^{13}C,^{15}N_2$]fU and [$^{13}C,^{15}N_2$]T; blue) and major natural (black) isotopologs of hmU, fU and T. AU, arbitrary units. (**b**) Overlaid chromatograms of heavy ([$^{13}C,D_3$]mC, [$^{13}C,D_2$]hmC, [$^{13}C,D_2$]hmU and [$^{13}C,D$]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[26]. 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 pulldown 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[28,29]. For this hmU study, we further developed this approach for the detection of proteins that 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[30,31]. 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**)[32]. 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 ε-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
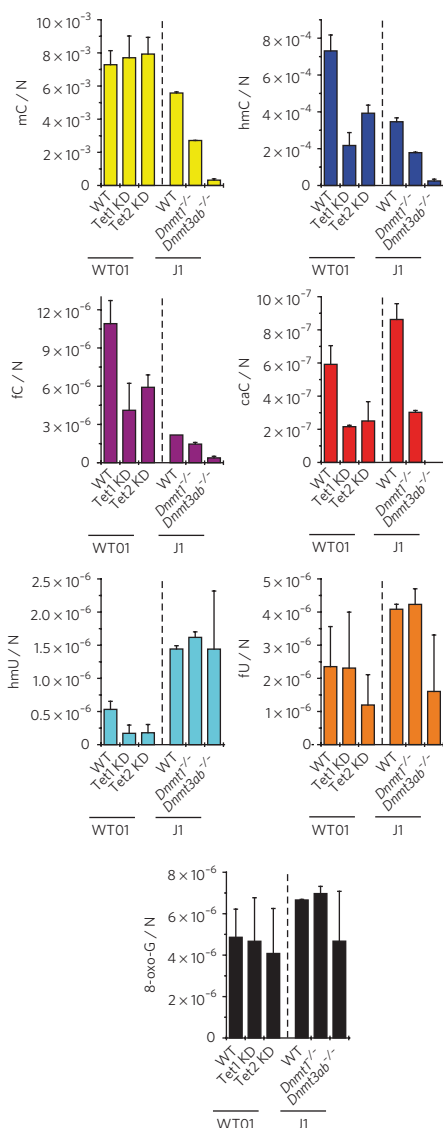
Figure 4 | Tet1 and Tet2 generate hmU in mESCs. Effect of Tet and Dnmt depletion on DNA modification levels in mESCs. J1 cell lines bearing homozygous *Dnmt*-null mutations and WT01 mESCs expressing Tet1 or Tet2 shRNAs (knockdown (KD)) were analyzed. Depicted are mean values per nucleoside (N) ± s.d. 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⁻/⁻* (J1, *n* = 2) and *Dnmt3a/b⁻/⁻* (J1, *n* = 6).

identification and quantification by MS[32,33]. 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 14
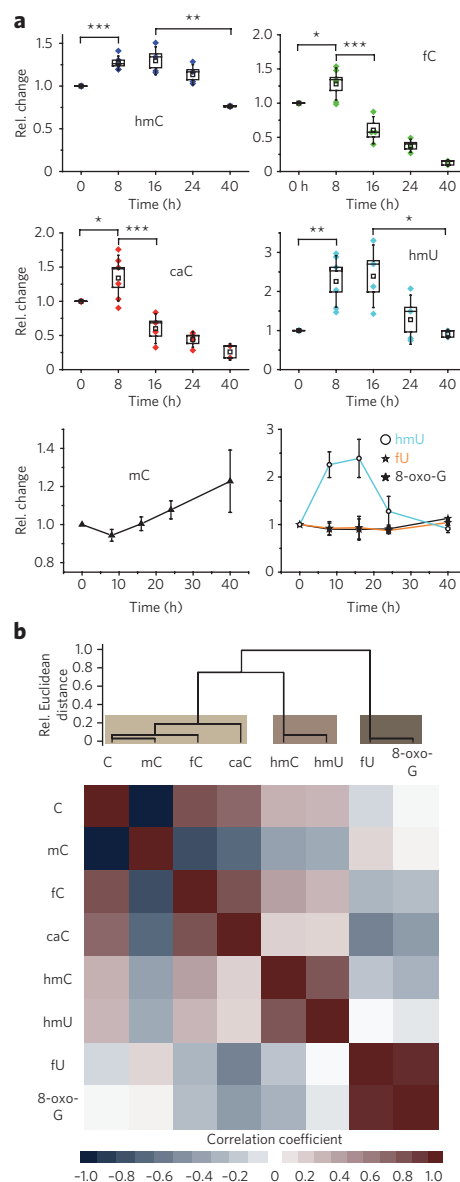
Figure 5 | hmU is produced during mESC differentiation. (**a**) Time course of DNA modification levels during early differentiation (0–40 h) of naive mESCs in the absence of growth factors. Box plot of relative modification changes for hmC, fC, caC and hmU averaged from six independent experiments using 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 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 ± 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.

(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 PHD finger
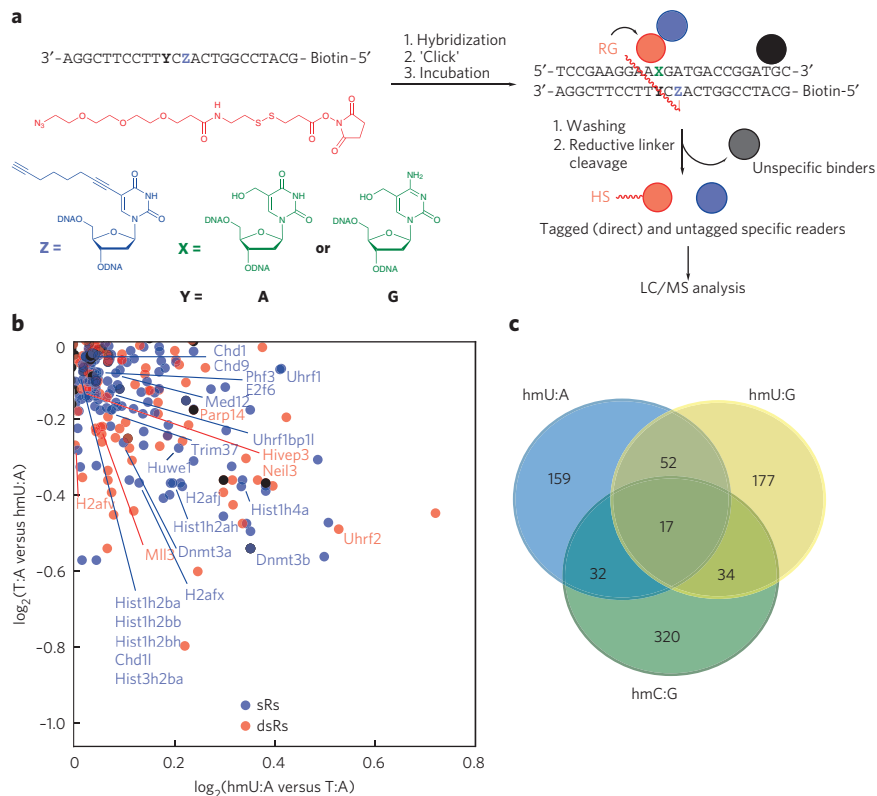
Figure 6 | Identification of hmU:A readers. (**a**) Workflow of pulldown experiments with hmU:A–containing oligonucleotides. DNA oligomers with the modified base Z containing an alkyne group for click reaction and biotin for the pulldown are hybridized with DNA oligomers containing hmU (X = hmU and Y = 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.

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[34], whereas we recently identified the homolog Uhrf2 as a specific binder of hmC[29]. 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[29] 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[35] 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[29], which further supports the validity of the covalent trapping method. Finally, we compared the results of the hmU:A pulldown with a pulldown 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 pulldown. 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[9,10], hmU in particular was recently discussed as an intermediate in active demethylation by deamination of hmC[11,12]. 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 [^{13}CD$_3$]L-Met and [^{13}C,^{15}N$_2$] 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[11,22]. To investigate the involvement of deamination[11–14], we studied $Tdg^{-/-}$ stem cells reconstituted with a catalytically inactive Tdg. In these cells, the putative deamination–repair complex[12] of Tdg and Aid could form, and we expected to see elevated hmU levels. 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[8,12], whereas hmU is efficiently recognized and repaired[36], 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[39]. Alternatively, hmU might influence transcription factor binding, which is known to be a hallmark of epigenetic reprogramming[40]. 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[41]. 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.

METHODS

Methods and any associated references are available in the online version of the paper.

References

1. Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
2. 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).
3. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
4. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
5. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
6. He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
7. Tan, L. & Shi, Y.G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895–1902 (2012).
8. 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).
9. 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).
10. Mouret, J.F., Polverelli, M., Sarrazini, F. & Cadet, J. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chem. Biol. Interact.* **77**, 187–201 (1991).
11. 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).
12. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
13. Nabel, C.S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
14. 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).
15. Jacobs, A.L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
16. 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).
17. Burrows, C.J. Surviving an oxygen atmosphere: DNA damage and repair. *ACS Symp. Ser. Am. Chem. Soc.* **2009**, 147–156 (2009).
18. 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).
19. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
20. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
21. Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
22. 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).
23. 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).
24. 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).
25. Leitch, H.G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
26. 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).
27. 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).
28. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
29. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
30. Burley, G.A. *et al.* Directed DNA metallization. *J. Am. Chem. Soc.* **128**, 1398–1399 (2006).
31. 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.* **41**, 2596–2599 (2002).
32. 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).
33. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
34. Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306 (2011).
35. 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).
36. 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).
37. 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).

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.K.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.

4 Unveröffentlichte Arbeiten

4.1 Substratselektivität der Tet-Enzyme zwischen 5mdC und dT

4.1.1 Einleitung

In murinen embryonalen Stammzellen (mES-Zellen) wird die genomische DNA durch Tet1 und Tet2 postreplikativ modifiziert.[5-8, 178, 319] Wie in Abschnitt 3.6 gezeigt wurde, oxidieren Tet1 und Tet2 dabei neben dem primären Substrat 5mdC[5-8] auch zu einem geringeren Anteil dT.[13] Ferner akkumulieren die Oxidationsprodukte 5hmdC, 5fdC, 5cadC und 5hmdU zu Beginn der frühen mES-Zelldifferenzierung intermediär.[13] Unklar ist bisher, ob Tet-Enzyme möglicherweise auch höhere Oxidationsprodukte von 5hmdU bilden. Darüber hinaus wurde die Selektivität von Tet-Enzymen gegenüber den Substraten 5mdC und dT sowie eine mögliche Modulation dieser Eigenschaften noch nicht quantitativ untersucht. Dies soll im Nachfolgenden anhand von Modellstudien im biologischen Kontext erörtert werden.

4.1.2 Ergebnisse und Diskussion

In der Vergangenheit nutzte man oftmals HEK-293T-Zellen, um die Aktivität der Tet-Enzyme zu untersuchen. Hierbei wurde gewöhnlich die gut exprimierbare C-terminale katalytische Domäne (englisch: *catalytic domain* = CD) des jeweiligen Tet-Enzyms mit einem N-terminalen GFP- oder Flag-Tag in die Zellen transfiziert.[5, 7-8, 13, 138, 177, 234, 313] Mit diesem System ließ sich die Bildung der drei Oxidationsprodukte von 5mdC nachweisen. **Abbildung 8A** stellt die quantitativen Ergebnisse dieses Experiments mit der katalytischen Domäne von Tet1 (GFP-mTet1CD)[I] dar, die mit Hilfe der in dieser Dissertation entwickelten UHPLC-UV-ESI-MS/MS-Methode gewonnen wurden (siehe Abschnitt 3.6). Die Expression des rekombinanten Enzyms verursachte eine starke Abnahme von 5mdC (-37%) in genomischer DNA und eine starke Zunahme der Oxidationsprodukte 5hmdC, 5fdC und 5cadC. Die Analysenmethode ermöglichte zudem auch das Endprodukt der DNA-Demethylierung per UHPLC-UV zu quantifizieren. So konnte eine globale Zunahme von dC (grau, **Abbildung 8A**) festgestellt werden. Da die Summe der quantifizierten Produkte (Δ5hmdC, Δ5fdC, Δ5cadC, ΔdC) der Abnahme von 5mdC im Betrag entsprach, konnte hiermit die Tet1CD-induzierte DNA-Demethylierung vollständig und exakt verfolgt werden.

[I] Bei dem GFP-mTet1CD-Konstrukt handelt es sich um eine N-terminale Fusion von GFP an die katalytische Domäne von Tet1 der Maus. Im nachfolgenden Text wird die Expression solcher Konstrukte nur noch mit Tet1 bezeichnet.

Die Demodifizierung von 5hmdC, 5fdC und 5cadC, die zu unmodifiziertem dC führt, kann dabei auf aktivem und passivem Wege geschehen (**Abbildung 8B**).[84]
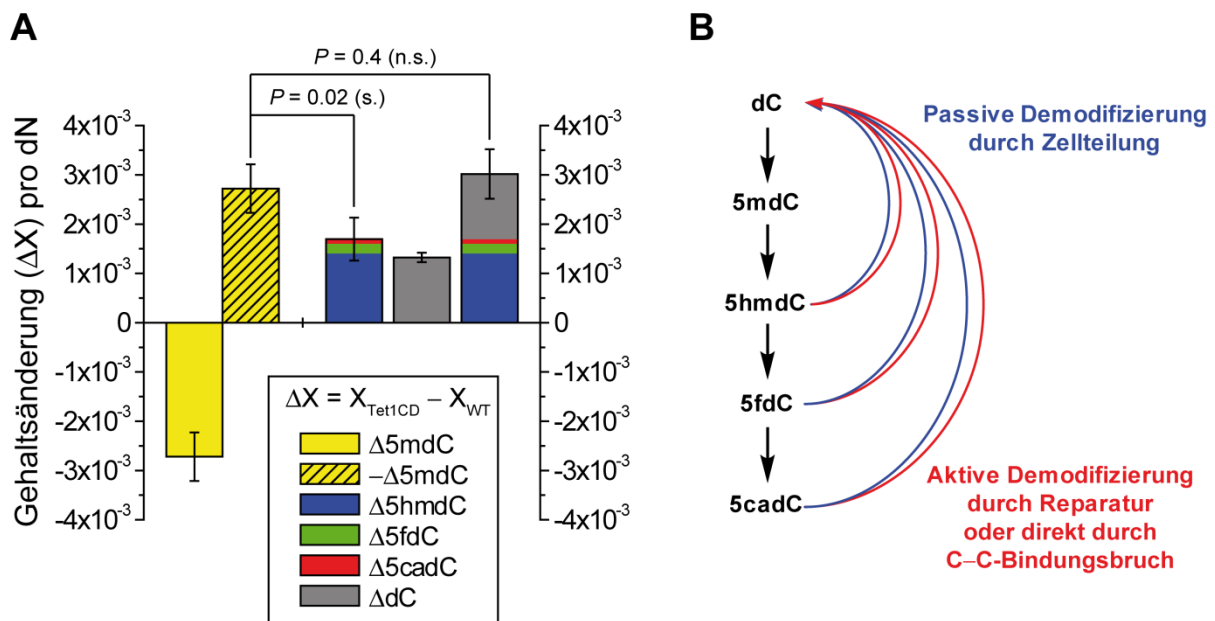
Abbildung 8 | DNA-Demethylierung in HEK-293T-Zellen durch ektopische Expression der katalytischen Domäne (englisch: *catalytic domain*; CD) von rekombinantem Tet1 (GFP-mTet1CD). (**A**) Dargestellt sind die absoluten Gehaltsänderungen (ΔX) pro dN von X = 5mdC, 5hmdC, 5fdC, 5cadC und dC, welche per quantitativer UHPLC-UV-ESI-MS/MS bestimmt wurden. ΔX ist die Differenz zwischen der Modifikationsmenge (X) aus der Tet1CD-Überexpression und der Modifikationsmenge (X) der Kontrollprobe (Wildtyp; WT). Die Balken repräsentieren biologische Mittelwerte und Standardabweichungen von vier unabhängigen Experimenten. Signifikanzwerte (*P*) wurden mit Hilfe eines ungepaarten zweiseitigen *t*-Tests berechnet. Signifikant (s.) sind Mittelwertsunterschiede mit $P < 0.05$; n.s. = nicht signifikant. (**B**) Darstellung des Reaktionsprofils der oxidativen DNA-Demethylierung.[57, 179, 187, 320]

Neben der Oxidation von 5mdC war in mTet1CD-transfizierten HEK-293T-Zellen auch eine Tet-vermittelte Oxidation von dT zu verzeichnen (**Schema 11**). Wie in Abschnitt 3.6 bereits beschrieben wurde, bildete sich dabei 5hmdU.[13] Die Expression von mTet1CD bewirkte eine absolute Zunahme von 5hmdU in Höhe von 2.9×10^{-5} pro dN (**Abbildung 9**). Die Zunahme von 5hmdU war relativ zur 5hmdC-Zunahme mit $2.1 \pm 0.3\%$ vergleichsweise gering. Darüber hinaus wurde auch eine Bildung von 5fdU beobachtet, die mit einer Zunahme von 3.2×10^{-6} pro dN zwar am Rande der Signifikanz ($P = 0.14$), aber dennoch über dem oxidativen Hintergrund (8oxodG)[II] lag. In entsprechender Weise lag die Zunahme von 5fdU bei $1.6 \pm 1.0\%$ relativ zur 5fdC-Zunahme. 5cadU als mögliches Oxidationsendprodukt[109] konnte

---

[II] 8oxodG diente, wie in Abschnitt 3.6 beschrieben, als Oxidationsmarker, dessen Gehaltsänderung in Korrelationen mit 5hmdU und 5fdU Aufschluss gibt, ob diese durch ROS-induzierte d.h. unspezifische Prozesse entstanden sind. In der Regel ereignen sich solche Oxidationsschäden mit einer Häufigkeit in der absteigenden Reihenfolge: 8oxodG > 5fdU > 5hmdU.[258]

mangels Detektionsempfindlichkeit nicht detektiert werden. In Analogie zur relativen Bildungsrate von 5hmdU und 5fdU könnte 5cadU ebenfalls mit 2% relativ zu 5cadC gebildet werden. Dies entspräche einer absoluten Zunahme von 1.5×10^{-6} pro dN.
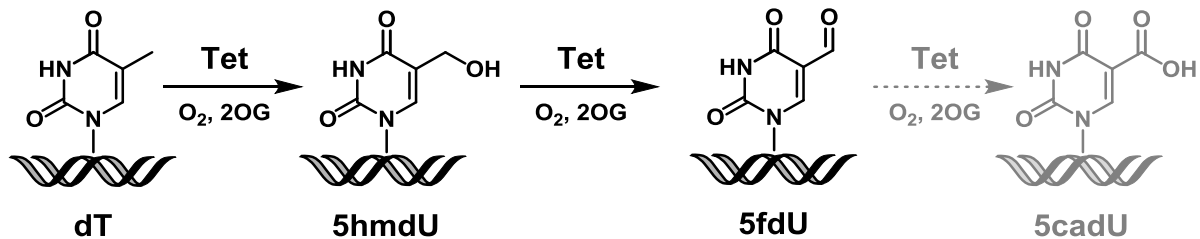
Schema 11 | Sukzessive Oxidation von dT zu 5hmdU und 5fdU durch Tet1CD in genomischer DNA. 5cadU konnte jedoch nicht detektiert werden. 2OG = 2-Oxoglutarat.
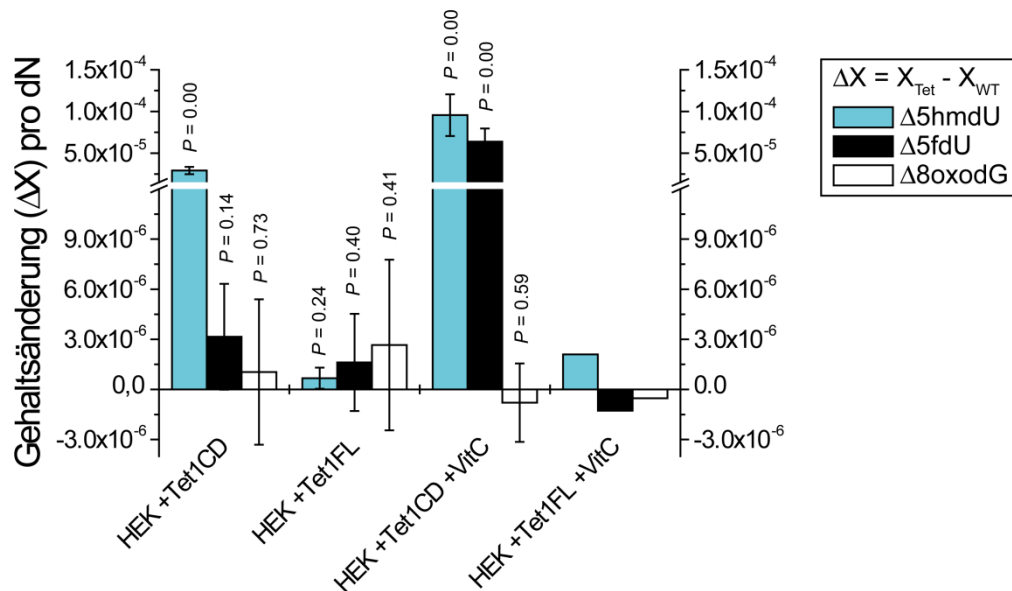
Abbildung 9 | Rekombinante Tet-Enzyme oxidieren dT zu 5hmdU und 5fdU in genomischer DNA von HEK-293T-Zellen. Absolute Gehaltsänderungen (ΔX) pro dN von X = 5hmdU, 5fdU und 8oxodG. ΔX ist die Differenz zwischen der Modifikationsmenge (X_{Tet}) aus der jeweiligen GFP-mTet1-Überexpression (±Vitamin C) und der Modifikationsmenge (X_{WT}) der Kontrollprobe (Wildtyp; WT). Die Balken repräsentieren biologische Mittelwerte und Standardabweichungen von $n = 4$, 2, 3 bzw. 1 unabhängigen Experimenten in der dargestellten Reihenfolge. Signifikanzwerte (P) wurden mit Hilfe eines ungepaarten zweiseitigen t-Tests berechnet. Signifikant sind Mittelwertsunterschiede mit $P < 0.05$.

Da sich kürzere Enzymkonstrukte in ihrem Reaktionsprofil von den nativen Enzymen unterscheiden können, wurden in HEK-293T-Zellen als nächstes volllängen Tet1 (GFP-mTet1FL; englisch: *full length* = FL) untersucht. Die Zunahme von 5hmdU war nur noch 0.2±1.0% relativ zur 5hmdC-Zunahme und eine Zunahme von 5fdU konnte nicht mehr festgestellt werden. Tet1FL ist damit deutlich selektiver als Tet1CD. Als nächstes stellte sich die Frage, ob sich die Selektivität möglicherweise ändert, wenn die Aktivität der Tet-Enzyme

gesteigert wird, wie z.B. durch Vitamin C.[78, 321-323] *Yin et al.* konnten mit höheren Vitamin C-Konzentrationen eine gesteigerte Oxidationsrate von 5mdC durch Tet1 und Tet2 beobachten.[323] Die Aktivitätssteigerung wurde durch eine allosterische Bindung von Vitamin C begründet, die vermutlich die Faltungsdynamik des Enzyms beeinflusst.[323] Wie in **Abbildung 9** zu sehen ist, verursacht die Zugabe von Vitamin C zu Tet1CD-transfizierten HEK-Zellen auch eine erhöhte Oxidationsrate von dT. Ferner wurde dabei ein deutlicher Verlust der Selektivität zwischen der 5mdC- und dT-Oxidation festgestellt. Die relative Zunahme von 5hmdU und besonders von 5fdU war stärker ausgeprägt, als die relative Zunahme von 5hmdC, 5fdC oder 5cadC (**Abbildung 10A**). Bei Tet1FL lag dagegen der umgekehrte Fall vor. Die relative Zunahme von 5hmdU blieb trotz Vitamin C-Behandlung gering und eine Bildung von 5fdU konnte noch immer nicht beobachtet werden (**Abbildung 9**). Mit dieser Beobachtung im Einklang war auch bei Vitamin C-behandelten mES-Zellen, die Tet1 und Tet2 exprimieren, die Zunahme von 5hmdC, 5fdC und 5cadC ausgeprägter, als die Zunahme von 5hmdU (**Abbildung 10B**).
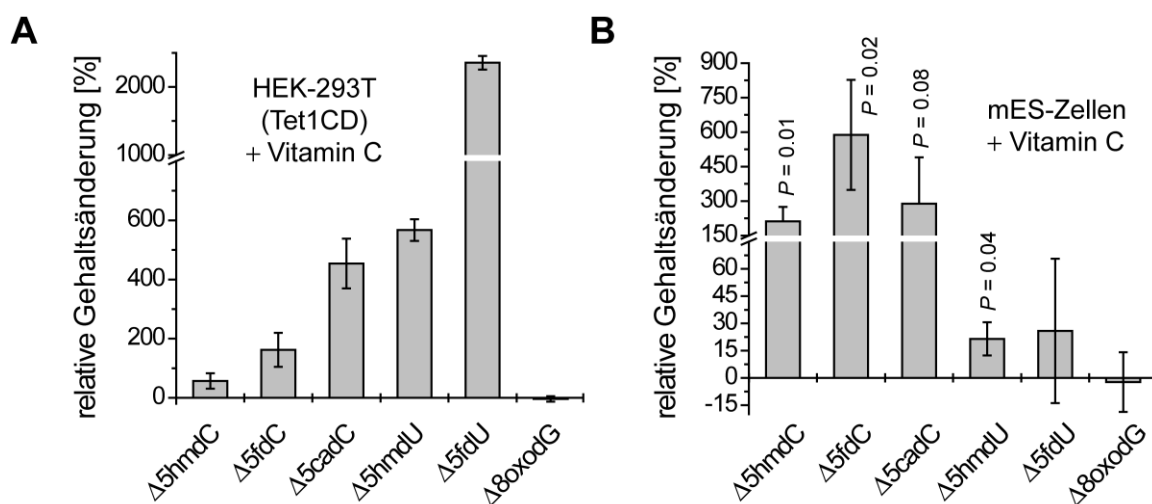
Abbildung 10 | Modulation der Aktivität und Selektivität von Tet-Proteinen in HEK-293T- und mES-Zellen durch Vitamin C-Behandlung. (**A**) Relative Gehaltsänderung der DNA-Modifikationslevel in HEK293T-Zellen mit ektopischer Expression von GFP-mTet1CD mit Vitamin C-Behandlung im Vergleich zu unbehandelten GFP-mTet1CD-exprimierenden Zellen. Die Daten wurden auf das GFP-Signal als Expressionsmaß beider Vergleichsproben normiert. Dargestellt sind die Mittelwerte und Standard-abweichungen von 3 technischen Replikaten. (**B**) Relative Gehaltsänderung in mES-Zellen nach Vitamin C-Behandlung. Dargestellt sind die Mittelwerte und Standardabweichungen von 3 biologischen Replikaten. Signifikanzniveaus: *$P < 0.05$; **$P < 0.01$ (gepaarter zweiseitiger *t*-Tests).

Zusammenfassend lässt sich somit festhalten, dass die N-terminalen, regulativen Domänen von Tet1FL und Tet2FL dem aktiven Zentrum des Enzyms zusätzliche Selektivität verleihen. Rekombinante Versionen der katalytischen Domäne der Enzyme können dT in

vergleichsweise höheren Raten oxidieren. Neben 5hmdU wird dabei auch 5fdU in signifikanten Mengen gebildet.

Untersuchung der Selektivitätsänderungen mit Hilfe eines quantitativen Parameters

Wie hoch die 5mdC/dT-Selektivität genau ist, und ob sie sich *in vivo* z.B. in differenzierenden mES-Zellen ändern kann, wird im Nachfolgenden mit Hilfe eines Selektivitätsparameters, dem sogenannten *E*-Wert, quantitativ untersucht. Der *E*-Wert[III] wurde ursprünglich für die quantitative Beschreibung kinetischer Racematspaltungen eingeführt[324-326] und entspricht formal dem Verhältnis der Geschwindigkeitskonstanten bzw. der Spezifitätskonstanten (k_{cat}/K_m) der Reaktionen zweier konkurrierender Substrate A und B eines Enzyms:[325, 327]

$$E \equiv \frac{v_{max}^A / K_m^A}{v_{max}^B / K_m^B} \equiv \frac{k_{cat}^A / K_m^A}{k_{cat}^B / K_m^B} \qquad (1)$$

Diese können entweder zwei Enantiomere sein oder, wie im Fall der Tet-Enzyme, die Substrate 5mdC und dT. Das Verhältnis der Spezifitätskonstanten ist der Parameter der Wahl, um die relativen Raten zweier konkurrierender enzymatischer Reaktionen auszudrücken und wird deswegen allgemein zur Berechnung von Enzymselektivitäten verwendet. *E* ist per Definition eine intrinsische Eigenschaft eines Enzyms.[325]

Diskussion zur Anwendbarkeit und Gültigkeit des *E*-Werts

Wichtige Voraussetzungen für die Berechnung von *E* sind unter anderem, dass eine *Michaelis-Menten*-Kinetik vorliegt, die Reaktion irreversibel ist, keine Produktinhibierung vorliegt und die Messwerte nicht durch andere Reaktionen, z.B. durch Überlagerung von Produktprofilen, verfälscht sind.[324-325, 328]

Bei Fe(II)- und 2-Oxoglutarat-abhängigen Dioxygenasen, zu denen die Tet-Enzyme zählen, verläuft die Oxidation von C−H-Bindungen nach einem komplexen ping-pong-Mechanismus (**Schema 12A**).[108-109, 111-112, 128, 132-135, 152] Die Gesamtreaktion kann *in vivo* allerdings aus zwei Gründen als pseudo uni-uni-Mechanismus nach *Michaelis-Menten* analysiert werden: Erstens, es ist anzunehmen, dass sich die Konzentration von Fe(II), 2-Oxoglutarat und O_2 in der Zelle nicht ändert; und zweitens, die Bildung des Fe(IV)-Oxo-Intermediats irreversibel und geschwindigkeitsbestimmend ist, welches die Oxidation der C−H-Bindung bewerkstelligt (**Schema 12B**).[133, 135, 147, 149] Die Reaktionsraten der Oxidation von 5mdC und dT werden

---

[III] In einer kinetischen Racematspaltung ist der Grad der Enantiomerenanreicherung vom Umsatz abhängig. Ein Vergleich hinsichtlich der Selektivität enzymatischer Racematspaltungen ist nur bei gleichem Umsatz möglich. Um dies zu umgehen, ist von *Chen* und *Sih* der dimensionslose Selektivitätsparameter *E* (*enantiomeric ratio; E-Wert*) eingeführt worden.[324]

schlussfolgernd nur durch die Affinität von Tet zum jeweiligen Substrat bestimmt, da die Substratbindung die irreversible Aktivierung von O_2 auslöst. Die sich an diesen irreversiblen Schritt anschließende Oxidation des gebundenen Substrats ist dann nicht mehr ratenbestimmend (**Schema 12**).[112, 132, 150-151]

A *Ping-pong-Mechanismus:*

Tet^{II}:S

+ S
− 2 H$_2$O

+ OG
− 2 H$_2$O

1) + RH, − H$_2$O
2) + O$_2$

irreversibel und ratenbestimmend!

Tet^{II} ⇌ Tet^{II}:OG ⇌ Tet^{III}:OG:O$_2$:RH ⟶ Tet^{IV}:S:CO$_2$:RH

+ F
− 2 H$_2$O

Tet^{II}:F

Tet^{II}:
His — Fe^{II} — OH$_2$
Asp / His \ OH$_2$
OH$_2$, OH$_2$

OG: HO—C(=O)—CH$_2$—CH$_2$—C(=O)—OH

RH: 5mdC, 5hmdC, 5fdC, dT, 5hmdU, 5fdU

− CO$_2$

H—O—Tet^{III}:S:R$^{\cdot}$

+ H$_2$O

Tet^{II}:S:ROH

− S, − ROH,
+ 2 H$_2$O

Tet^{II}
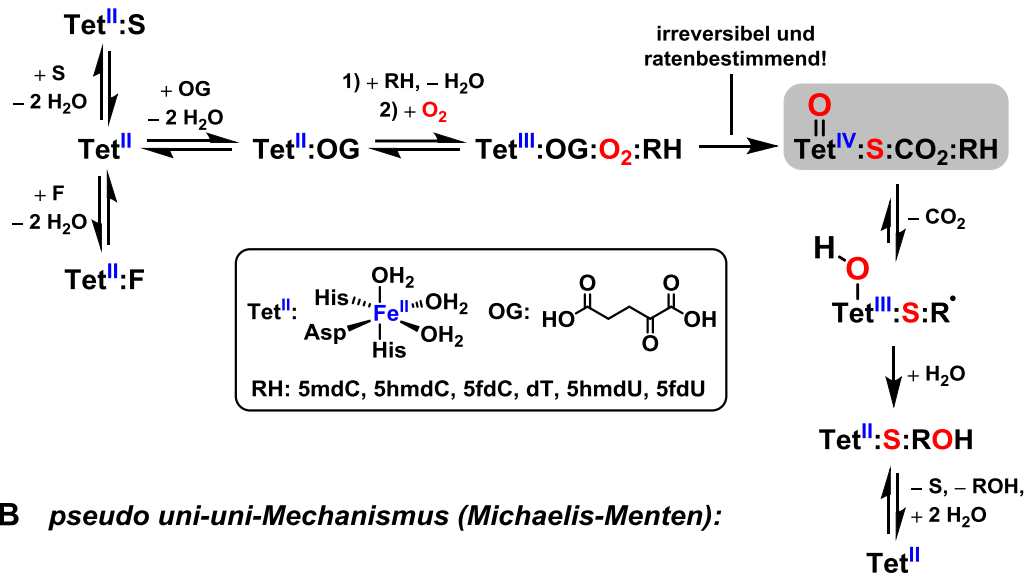
B *pseudo uni-uni-Mechanismus (Michaelis-Menten):*

1) + OG, − 2 H$_2$O
2) + RH, − H$_2$O
3) + O$_2$

irreversibel und ratenbestimmend!

Tet^{II} ⇌ Tet^{III}:OG:O$_2$:RH ⟶ Tet^{IV}:S:RH + CO$_2$

k_+ / k_-

k_{cat}
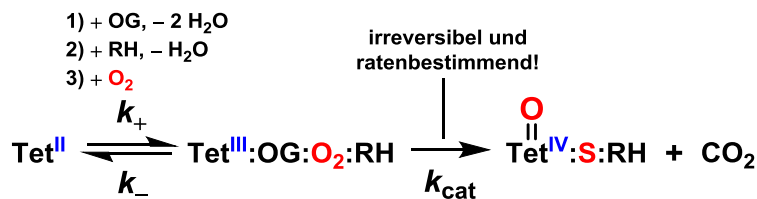
Schema 12 | Reaktionsmechanismus von Tet-Enzymen, abgeleitet anhand verwandter Fe(II)- und 2-Oxoglutatrat (OG)-abhängiger Dioxygenasen.[108-109, 111-112, 128, 132-135, 152] (**A**) Die stufenweise Oxidation von 5mdC oder dT verläuft nach einem Ping-pong-Mechanismus.[146] Durch die oxidative Decarboxylierung von OG wird das Enzym hierbei zunächst chemisch verändert. In diesem irreversiblen und ratenbestimmenden Schritt entsteht ein Fe(IV)-Oxo-Intermediat (grau markiert), welches anschließend die Oxidation der C−H-Bindung des RH-Substrats bewerkstelligt.[133, 135, 147, 149] (**B**) Kinetisch gesehen kann die Gesamtreaktion nach *Michaelis-Menten* als pseudo uni-uni-Mechanismus analysiert werden, da sich die Konzentration von Fe(II), OG und O$_2$ in der Zelle nicht ändert und die Bildung des Fe(IV)-Oxo-Intermediats irreversibel und geschwindigkeitsbestimmend ist. Für eine detaillierte Darstellung des Katalysezyklus siehe auch Abschnitt 1.4.1. Doppelpunkte symbolisieren intermolekulare Wechselwirkungen innerhalb eines Enzym-Substrat-Komplexes.

Desweiteren muss für die Gültigkeit des *E*-Werts gegeben sein, dass keine Substrat-, Intermediat- oder Produkt-Hemmung im betrachteten System vorliegt. Der Absolutwert des *E*-Wertes kann sich sonst während der fortschreitenden Reaktion ändern.[329] Dies geschieht allerdings nur dann, wenn nur eines oder wenige Substrate des Substratspektrums eine selektive Inhibierung bewirken.[329] Auszuschließen ist dies bei Tet, da die größte Affinität zu

5mdC besteht ($k_{5mdC} \gg k_{5hmdC} > k_{5fdC}$; siehe auch Abschnitt 4.2.2)[8] und darüber hinaus ein vollständiger Reaktionsumsatz von 5mdC zu 5cadC *in vitro* berichtet wurde.[174]

Ferner können Protein-Protein-Interaktionspartner,[13-14, 208] posttranslationale Modifikationen,[193] Autoinhibitionsmechanismen,[140] nicht-kompetitive Inhibitoren oder allosterisch-bindende Faktoren (Vitamin C, ATP)[7, 78, 321-323, 330] ebenfalls Einfluss auf den Selektivitäts-parameter haben. Diese Einflüsse sind aus epigenetischer Sicht von höchstem Interesse und deshalb Gegenstand dieser Untersuchung. Da diese und andere multiple Faktoren (z.B. Entfernung der Produkte durch Zellteilung und Reparaturprozesse) je nach Zelltyp sehr unterschiedlich sein können,[14] werden im Nachfolgenden Selektivitätsunterschiede nur innerhalb der gleichen Zelllinie (gleicher genetischer Hintergrund) und bei gleichen Kultivierungsbedingungen diskutiert.

Mathematische Herleitungen

Chen et al. zeigten bei der Einführung des *E*-Wertes, dass das Verhältnis der Geschwindig-keitskonstanten (k^A/k^B) zweier konkurrierender Reaktionen als Selektivitätsparameter, wie er beispielsweise zur Quantifizierung von enantioselektiven photochemischen und chemischen Katalysen zur Anwendung kam, auch in der Enzymkatalyse gültig ist.[324-325] In der Abwesenheit von Nebenreaktionen kam somit Gleichung 2 zur Anwendung:

$$\frac{dc_S^A / dt}{dc_S^B / dt} = E \times \frac{c_S^A}{c_S^B} \tag{2}$$

Nach Integration mit den Anfangskonzentrationen der Substrate A und B ($c_{S,0}^A$ und $c_{S,0}^B$) zum Zeitpunkt $t = 0$ h ergibt sich Gleichung 3:

$$\ln\frac{c_{S,t}^A}{c_{S,0}^A} = E \times \ln\frac{c_{S,t}^B}{c_{S,0}^B} \tag{3}$$

Gleichung 3 kann entsprechend umgeformt werden und man erhält einen Ausdruck, bei dem die Stoffmengen bzw. die relativen Stoffmengen (pro dN) der Substrate 5mdC und dT einsetzbar sind:

$$E = \frac{\ln\left(n_{S,t}^{5mdC}/n_{S,0}^{5mdC}\right)}{\ln\left(n_{S,t}^{dT}/n_{S,0}^{dT}\right)} \tag{4}$$

Gleichung 4 berücksichtigt somit das statistische Verhältnis in dem die Substrate 5mdC und dT vorliegen. Der 5mdC-Gehalt lag in den hier untersuchten Zellen bei durchschnittlich

5.8×10^{-3} pro dN, während dT mit 0.23 pro dN in großem Überschuss vorlag (Abbildung 11). Allerdings geht nur die relative Änderung der jeweiligen Substratmenge ($n_{S,t}/n_{S,0}$) logarithmisch in den *E*-Wert ein. Obwohl dT in der genomischen DNA von HEK-293T-Zellen einen Anteil von 96.9% an der Gesamtsubstratmenge (5mdC + dT) einnimmt, überwiegt der Anteil der 5mdC-Oxidationsprodukte (5hmdC, 5fdC, 5cadC und dC) an der Gesamtproduktmenge mit 97.6% in Tet1CD-transfizierten Zellen. Dies illustrieren die Kreissektoren nach Äquivalenzierung der Substrat- und Produktmengen in Abbildung 11 rechts unten. Zum Vergleich: In Vitamin C behandelten Tet1CD-transfizierten Zellen war der Produktüberschuss 89.1% und in Tet1FL-transfizierten Zellen 99.8%. Da, wie oben erwähnt, nur die relativen Änderungen der jeweiligen Substratmengen in den *E*-Wert (Gleichung 4) eingehen, ist eine Äquivalenzierung der Substratmengen zur Berechnung des *E*-Werts nicht nötig.[325]
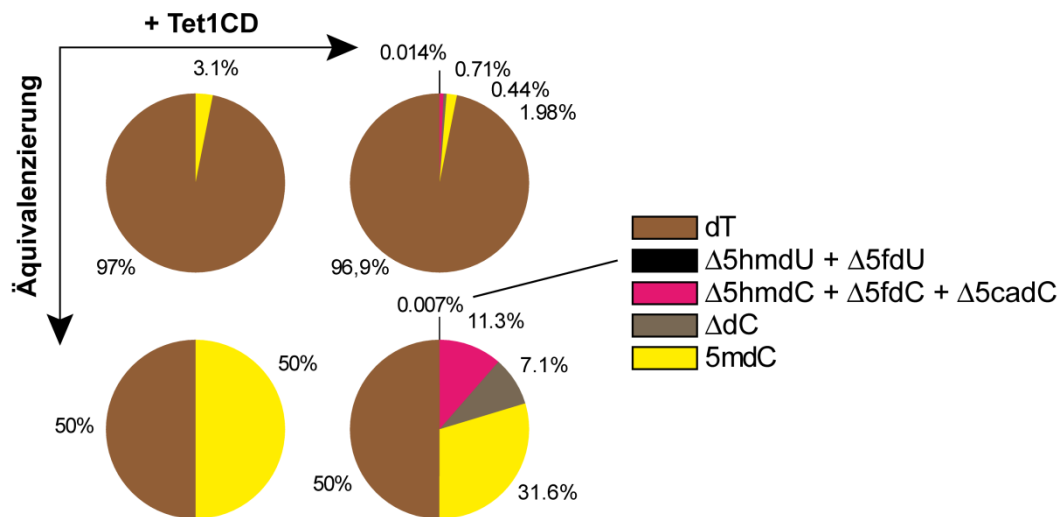
Abbildung 11 | Vergleich der 5mdC- und dT-Oxidation durch ektopische Expression von GFP-mTet1CD in HEK-293T-Zellen unter Berücksichtigung der Häufigkeit von dT bzw. 5mdC in der untersuchten DNA. Dargestellt sind die absoluten Gehaltsänderungen (∆X) pro dN von X = 5mdC, 5hmdC, 5fdC, 5cadC und dC. Die Kreissektoren repräsentieren biologische Mittelwerte von vier unabhängigen Experimenten.

Ausgehend von Gleichung 4 hängt die Genauigkeit der *E*-Wert-Bestimmung von der Genauigkeit der Quantifizierung der Stoffmengenänderung von 5mdC und dT ab. Da die prozentualen Änderungen der Stoffmengen von 5mdC groß sind, können mit der verwendeten Analysenmethode signifikante Unterschiede direkt bestimmt werden. Im Gegensatz dazu ist die prozentuale Änderung der Stoffmengen von dT sehr klein. Eine direkte Messung von dT schied daher aus. Deswegen wird die Gesamtstoffmenge von dT (n_{ges}^{dT}) per Definition gleich

0.23 pro dN gesetzt und die Stoffmengenänderung von dT aus der Differenz zur Stoffmenge von 5hmdU und 5fdU ermittelt:

$$E = \frac{\ln\left(n_{S,t}^{5mdC}/n_{S,0}^{5mdC}\right)}{\ln\left[\left(n_{ges}^{dT} - n_{P,t}^{5hmdU} - n_{P,t}^{5fdU}\right)/\left(n_{ges}^{dT} - n_{P,0}^{5hmdU} - n_{P,0}^{5fdU}\right)\right]} \tag{5}$$

Gleichung 5 ist gültig, sofern Nebenreaktionen ausgeschlossen werden können, die ebenfalls zu einer Verminderung des 5mdC- oder dT-Gehaltes führen. Da deren Oxidationsprodukte, insbesondere 5hmdU und 5fdU, durch reaktive Sauerstoffspezies entstehen können (siehe Abschnitt 3.4 und 3.6),[12-13, 164, 257, 260] wurden nur solche Proben zur Auswertung herangezogen, bei denen die Level des Oxidationsmarkers 8oxodG nahezu unverändert blieben ($\Delta n_{rel}^{8oxodG} \cong 0\%$ oder $\Delta n_{rel}^{8oxodG} < \Delta n_{rel}^{5fdU} < \Delta n_{rel}^{5hmdU}$)[IV]. Falls die relative Änderung von 5fdU kleiner als die Änderung von 8oxodG oder gleich war ($\Delta n_{rel}^{5fdU} \leq \Delta n_{rel}^{8oxodG} < \Delta n_{rel}^{5hmdU}$), wurde die relative Änderung von 5fdU durch Tet per Definition gleich null gesetzt ($\Delta n_{P}^{5fdU}(Tet) \stackrel{def}{=} 0$). 5hmdU und 5fdU könnten ferner auch durch Desaminierung von 5hmdC und 5fdC entstehen (siehe Abschnitt 3.4).[12] Dieser Fall ist allerdings unter normalen Bedingungen in genomischer DNA von HEK-293T- und mES-Zellen nicht messbar (siehe Abschnitt 3.6).[13]

Ein Problem von Gleichung 5 ist die Tatsache, dass die Oxidationsprodukte 5hmdU und 5fdU durch Reparatur und Zellteilung entfernt, d.h. durch dT wieder ersetzt werden. Daher geht eine gewisse Fraktion der dT-Oxidation in die Berechnung des Selektivitätsparameter nicht ein. Bei der Evaluierung der Gesamtproduktmenge der Oxidationsreaktionen von 5mdC ist dies kein Problem, da die Oxidationsprodukte bei Reparatur oder Zellteilung durch dC ersetzt werden (und nicht durch 5mdC).

Die Ergebnisse der Selektivitätsparameter mittels Gleichung 5 sind in **Abbildung 12** zusammengefasst. Tet1CD oxidiert in HEK-Zellen dT und 5mdC mit einer relativen Rate von 1:3400. Das Volllängen-Enzym (Tet1FL) weist, wie bereits erwähnt, eine beträchtlich höhere Selektivität auf und oxidiert dT und 5mdC mit einer Rate von 1:64000. Auf diesem Niveau befindet sich auch die Selektivität von Tet3FL. Die aktivitätssteigernde Wirkung von Vitamin C veränderte die Selektivität von Tet1FL nicht signifikant, wohingegen die Selektivität von Tet1CD beträchtlich abnahm (1:700). Die N-terminalen Domänen tragen somit bis zu zwei Größenordnungen zur Selektivität der katalytischen Domäne bei. Diese Ergebnisse

---

[IV] $\Delta n_{rel} = (n_t - n_0)/n_0$

untermauern nun quantitativ die oben getroffenen qualitativen Schlussfolgerungen (siehe Diskussion zu **Abbildung 9** und **Abbildung 10**).
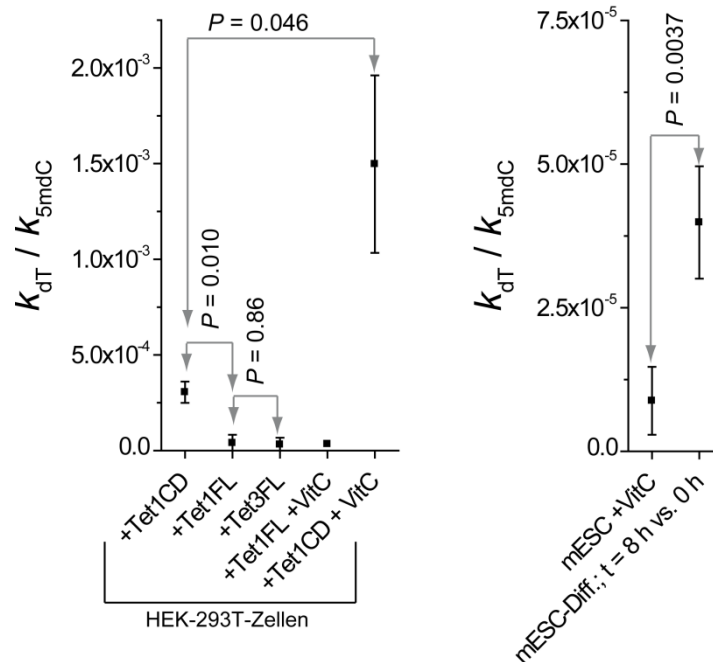
Abbildung 12 | Selektivitätsparameter von Tet-Enzymen in verschiedenen biologischen Kontexten: E-Werte wurden reziprok abgebildet ($E^{-1} = k_{dT} / k_{5mdC}$). Dargestellt sind Mittelwerte biologischer Replikate ± Standardabweichung (SD). Die Angabe von SD fehlt, wenn nur ein Replikat untersucht wurde. Signifikanzwerte (P) wurden mit Hilfe eines ungepaarten zweiseitigen t-Tests berechnet.

In Vitamin C-behandelten mES-Zellen war die relative Rate zwischen der dT und 5mdC-Oxidation 1:170000. Erstaunlicherweise nahm diese Selektivität in den ersten 8 h der mES-Zelldifferenzierung signifikant um eine Größenordnung ab (1:26000; **Abbildung 12**). Festzuhalten ist daher, dass einerseits eine hohe Selektivität für 5mdC im Vergleich zu dT vorliegt, diese aber *in vivo* bei Differenzierungsprozessen moduliert wird. Darüber hinaus ist die absolute Selektivität wohl geringer. Erstens kann mit diesem Ansatz der Verlust der Oxidationsprodukte von dT durch Reparatur und Zellteilung nicht evaluiert werden. Ferner kennt man Substratpräferenzen von Tet für 5mCpG- gegenüber 5mCpH-Dinukleotiden (H = A, C, T),[128-129] die vermutlich auch für TpG- bzw. TpH-Dinukleotide gelten könnten und ferner spielen Heterochromatinbereiche, Histonmodifikationen und damit die genomische Lokalisation von Tet-Enzymen eine Rolle.[140, 176, 187, 191, 197, 204] Deshalb kann nur eine begrenzte Fraktion von dT im Genom zur Oxidation durch Tet zur Verfügung stehen. Der reale Selektivitätswert (E-Wert) ist daher von diesen lokalen Substratkonzentrationen abhängig.[325] In dem hypothetischen Fall, dass die zur Verfügung stehenden Substratmengen

von dT und 5mdC gleich sind, könnte die relative Rate zwischen der Oxidation von dT und 5mdC in differenzierenden mES-Zellen 1:360 nach Gleichung 5 betragen.

4.1.3 Zusammenfassung

Tet-Enzyme oxidieren 5mdC in genomischer DNA mit einer hohen Selektivität zu 5hmdC, 5fdC und 5cadC.[5-8] dT wird in analoger Weise in einem geringeren Umfang zu 5hmdU,[13] 5fdU und vermutlich auch zu 5cadU oxidiert. Während der Substratüberschuss von dT in genomischer DNA von HEK-293T-Zellen 96.9% ist, war der Produktüberschuss der 5mdC-Oxidationsprodukte 99.8% in Volllängen-mTet1-transfizierten Zellen. Die relative Rate zwischen der dT- und 5mdC-Oxidation liegt nach den vorliegenden Schätzungen für das rekombinante Volllängen-mTet1 bei 1:64000. Diese Schätzung beruht allerdings im Wesentlichen auf der Annahme, dass alle dTs und 5mdCs des Genoms gleichermaßen zur Reaktion zur Verfügung stehen. Da durch Heterochromatinbereiche, die genomische Lokalisation von Tet und etwaige sequenzabhängige Substratpräferenzen die zur Verfügung stehende dT-Menge wesentlich geringer sein muss, könnte die tatsächliche Selektivität von Tet1 um bis zu zwei Größenordnungen geringer sein. Ferner konnte gezeigt werden, dass die primär regulativen N-terminalen Domänen bis zu zwei Größenordnungen zur Gesamtselektivität des Enzyms beitragen. Während der frühen Differenzierung von mES-Zellen[13] konnte des Weiteren gezeigt werden, dass die Selektivität in den Zellen intermediär um eine Größenordnung herabgesetzt wurde. Dies verdeutlicht die Modulierbarkeit der Selektivität im biologischen System.

In Zukunft bleibt zu klären, ob lediglich keine bessere Selektivität für 5mdC erreicht werden konnte und die Oxidation von dT als Nebenreaktion von der Zelle in Kauf genommen wird oder ob zusätzliche biologische Funktionen in den betreffenden Organismen etabliert wurden (siehe Abschnitt 1.5 und 3.6).[13] Wertvolle Informationen könnten hierzu über die Sequenzierung von 5hmdU im Genom erhalten werden.

4.1.4 Projektbeiträge in diesem Abschnitt

Ich führte die LC-MS/MS-Quantifizierung der DNA-Modifikationen von mES-Zellen mit Hilfe der von mir in Abschnitt 3.6 entwickelten Methode durch. Ferner führte ich die Bestimmung des Selektivitätsparameters durch. *Fabio Spada* führte in Zusammenarbeit mit *Nada Raddaoui* die Kultivierung der mES-Zellen durch. *Mirko Wagner* und *Caterina Brandmayr* führten die Kultivierung der HEK-293T-Zellen sowie die Bestimmung der DNA-

Modifikationslevel durch. Zur Berechnung der Selektivitätswerte der Tet-Enzyme erhielt ich in Zusammenarbeit die entsprechenden LC-MS/MS-Daten.

4.1.5 Materialien und Methoden

mES-Zellkultur

mES-Zellen (R1-Linie; *Serum/primed*) wurden kultiviert wie im Experimentalteil von Abschnitt 3.6 beschrieben:[13] R1-Zellen, die in DMEM mit FBS, LIF und 2i-Bedingungen wuchsen, wurden zweimal im gleichen Medium aber ohne 2i passagiert (5 Tage). 24 h vor der DNA-Isolation wurde Vitamin C mit einer Endkonzentration von 100 μM hinzugegeben. Die LC-MS/MS-Daten der mES-Zelldifferenzierung wurden von Abschnitt 3.6 verwendet (*Figure 5*).[13]

HEK-293T-Zellenkultur und Transfektionsbedingungen

Die Zellkultivierung und die Transfektion von GFP-mTet1CD, GFP-mTet1FL oder GFP-mTet3FL erfolgte wie im Experimentalteil von Abschnitt 3.6 beschrieben.[13] 16 h nach der Transfektion wurde das Medium gewechselt, wobei Vitamin C mit einer Endkonzentration von 100 μM hinzugefügt wurde. Nach 40 h wurde das Medium (+100 μM Vitamin C) erneut gewechselt und nach 48 h wurden die Zellen geerntet.

DNA-Isolierung, DNA-Verdau und UHPLC-UV-ESI-MS/MS-Analyse

Siehe Experimentalteil von Abschnitt 3.6.[13]

4.2 Untersuchung der Tet-vermittelten DNA-Demethylierung mittels eines neuen Zellkern-basierten *Assays*

4.2.1 Einleitung

Wie im vorherigen Abschnitt 4.1 gezeigt wurde, bewirkt die ektopische Expression der katalytischen Domäne von mTet1 (GFP-mTet1CD) in HEK-293T-Zellen eine ausgeprägte DNA-Demethylierung, die sich sowohl durch passive Verdünnung der Oxidationsprodukte von 5mdC bei der Zellteilung als auch durch aktive Prozesse erklären lassen könnte (siehe 4.1.2, **Abbildung 8**).[84] Bei den aktiven Prozessen geht man bislang hauptsächlich von einer Tdg-vermittelten Basenexzisionsreparatur von 5fdC und 5cadC aus.[7, 9, 13, 84, 312] Da dieser Mechanismus allerdings die intermediäre Bildung abasischer Stellen beinhaltet, könnte dies bei globalen Demethylierungsereignissen über vermehrte Strangbrüche zu genomischer Instabilität führen. Aus diesem Grund sind die direkten C−C-Bindungsbruchreaktionen von 5hmdC, 5fdC und insbesondere von 5cadC nachwievor eine attraktive Alternative.[3, 11-12, 76, 84, 331] Trotz der in Abschnitt 3.3[11] gefundenen decarboxylierenden Aktivität von mES-Zellkernextrakten, fehlen entsprechende Nachweise, ob solche Reaktionen *in vivo* in einer beträchtlichen Menge auftreten und ob ein entsprechendes Enzym existiert. Im Hinblick auf diesen alternativen Mechanismus wird im nachfolgenden Abschnitt der Ursprung der globalen DNA-Demethylierung in HEK-293T-Zellen untersucht, die bei der Überexpression von GFP-mTet1CD beobachtet wurde (**Abbildung 8**).

Bisherige Erkenntnisse zum Thema des direkten C−C-Bindungsbruchs

Wie in Abschnitt 3.3 und 3.4 gezeigt wurde, kann der direkte C−C-Bindungsbruch der oxidierten Derivate von 5mdC über eine Thiol-vermittelte transiente Absättigung der C(5)=C(6)-Doppelbindung ablaufen. Dies ermöglicht dann die Dehydroxymethylierung von 5hmdC, die Deformylierung von 5fdC bzw. die Decarboxylierung von 5cadC. Über diesen Mechanismus kann die jeweilige Nukleobase unter Freisetzung von Formaldehyd, Ameisensäure respektive Kohlenstoffdioxid unter Eliminierung des Thiols wieder rearomatisieren (**Schema 13**).[11-12] Die Methylierung von dC zu 5mdC durch die Dnmts verläuft nach einem sehr ähnlichen Mechanismus, weswegen diese Enzyme ebenfalls als potentielle Kandidaten für die Rückreaktion, also die DNA-Demethylierung, angesehen wurden.[11-12, 309] Unter den in Abschnitt 3.4 beschriebenen chemisch-synthetischen Bedingungen, die eine entsprechende enzymatische Aktivität simulieren sollte, war die Decarboxylierung von 5cadC etwa 10mal effizienter als die Deformylierung von 5fdC und

60mal effizienter als die Dehydroxymethylierung von 5hmdC.[11-12] Kürzlich konnte die Gruppe von *Klimasauskas* zeigen, dass bakterielle Dnmts diese Reaktionen *in vitro* tatsächlich katalysieren können.[331] Allerdings waren die Enzyme nur in der Lage 5hmdC und 5cadC direkt in dC umzuwandeln, während dies für 5fdC nicht beobachtet werden konnte. Als Grund wurde angeführt, dass die aktiven Taschen der Dnmts für das Hydrat der Formylgruppe von 5fdC zu klein sind, welches für die Deformylierung unter Freisetzung von Ameisensäure benötigt wird (**Schema 13**). Im Gegensatz zu den bakteriellen Dnmts zeigten die orthologen Enzyme in der Maus und im Menschen (Dnmt1, Dnmt3a/3b) jedoch kaum eine Reaktivität, die durch physiologische Konzentrationen des Co-Faktors SAM sogar gänzlich verloren ging. Schlussfolgernd kommen die Dnmts von Säugetieren für die C–C-Bindungsbruchreaktionen *in vivo* wahrscheinlich nicht in Frage.[331-332] Die Erkenntnisse zeigen aber zusammenfassend, dass verwandte Enzyme prinzipiell über diesen Thiol-vermittelten Mechanismus die C–C-Bindungsbruchreaktionen der oxidierten Derivate bewerkstelligen könnten.
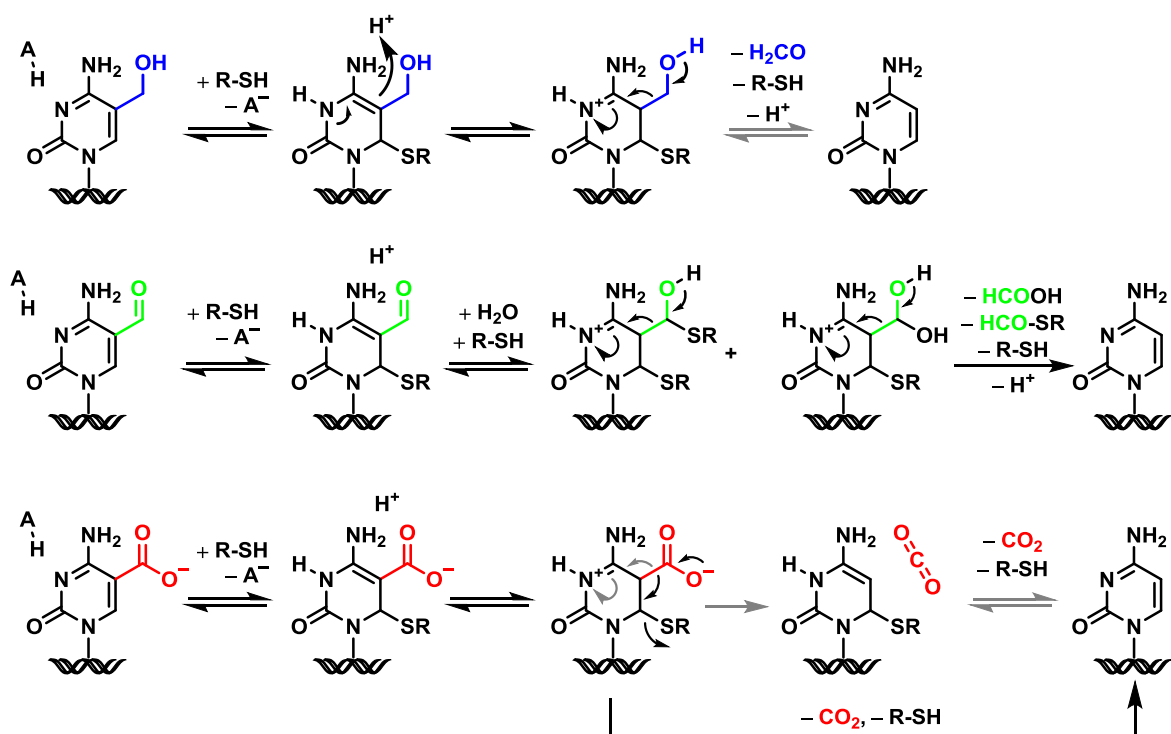
Schema 13 | Vorgeschlagene Mechanismen zum Thiol-vermittelten C–C-Bindungsbruch von 5hmdC, 5fdC oder 5cadC z.B. durch DNA-Methyltransferasen.[11-12, 309, 331]

Ein stichhaltiger Hinweis für alternative Entfernungsmechanismen, wurde jüngst von der Gruppe von *Guo-Liang Xu* veröffentlicht. Die neue Studie lieferte Beweise, dass nach der Befruchtung der Eizelle das Genom durch Tet3 oxidiert wird, wobei neben der passiven

103

Verdünnung der Oxidationsprodukte auch ein aktiver Entfernungsmechanismus vorhanden ist, der unabhängig von der Tdg-vermittelten Basenexzision ist. Als alternative Mechanismen könnten deshalb entweder andere Glykosylasen verantwortlich sein oder die bereits genannten C–C-Bindungsbruchreaktionen in Frage kommen.[95] Naheliegend wäre, dass Tet-Enzyme möglicherweise selbst diese C–C-Bindungsbruch-Reaktivität innehaben könnten. Dies wurde bisher noch nicht hinlänglich untersucht.[333] In Analogie zur Isoorotatdecarboxylase,[125, 253, 334] könnte ein ähnliches Enzym oder die Tet-Enzyme mit Hilfe eines aktivierten Wassermoleküls den C–C-Bindungsbruch bewerkstelligen (**Schema 14A,B**). Darüber hinaus wäre ein radikalischer Mechanismus denkbar, in dem das Hydrat von 5fdC oxidativ decarboxyliert wird (**Schema 14C**). Das aktive Zentrum der Tet-Enzyme würde für diese Reaktion im Gegensatz zu den Dnmts genügend Platz bieten.[128]
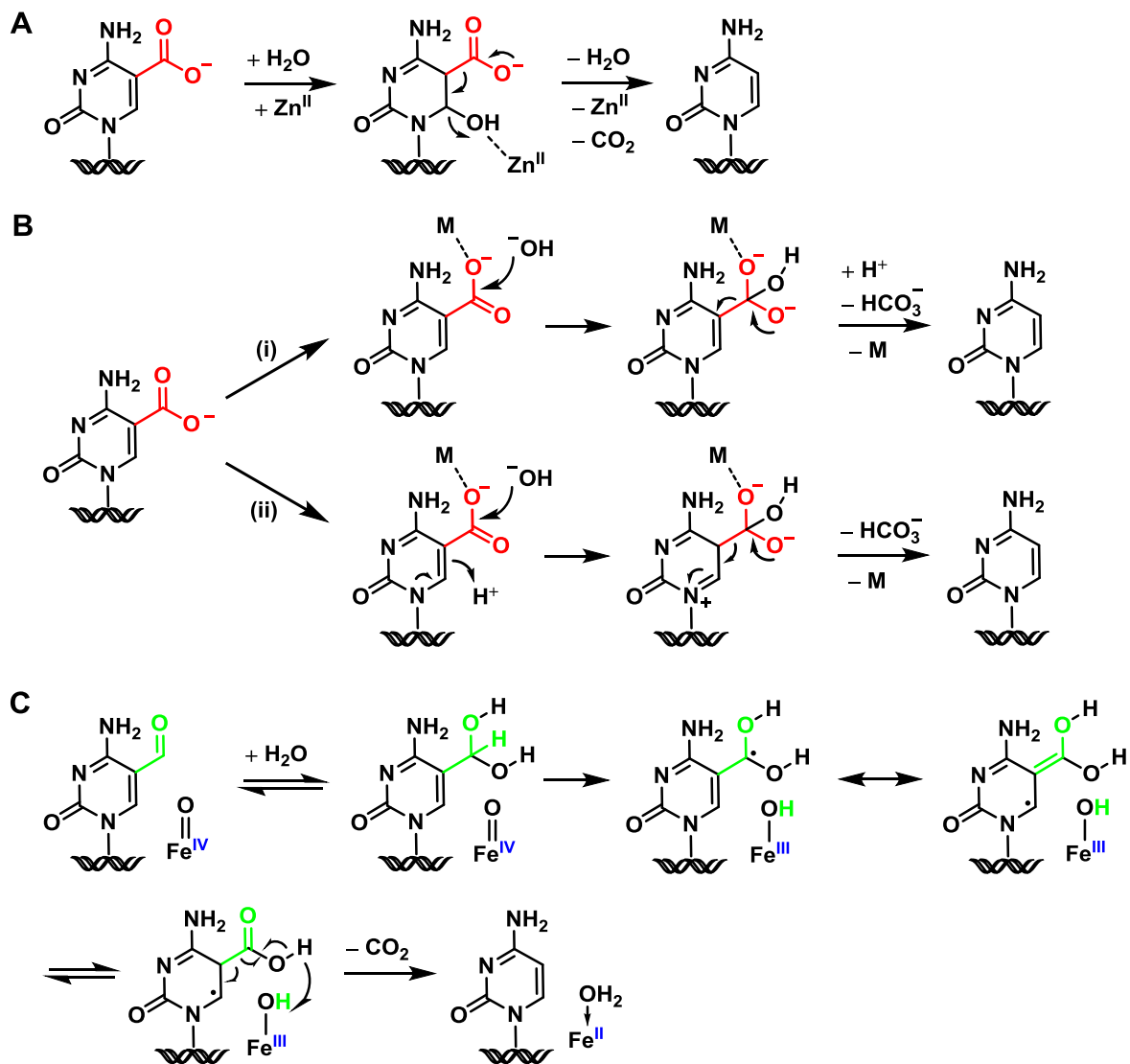
Schema 14 | Mögliche Thiol-unabhängige Mechanismen der C–C-Bindungsbruchreaktionen. (**A**) Möglicher Mechanismus von Übergangsmetal-katalysierten nicht-oxidativen Reaktionen adaptiert

nach *Martynowski et al.*[334-335] (**B**) Möglicher Mechanismus in Analogie zur Decarboxylierung von 5-Carboxyuracil (Isoorotat) durch die Isoorotatdecarboxylase nach *Xu et al* (Weg (i)).[125] Dieser Mechanismus würde allerdings über die intermediäre Bildung eines hochenergetischen Vinylanions ablaufen (siehe Abschnitt 3.4).[12] Als Alternative könnte diese Reaktion auch nach Weg (ii) im Sinne von *Martynowski et al.* ablaufen.[335] Bei diesem Mechanismus hätte der Stickstoff (N1) der glykosidischen Bindung eine formal positive Ladung, wodurch als Begleitreaktion die Deglykosylierung der Nukleobase von der DNA eintreten könnte. M symbolisiert ein Übergangsmetallion (z.B. Zn oder Fe). (**C**) Alternativ könnte 5fdC durch Tet-Enzyme oxidativ decarboxyliert werden. Die Oxidation von 5fdC könnte chemisch gesehen eine vorhergehende Hydrat-Bildung voraussetzen.[6, 134] Nach H-Abstraktion durch die Fe(IV)-Oxo-Spezies ist das sp^3-Radikal des 5fdC-Hydrats durch Nachbargruppeneffekte stabilisiert (allylische Doppelbindung und vicinale Sauerstoffatome). Dieses stabilisierte Radikal kann dann entweder zu 5cadC weiter oxidieren[8] oder könnte nach Enol/Keto-Tautomerisierung oxidativ decarboxylieren. Auch bei diesem Mechanismus könnten durch radikalischen Zerfall abasische Stellen entstehen (siehe Abschnitt 4.3, Schema 16).

4.2.2 Ergebnisse und Diskussion

Entwicklung eines neuen Zellkern-basierten *Assays*

Für die Untersuchung möglicher C−C-Bindungsbruchreaktionen in GFP-mTet1CD-transfizierten HEK-293T-Zellen wurde der in Abschnitt 3.3 erstmals vorgestellte Isotopenverfolgungs-*Assay* weiterentwickelt. Dieser beruhte auf der Inkubaton Isotopen-markierter 5cadC-enthaltender Oligodesoxynukleotide mit mES-Zellkernextrakten, anschließender Reisolation der DNA per Biotin/Streptavidin-Anreicherung und massen-spekrometrischer Analyse.[11] **Abbildung 13** fasst den neuen Arbeitsablauf schematisch zusammen. Als Reportermoleküle wurden erneut Oligodesoxynukleotide synthetisiert, die [$^{15}N_2$]-markierte Derivate von 5fdC und 5cadC enthielten. Durch C−C-Bindungsbruch-reaktionen entstandenes [$^{15}N_2$]-dC ist so von natürlichem dC aus der Zelle per Massenspektrometrie zweifelsfrei unterscheidbar. Im Gegensatz zu der früheren Methode in Abschnitt 3.3[11] wurden intakte Zellkerne eingesetzt.[336] Moleküle mit einer Größe von bis zu 40 kDa und somit auch die synthetisierten Oligodesoxynukleotide können durch die Kernporen passiv diffundieren (**Abbildung 13**).[337-338] Dies bot zwei Vorteile: Zum einen befanden sich die Zellkerne nach ihrer Gewinnung in einem sehr nativen Zustand, d.h. der mögliche Verlust von essenziellen Cofaktoren oder die Denaturierung von Proteinkomplexen bei der Präparation von Zellkernextrakten konnte so minimiert werden. Zum anderen konnten die Zellkerne nach der Reaktion durch einfaches Zentrifugieren abgetrennt werden, während sich der Großteil der synthetischen Oligodesoxynukleotide im Überstand befand. Auf diese Weise konnte der Hauptteil der genomischen DNA in einem Schritt abgetrennt werden, die

eine potenzielle Störquelle[V] für die spätere MS-Analytik der isotopenmarkierten Oligodesoxynukleotide darstellt. Die Biotin/Streptavidin-Anreicherung aus Abschnitt 3.3[11] war deshalb nicht notwendig.
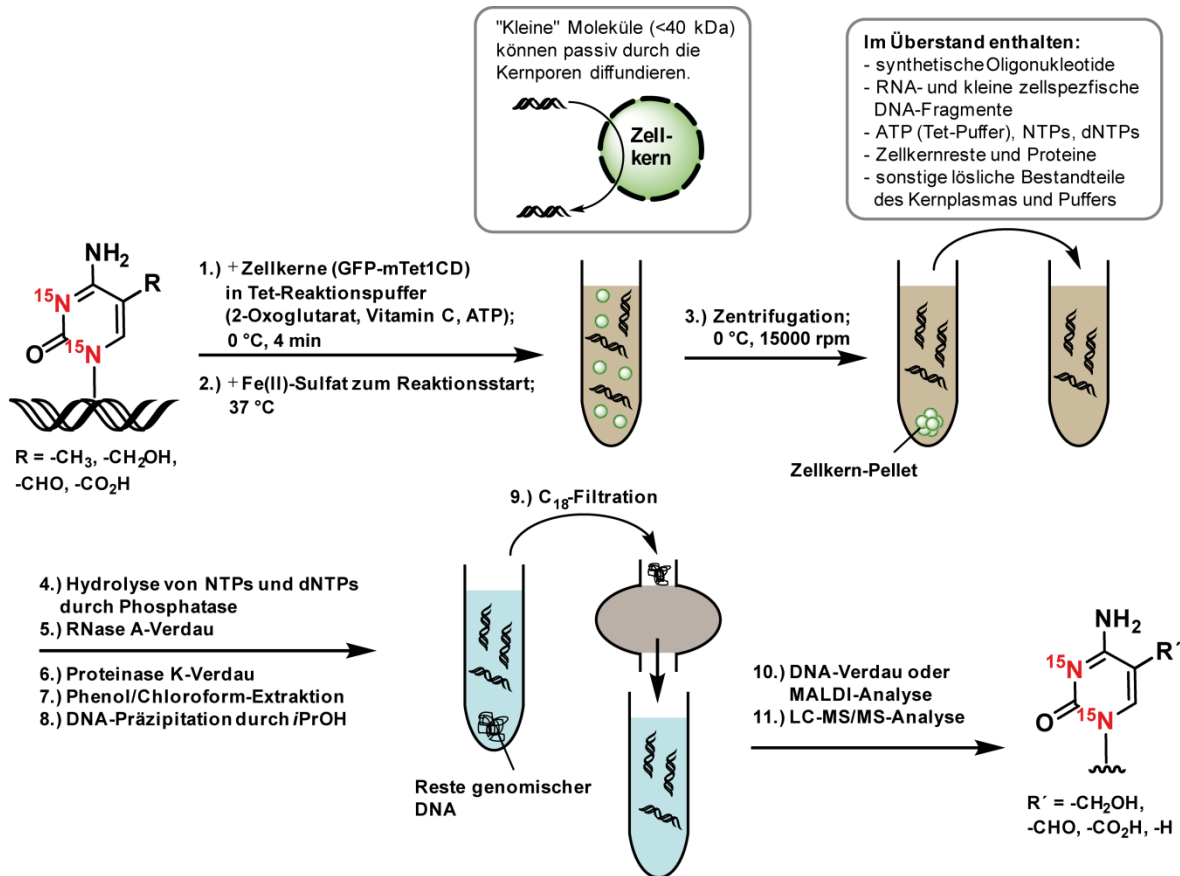
Abbildung 13 | Schematischer Arbeitsablauf des neu-entwickelten Zellkern-*Assays*. Synthetische Oligodesoxynukleotide, die [$^{15}N_2$]-markierte Cytosin-Derivate enthielten, wurden mit HEK-293T-Zellkernen inkubiert, die mit GFP-mTet1CD überexprimiert waren.

Optimierung der ektopischen Expression von GFP-mTet1CD

Da GFP-mTet1CD möglicherweise direkt für die DNA-Demethylierung in HEK-Zellen per C–C-Bindungsbruchreaktion verantwortlich sein könnte, wurde zunächst dessen Protein-konzentration durch Optimierung der Transfektions- und Expressionsbedingungen maximiert. Dabei bot das GFP-Fusionsprotein eine einfache Möglichkeit die Expressions-menge per Fluoreszenzmessung mit Hilfe eines *Tecan-Readers* und per FACS-Analyse zu kontrollieren. **Abbildung 14** zeigt links eine konfokal-mikroskopische Aufnahme von HEK-293T-Zellen mit einer maximalen Transfektionseffizienz von 75%. GFP-mTet1CD ist dabei hauptsächlich im Zellkern lokalisiert. Zusätzlich konnte die Expression durch eine

---

[V] Hohe Mengen genomischer DNA verursachen unter anderem durch die natürliche Isotopenverteilung (^{13}C, ^{15}N, ^{18}O) eine Signalüberlagerung mit der [$^{15}N_2$]-Isotopenmarkierung der synthetischen Oligodesoxynukleotide. Ferner führen hohe Mengen genomischer DNA zu einer Signalsuppression.

Behandlung der transfizierten Zellen mit Natriumbutyrat vervierfacht werden. Mechanistisch wird diese Expressionszunahme durch Inhibierung von Histondeacetylasen erklärt.[339-340] Unter diesen Bedingungen war GFP-mTet1CD auch im Zytosol in größeren Mengen vorhanden. Die Zellkernisolation wurde durch Zellaufschluss mit einem *Dounce*-Homogenisator und anschließender Zentrifugation bewerkstelligt. Typischerweise wird dies unter Verwendung einer Protease-Inhibitor-Cocktail-Tablette (*Roche* cOmplete, EDTA-frei) durchgeführt. Die Konzentration einer Tablette bezogen auf ein Puffervolumen von 40 mL darf hierbei nicht überschritten werden, da GFP-mTet1CD andernfalls irreversibel inhibiert wird. In **Abbildung 14** sind rechts die erhaltenen Zellkerne zu sehen.
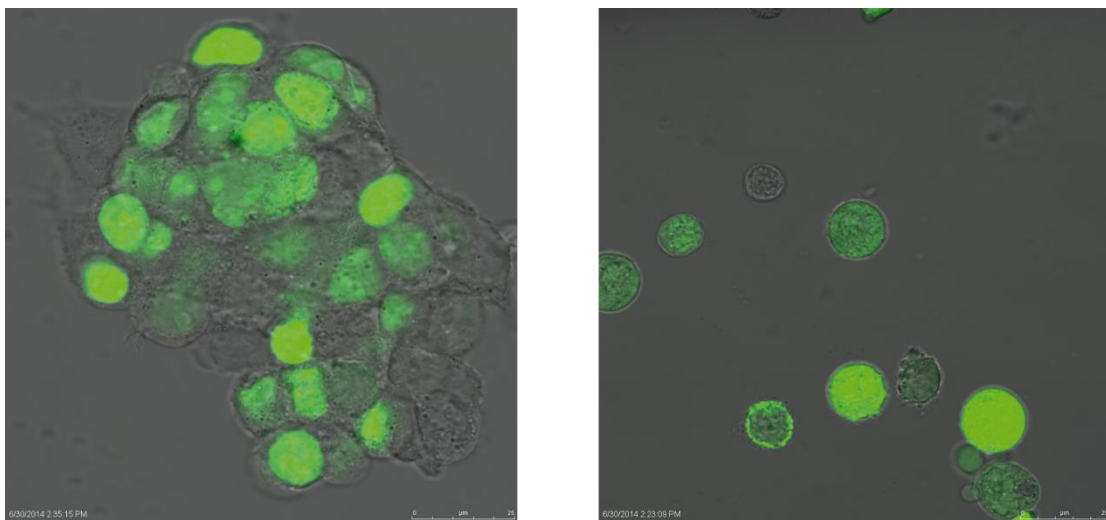
Abbildung 14 | Konfokalmikroskopie von HEK-293T-Zellen mit ektopischer Expression von GFP-mTet1CD. Fluoreszenzmikroskopische Aufnahme mit Lichtkanal-Überlagerung. Links: Intakte Zellen 36 h nach Transfektion mit einer Effizienz von 75% (FACS-Analyse); rechts: Zellkerne nach Isolation mit Hilfe eines *Dounce*-Homogenisators.

Probenaufreinigung

Wie bereits erwähnt, ermöglichte der Einsatz von intakten Zellkernen nach der Inkubation mit synthetischen Oligodesoxynukleotiden die einfache Abtrennung durch Zentrifugation (**Abbildung 13**). Nach der Abtrennung der Zellkerne wurde der Überstand dann schrittweise weiter aufgereinigt, um die Oligodesoxynukleotide in ausreichender Reinheit für die empfindliche massenspektrometrische Analyse zu erhalten. Zunächst wurden die dNTPs und NTPs mit Hilfe einer Phosphatase auf Nukleosid-Ebene hydrolysiert. Dies ist wichtig, da die Triphosphate bei der späteren alkoholischen DNA-Fällung co-präzipitieren und Reaktionspuffer von Tet-Enzymen viel ATP enthalten.[7, 330] Anschließend erfolgte eine Standard-Reinigung per RNase A- und Proteinase K-Verdau sowie Phenol/Chloroform-Extraktion und DNA-Präzipitation. Nach der DNA-Präzipitation befanden sich im DNA-

Pellet noch Reste von genomischer DNA, da während dem *Assay* Zellkerne zum Teil platzen oder die Abtrennung per Zentrifugation nicht vollständig ist. **Abbildung 15A** zeigt die Analyse dieser Reinigungsstufe mittels denaturierender Polyacrylamid-Gelelektrophorese. Der *Assay* wurde mit einem einzelsträngigen 13mer-Oligodesoxynukleotid durchgeführt, dessen Bande bei Inkubation mit 2 Millionen Zellkernen deutlich zu sehen ist. Bei Verwendung größerer Zellkernmengen (8 Millionen) und längeren Inkubationszeiten war an der Geltasche eine große Fraktion genomischer DNA zu erkennen. Zur Abtrennung dieser genomischen DNA wurde anschließend eine Filtration mit C_{18}-Sep-Pak-Kartuschen durchgeführt (**Abbildung 13**, Schritt 9). Molekulargewichtsfilter (*cut-off*-Filter) konnten nicht verwendet werden, da die synthetische DNA mit der genomischen DNA zurückgehalten wurde. Bei der Verwendung von 0.5 nmol eines einzelsträngigen Oligodesoxynukleotids (2 µg) und 4–6 Millionen Zellkernen (ca. 12–18 µg genomische DNA) wurden mit diesem Protokoll nach massenspektrometrischer Quantifizierung 10–20% der Oligodesoxynukleotide zurückgewonnen. Der Massenanteil von endogener DNA war dabei 20–40%.

Optimierung der Tet-vermittelten Oxidationsreaktion

Um sicherzustellen, dass die Aktivität der transfizierten Zellkerne bei deren Präparation oder unter den Reaktionsbedingungen nicht verloren ging, wurde zunächst die Oxidationskraft des Fe(II)- und 2-Oxoglutarat-abhängigen Tet-Enzyms innerhalb der Zellkerne untersucht und optimiert. Hierzu wurde ein kurzes 5mdC-enthaltendes einzelsträngiges Oligodesoxynukleotid mit der Sequenz 5′-GTA ATG XGG TAG G-3′ (X = 5mdC) gewählt, da kurze einzelsträngige Oligodesoxynukleotide durch Tet1CD *in vitro* mit hoher Effizienz oxidiert werden.[330] Die isolierten Zellkerne wurden zusammen mit dem Oligodesoxynukleotid und dem Tet-Reaktionspuffer (2-Oxoglutarat, Vitamin C, ATP, Hepes pH 7.9)[330] für 4 min auf Eis präinkubiert. Da Fe(II)-Lösungen durch Luftsauerstoff sehr schnell zu Fe(III) oxidieren, welches eine katalytisch inaktive Spezies darstellt, wurde Fe(II)-Sulfat erst zum Reaktionsstart zur Reaktionsmischung hinzugegeben. Ferner wurde der Reaktionspuffer und die Fe(II)-Lösung unter Sauerstoff-freier Atmosphäre (*Schlenk*-Bedingungen) hergestellt und bei −80 °C gelagert, um eine gleichbleibende Qualität für jeden Versuchstag zu gewährleisten. Zur Qualitätskontrolle wurden am Versuchtag gleiche Anteile der konzentrierten Lösungen von Fe(II)-Sulfat und des Reaktionspuffers (25x Stocklösungen) zusammengegeben. Pure Fe(II)-Gemische sind dabei zu Anfangs farblos. Luftsauerstoff bewirkt dann einen dunkelbraunen Farbumschlag, gefolgt von der Bildung eines dunkelgrünen Fe(III)-Niederschlags.

A

Marker | 2 Mio Zellkerne, 30 min | 8 Mio Zellkerne, 45 min

genom. DNA

13mer —
12mer —

C

5′-GTA ATG **X**GG TAG G-3′ X= 5mC

unbehandelter DNA-Strang

4065
[M–H]⁻
berechnet: 4066

3946*

4087
[M–2H+Na]⁻
berechnet: 4088

MS-Signal-Intensität [a.u.]

m/z

B

MS-Signal-Intensität [a.u.]

2789 2805 3094 3109
–T   –A
3407 3423
–G
3736
3752
–G
3945* 3960*
4065
4081

m/z

D

Modifikationsgehalt [%]

5mdC 5hmdC 5fdC 5cadC

Exonuclease-Aktivität 3′→5′	X=5mC	5hmC
	m/z [M–H]⁻	m/z [M–H]⁻
5′-GTA ATG **X**GG TAG G-3′	4066	4082
5′-GTA ATG **X**GG TAG	3737	3753
5′-GTA ATG **X**GG TA	3408	3424
5′-GTA ATG **X**GG T	3095	3111
5′-GTA ATG **X**GG	2791	2807

Abbildung 15 | Stabilität und Reaktivität eines 5mdC-enthaltenden Oligodesoxynukleotids bei Inkubation mit HEK-293T-Zellkernen. (**A**) Denaturierende Polyacrylamid-Gelelektrophorese (20%iges Gel) von Zellkerninkubationen mit 5′-GTAATGXGGTAGG-3′ (X=5mC). Ohne C₁₈-Sep-Pak-Filtration (Schritt 9 in Abbildung 13) ist bei Verwendung höherer Mengen von Zellkernen und längeren Inkubationszeiten genomische DNA zu sehen. Das Oligodesoxynukleotid wird während der Inkubation in zum Teil per Exonukleaseaktivität abgebaut. (**B**) MALDI-TOF-Spektrum des isolierten Produkt-gemisches nach Inkubation von 1 nmol Oligodesoxynukleotid mit 15 Millionen Zellkernen für 60 min bei 37 °C und 300 rpm. Transfizierte HEK-Zellen wurden mit Natriumbutyrat zur Expressions-steigerung von GFP-mTet1CD behandelt. Neben einer deutlichen Oxidation von 5mdC zu 5hmdC ist eine Exonuklease-Aktivität von 3′→5′ für den Abbau des Oligodesoxynukleotids verantwortlich. In der Tabelle sind die berechneten Molekülmassen aufgelistet. Diese entsprechen den Signalen des Spektrums mit einer akzeptablen Abweichung von 1 amu. Die mit (*) gekennzeichneten Massen

entsprechen einer synthetischen Verunreinigung des eingesetzten Oligodesoxynukleotids. (**C**) MALDI-TOF-Spektrum des unbehandelten Oligodesoxynukleotids zum Vergleich. (**D**) LC-ESI-MS/MS-Quantifizierung der Nukleosid-Zusammensetzung nach enzymatischem DNA-Verdau des Reaktionsgemisches von (B). Dargestellt sind die technischen Mittelwerte ± Standardabweichung von Duplikatsmessungen.

Die Reaktion der Zellkerne mit dem Oligodesoxynukleotid wurde durch Abkühlen auf Eis beendet. Die Kerne wurden anschließend abzentrifugiert und das eingesetzte Oligodesoxynukleotid, wie oben skizziert, aufgereinigt (**Abbildung 13**). In **Abbildung 15B** ist ein typisches MALDI-TOF-Spektrum des isolierten Produktgemisches nach 60 min Reaktionszeit zu sehen; **Abbildung 15C** zeigt das Spektrum des unbehandelten DNA-Stranges zum Vergleich. Rechts neben dem Edukt-Signal bei *m/z* 4065 wurde ein neues Signal bei *m/z* 4081 mit einer relativen Intensität von ca. 2/3 erhalten, welches der berechneten Molekülmasse (*m/z* 4082) des 5hmdC-enthaltenden Produktstranges entspricht. Daneben zeigen wiederkehrende Signalmuster im linken Teil des Spektrums den graduellen Abbau des Oligodesoxynukleotids während der Zellkerninkubation an. Der beobachtete Massenverlust jeweils eines Nukleotids entspricht einer Exonuklease-Aktivität, die das Oligodesoxynukleotid vom 3′-Ende zum 5′-Ende hydrolysiert. Bei 15 Millionen Zellkernen im Verhältnis zu 1 nmol Oligodesoxynukleotid, konnte in 60 min der partielle Verlust von vier endständigen Nukleotiden beobachtet werden. Aus quantitativer Sicht fiel jedoch nur der Abbau des ersten Nukleotids ins Gewicht (man vergleiche mit **Abbildung 15A**), weshalb die Stabilität des Oligodesoxynukleotids den Anforderungen genügte.

Das Produktgemisch der Oligodesoxynukleotide wurde anschließend enzymatisch verdaut und das erhaltene Nukleosidgemisch per UHPLC-ESI-MS/MS nach der in Abschnitt 3.4 beschriebenen Isotopenverdünnungsmethode exakt quantifiziert. Die LC-MS/MS-Ergebnisse der quantitativen Produktanalyse bestätigten die MALDI-TOF-Analyse: 5mdC und 5hmdC lagen in Anteilen von 59.5% bzw. 38.5% bezogen auf die Gesamtmenge modifizierter Cytidine vor (**Abbildung 15D**). Daneben konnte auch die Bildung von 5fdC und 5cadC mit je 1.6% und 0.4% festgestellt werden.

Der Umsatz von 5mdC konnte durch Verlängerung der Reaktionszeit und durch mechanische Rotationsbewegung des Gemisches bis auf ~60% gesteigert werden (**Abbildung 16**). Zusammenfassend konnte somit ein funktionsfähiger *in vitro* Oxidations-*Assay* unter Verwendung von Zellkernen von GFP-mTet1CD-transfizierten HEK-293T-Zellen etabliert werden.
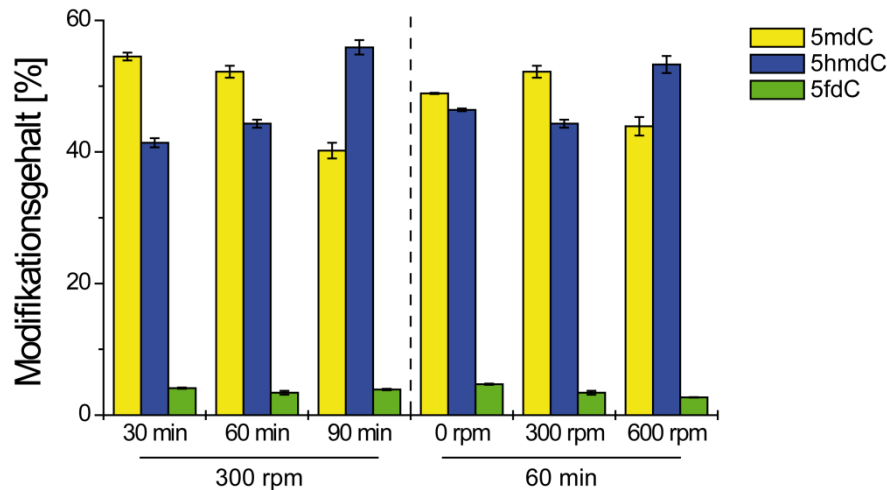
Abbildung 16 | Reaktionsoptimierung der Zellkern-basierten Oxidationsreaktion mit GFP-mTet1CD. LC-ESI-MS/MS-Ergebnisse der isolierten Produktgemische nach Inkubation von 500 pmol Oligodesoxynukleotid (5′-GTA ATG XGG TAG G-3′; X=5mC) mit 6 Millionen Zellkernen bei 37 °C. Transfizierte HEK-Zellen wurden mit Natriumbutyrat zur Expressionsteigerung von GFP-mTet1CD behandelt. Dargestellt sind die technischen Mittelwerte ±Standardabweichung von Duplikats-messungen.

Analyse der DNA-Demethylierung per Isotopenverfolgung

Nach der Optimierung des *Assays* wurde die Analyse der DNA-Demethylierungsreaktionen per Isotopenverfolgung durchgeführt. Als Reportermoleküle kamen Oligodesoxynukleotide zum Einsatz, die $[^{15}N_2]$-markierte Derivate von X = 5mdC, 5fdC und 5cadC enthielten (5′-GTA ATG XGG TAT G-3′). Durch C–C-Bindungsbruchreaktionen entstandenes $[^{15}N_2]$-dC ist per Massenspektrometrie von natürlichem dC aus der Zelle unterscheidbar. In Abschnitt 3.3 wurde für die Detektion von $[^{15}N_2]$-dC ein hochauflösendes Massenspektro-meter (Orbitrap) verwendet, um die Unterscheidung von ^{13}C-, ^{15}N- oder ^{18}O-Spezies der natürlichen Isotopenverteilung zu gewährleisten.[11] Da die verwendete LC-MS-Methode für eine Quantifizierung von $[^{15}N_2]$-dC nicht sensitiv genug war,[11] erfolgte die Analyse der Reaktionsprodukte in diesem Abschnitt mit Hilfe eines Triple-Quadrupol-Tandemmassen-spektrometers (UHPLC-ESI-MS/MS). Die niedrigere Auflösung des Massenspektrometers war für dieses Experiment trotz eines Massenunterschieds von nur 2 amu durch die $[^{15}N_2]$-Markierung geeignet, da die genomische DNA der Zellkerne ausreichend abgetrennt werden konnte (siehe oben). Ferner nahmen die [M+2]-Isotopologen des endogenen dCs bei Verwendung des MS-Übergangs *m/z* 230/114 in Kontrollproben nur einen Anteil von 0.3% ein.

Die LC-MS/MS-Quantifizierung der $[^{15}N_2]$-markierten Reaktionsprodukte erfolgte durch Zugabe von internen Standards mit einer höheren oder niedrigeren Isotopenmarkierung (siehe Experimentalteil). **Abbildung 17** fasst die Ergebnisse der Isotopenverfolgungsexperimente

zusammen. Die Oxidation von [^{15}N$_2$]-5mdC-enthaltenden Oligodesoxynukleotiden lieferte als Hauptprodukt erneut [^{15}N$_2$]-5hmdC. **Abbildung 17A** zeigt den direkten Vergleich der Tet-vermittelten Oxidation von einzelsträngiger, hemi-methylierter und vollmethylierter doppelsträngiger DNA. Die Oxidation des hemi-methylierten Stranges war zwar relativ gesehen am stärksten, jedoch war die Anzahl an absoluten Oxidationsereignissen bei dem vollmethylierten Konstrukt am größten.
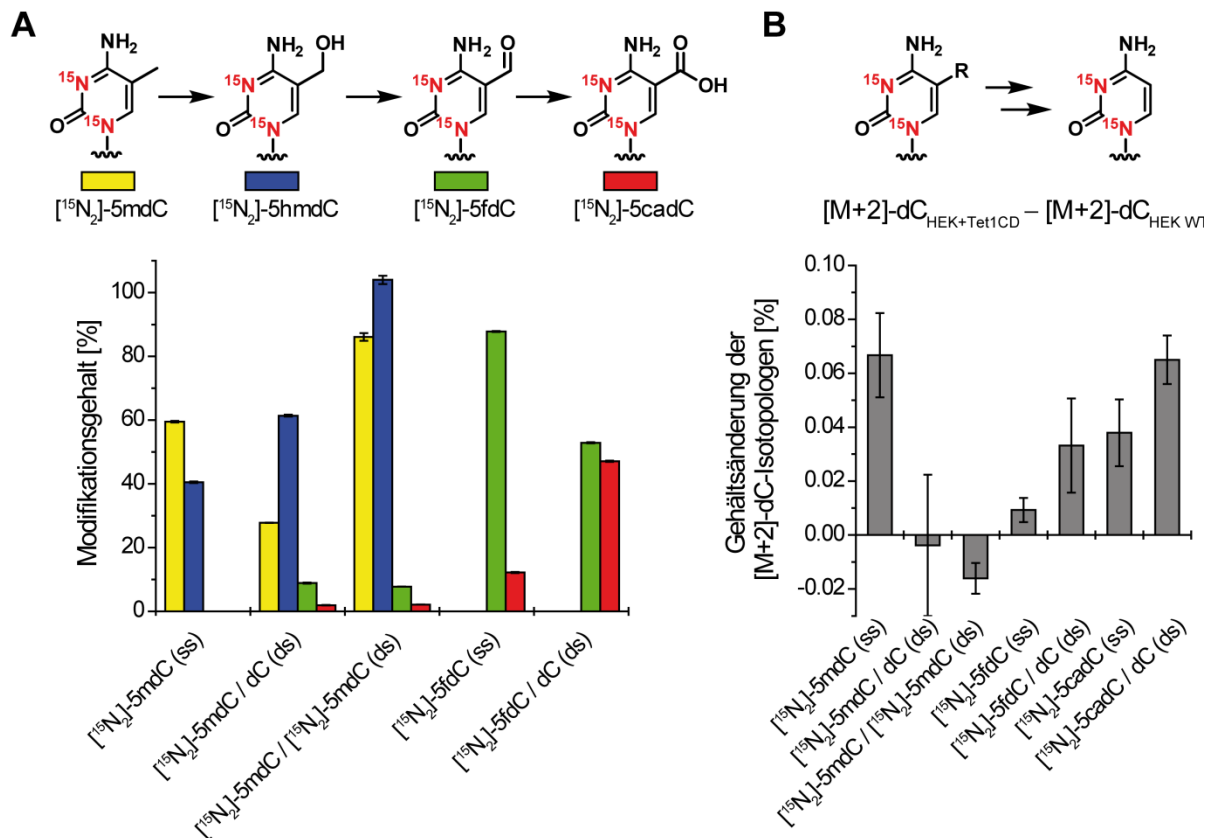
Abbildung 17 | Isotopenverfolgungsexperiment mit GFP-mTet1CD-Zellkernen. LC-ESI-MS/MS-Ergebnisse der isolierten Produktgemische nach Inkubation von 350 pmol einzelsträngigen (ss) oder doppelsträngigen (ds) DNA-Konstrukten mit 4 Millionen Zellkernen bei 37 °C/600 rpm für 90 min. Transfizierte HEK-Zellen wurden mit Natriumbutyrat zur Expressionsteigerung von GFP-mTet1CD behandelt. ss-DNA: 5′-GTA ATG XGG TAT G-3′ mit X = [^{15}N$_2$]-5mC, [^{15}N$_2$]-5fC und [^{15}N$_2$]-5caC); komplementäre DNA: 5′-CAT ACX GCA TTA C-3′ mit X = [^{15}N$_2$]-5mC und dC) (**A**) Quantifizierung der Oxidationsprodukte nach Umsatz der jeweiligen DNA-Konstrukte. Der Modifikationsgehalt wurde auf 100% mit Ausnahme des vollmethylierten Doppelstranges (200%) normalisiert. Dargestellt sind technische Mittelwerte ± Standardabweichung von Duplikatsmessungen. (**B**) Quantifizierung der C−C-Bindungsbruchreaktionen. Dargestellt ist die Differenz zwischen der Menge der [M+2]-dC-Isotopo-logen in HEK-293T-Zellen mit ektopischer Expression von GFP-mTet1CD und Wildtypzellen (WT). Fehlerbalken repräsentieren die Gaußsche Fehlerfortpflanzung der Standardabweichung aus je zwei technischen Replikaten.

[^{15}N$_2$]-5fdC-enthaltende Oligodesoxynukleotide wurden durch GFP-mTet1CD im Vergleich weniger häufig oxidiert. Dieses Ergebnis steht in Übereinstimmung mit den von *Ito et al.*

berichteten Initialgeschwindigkeiten der Oxidationsreaktionen von mTet2CD, die in der Reihenfolge 5mdC >> 5hmdC > 5fdC abnehmen.[8] Auch bei $[^{15}N_2]$-5fdC-enthaltenden Oligodesoxynukleotiden war die Oxidation des hemi-modifizierten Doppelstranges deutlich ausgeprägter als die des Einzelstranges (**Abbildung 17A**). Zusammenfassend waren demnach die doppelsträngigen DNA-Konstrukte ein besseres Oxidationssubstrat für mTet1CD als die einzelsträngigen Konstrukte. Dies steht in Widerspruch zu den Ergebnissen von *Kizaki* und *Sugiyama*, die *in vitro* eine höhere Reaktivität gegenüber Einzelsträngen berichteten.[330]

Obwohl eine hohe Oxidationsaktivität der Zellkernpräparation beobachtet werden konnte (bis zu ~75% 5mdC-Oxidation und ~50% 5fdC-Oxidation), war die Bildung von $[^{15}N_2]$-dC per C−C-Bindungsbruchreaktion durch GFP-mTet1CD nur im Spurenbereich (<0.1%) feststellbar (**Abbildung 17B**). Aus diesem Grund kommt die katalytische Domäne von Tet1 als Deformylase oder Decarboxylase unter den gegebenen Bedingungen nicht in Frage. In Wildtyp-Zellen konnte ebenfalls keine C−C-Bindungsbruch-Reaktivität von $[^{15}N_2]$-5fdC oder -5cadC beobachtet werden, sowie keine sich anschließend ereignende, potenziell denkbare Remethylierung zu $[^{15}N_2]$-5mdC (Daten nicht gezeigt).

4.2.3 Zusammenfassung

In diesem Abschnitt wurde die Etablierung eines neuen Zellkern-basierten *in vitro Assays* vorgestellt. Dieser ermöglichte die Untersuchung von Oxidations- und C−C-Bindungsbruch-reaktionen in GFP-mTet1CD-transfizierten HEK-293T-Zellen mit Hilfe von synthetischen Oligodesoxynukleotiden per Isotopenverfolgung. Während die Oxidation von 5mdC durch Tet1CD in hohen Ausbeuten beobachtet werden konnte, waren C−C-Bindungsbruch-reaktionen höchstens im Spurenbereich präsent. Tet1CD hat übereinstimmend mit Tet2CD[8] eine Substratpräferenz für 5mdC, wobei doppelsträngige DNA mit höherer Effizienz oxidiert wird, als einzelsträngige DNA. Die Untersuchungen der katalytischen Domänen von Tet2 und Tet3 sowie der Volllängen-Enzyme stehen bezüglich der C−C-Bindungsbruchreaktionen noch aus. In zukünftigen Arbeiten könnte der *Assay* generell für die Aktivitätsuntersuchung identifizierter DNA-bindender Proteine verwendet werden, um DNA-modifizierende oder DNA-reparierende Mechanismen in einer nativen Umgebung untersuchen zu können.[336, 338, 341]

4.2.4 Projektbeiträge in diesem Abschnitt

Unter enger Zusammenarbeit mit *Olga Kotljarova*, *Edris Parsa*, *Benjamin Hackner* und *Thomas Wildenhof* entwickelte ich den *Assay*. Meine maßgeblichen Beiträge waren die

Entwicklung der Zellkerninkubation, die Etablierung des DNA-Isolationsprotokolls, die Aktivitätsoptimierung und die LC-MS-Analytik. *Edris Parsa* und *Benjamin Hackner* optimierten die Zellkultur- und Transfektionsbedingungen. Sie wurden durch *Thomas Wildenhof* und *Olga Kotljarova* unterstützt. *Olga Kotljarova* und *Benjamin Hackner* führten die Zellkernisolation durch. *Arne S. Schröder* und *Barbara Steigenberger* synthetisierten die modifizierten Oligodesoxynukleotide. *Arne S. Schröder* synthetisierte ferner [D$_3$,^{15}N$_2$]-5mdC als internen MS-Standard.

4.2.5 Materialien und Methoden

Zellkernpräparation

Intakte Zellkerne wurden nach einer geringfügig modifizierten Methode von *Dignam et al.* präpariert.[342-343] Das Zellpellet wurde mit einem fünffachen Volumen von Puffer A resuspendiert, der 10 mM Hepes pH 7.9, 1.5 mM MgCl$_2$, 10 mM KCl, 0.5 mM frisches DTT und eine EDTA-freie Protease-Inhibitor-Cocktail-Tablette (*Roche* cOmplete) enthielt. Eine Inhibitor-Tablette wurde in 40 mL Puffer gelöst. Die Zellsuspension wurde für 10 min auf Eis inkubiert, anschließend mit einer *Eppendorf*-Zentrifuge (5427 R) pelletiert (2000 rpm, 15 min) und erneut mit zweifachem Volumen von Puffer A resuspendiert. Die Zellen wurden mit einem *Dounce*-Homogenisator (25 Züge) aufgeschlossen und die Lyse durch Färbung mit Trypanblau geprüft. Die Zellkerne wurden pelletiert (2000 rpm, 15 min) und der Überstand entfernt. Nach Resuspensdierung der Zellkerne in Puffer A, wurden die Zellkerne auf die Reaktionsgefäße aliquotiert und Puffer A durch erneute Zentrifugation (2000 rpm, 15 min) entfernt.

Reaktionspuffer

Der Tet-Reaktionspuffer (25faches Konzentrat: 1.25 M Hepes pH 7.9, 2.5 M NaCl, 50 mM Vitamin C, 25 mM 2-Oxoglutarat-Mononatriumsalz, 30 mM ATP, 62.5 mM DTT) und Fe(NH$_4$)$_2$(SO$_4$)$_2$×6H$_2$O (25faches Konzentrat: 2.63 mM) wurden separat unter *Schlenk*-Bedingungen in einer Stickstoffatmosphäre hergestellt und bei −80 °C gelagert. Die Substanzen wurden in einem *Schlenk*-Kolben in festem Zustand eingewogen und mit entgastem H$_2$O gelöst. Entgastes H$_2$O (*MilliQ*-Reinheit) wurde in drei Zyklen durch Einfrieren in flüssigem N$_2$, Entgasen im Hochvakuum und Auftauen im Unterdruck erhalten.

Durchführung des *in vitro Assays* und DNA-Reinigung

Zu den in einem 1.5 mL Mikroreaktionsgefäß vorgelegten GFP-mTet1CD-transfizierten Zellkernen (3−15 Millionen) wurden auf Eis schrittweise der vorverdünnte Reaktionspuffer

und das Oligodesoxynukleotid (10 µL, 0.35−1 nmol) gegeben und anschließend für 4 min auf Eis inkubiert. Die Fe(II)-Lösung (6 µL, 25x) wurde hinzugegeben und das Reaktionsgemisch durch drei- bis viermaliges Auf- und Abpipettieren gemischt ohne Luftblasen zu erzeugen. Das Reaktionsvolumen betrug 150 µL. Die Endkonzentration des Reaktionspuffers war 50 mM Hepes (pH 7.9), 100 mM NaCl, 2 mM Vitamin C, 1 mM 2-Oxoglutarat-Mononatriumsalz, 1.2 mM ATP, 2.5 mM DTT und 0.11 mM $Fe(NH_4)_2(SO_4)_2 \times 6H_2O$.[330] Zum Reaktionsstart wurden die Proben in einen 37 °C-temperierten Thermomixer (*Eppendorf comfort*) überführt (300 rpm). Der Reaktionsstopp erfolgte durch Zugabe von eiskaltem H_2O (75 µL) und Abkühlen auf Eis. Die Zellkerne wurden pelletiert (45 min, 15000 rpm, 0 °C), der Überstand überführt und das Zellkernpellet erneut mit eiskaltem H_2O (125 µL) versetzt und pelletiert (15 min, 15000 rpm, 0 °C). Die beiden Überstände wurden vereinigt und erneut zentrifugiert (45 min, 15000 rpm, 0 °C). Zu dem finalen Überstand (300 µL) wurde $ZnSO_4$ (7.5 µL, 4 mM) und $MgCl_2$ (10 µL, 30 mM) und *Antarctic Phosphatase* (10 µL, 5 U/µL, *New England Biolabs*) hinzugegeben und die Mischung bei 37 °C für 1.5 h inkubiert (0 rpm). Anschließend erfolgte ein RNase A-Verdau (10 µL, 7 U/µL, DNase-frei, *Qiagen*) bei 50 °C für 1 h, gefolgt von einem Proteinase K-Verdau (30 µL, 600 mAU/mL, *Qiagen*) bei 50 °C für 2 h. Das Reaktionsgemisch wurde zweimal mit *Roti*-Phenol/Chloroform/Isoamylalkohol (350 µL, *Carl Roth*) extrahiert (Phasentrennung durch Zentrifugation für 10 min bei 0 °C, 15000 rpm). Bei HEK-293T-Proben, deren Expression mit Natriumbutyrat gesteigert wurde,[339-340] musste die Extraktion viermal durchgeführt werden, um die Proteinreste aus der wässrigen Phase zu entfernen. Danach wurde zweimal mit $CHCl_3$ (350 µL) extrahiert (Phasentrennung durch Zentrifugation für 5 min bei 0 °C, 15000 rpm) und die wässrige Phase in einer *Speedvac* (RVC-2-33 IR, *Christ*) bei 60 °C auf 200 µL eingeengt.

Für MALDI-TOF-Analysen wurde eine 7 µL-Probe über Nacht auf einer 0.025 µm-Membran (*VSWP, Millipore*) gegen H_2O dialysiert. Anschließend erfolgte die Präzipitation der kurzen einzelsträngigen Oligodesoxynukleotide nach Zugabe von NaOAc (20 µL, 3 M) und absolutem *i*PrOH (600 µL) durch Zentrifugation über Nacht (15000 rpm, 0 °C). Das DNA-Pellet wurde mit −20 °C-kaltem EtOH (75% v/v) zweimal gewaschen und für je 30 min bei 15000 rpm und 0 °C zentrifugiert. Das DNA-Pellet wurde für 3 min in einer *Speedvac* (40 °C) getrocknet und anschließend in 1 mL H_2O gelöst. Zur Abtrennung von genomischen DNA-Resten wurde die Lösung auf 65 °C für 5 min (1400 rpm) erhitzt, auf Eis abgeschreckt und durch eine konditionierte Sep-Pak-C_{18}-Kartusche (*Waters*) mit einem Fluss von 1 mL min^{-1} gefiltert. Die Kartusche wurde mit MeCN (80% v/v, 5 mL) gespült und die vereinigten

Filtrate in einer *Speedvac* (50 °C) bis zur Trockne eingeengt. Das DNA-Pellet wurde für die enzymatische Hydrolyse in H_2O gelöst (ca. 2–10 µM Endkonzentration) und bei −20 °C gelagert.

MALDI-TOF-Analyse intakter Oligodesoxynukleotide

MALDI-TOF Spektren wurden mit einem *Bruker*-Autoflex II-Massenspektrometer und einem AnchorChip™-Target (*Bruker MTP* var/384) generiert. Entsalzte Oligodesoxynukleotide wurden auf einer HPA-Matrix co-kristallisiert. Zusammensetzung der Matrix: 3-Hydroxypicolin-Säure (50 mg), 15-Krone-5 (10 µL), Ammonium-Hydrogencitrat (10 mg) in 500 µL H_2O und 500 µL MeCN.

DNA-Verdau für die LC-ESI-MS/MS-Analyse

Die gereinigten Oligodesoxynukleotid-Proben (10–30 pmol) wurden nach dem Protokoll im Experimentalteil von Abschnitt 3.6 enzymatisch hydrolysiert.[13] Die erste Inkubationsrunde (*Antarctic Phosphatase*, Nuklease S1) dauerte 3 h und erfolgte unter vorheriger Zugabe der internen Standards für die LC-ESI-MS/MS-Quantifizierung (siehe unten), die zweite Inkubationsrunde (*Snake Venom Phosphodiesterase I*) erfolgte über Nacht. Die Proben wurden vor der LC-ESI-MS/MS-Analyse filtriert (AcroPrep^{TM} Advance 96 Filterplatte, 0.2 µm Supor®, *Pall Life Sciences*).

Für die Quantifizierung der Reaktionsprodukte des 5mdC-enthaltenden Oligodesoxynukleotids wurden folgende interne Standards zu der DNA-Verdau-Lösung gegeben: 29.4 pmol $[^{13}C_{10},^{15}N_2]$dT, 10.8 pmol $[^{13}C_9,^{15}N_3]$dC, 13.7 pmol $[D_3]$5mdC, 10.2 pmol $[D_2,^{15}N_2]$5hmdC, 1.52 pmol $[^{15}N_2]$5fdC, 1.08 pmol $[^{15}N_2]$5cadC und 109 fmol $[^{15}N_5]$8oxodG.

Für die Quantifizierung der Reaktionsprodukte der $[^{15}N_2]$5mdC-, $[^{15}N_2]$5fdC- und $[^{15}N_2]$5cadC-enthaltenden Oligodesoxynukleotide wurden folgende interne Standards zu der DNA-Verdau-Lösung gegeben: 121.9 pmol $[^{13}C_{10},^{15}N_2]$dT, 1.7 pmol $[^{13}C_9,^{15}N_3]$dC, 14.5 pmol $[D_3,^{15}N_2]$5mdC, 10.2 pmol $[D_2,^{15}N_2]$5hmdC, 10.3 pmol 5fdC, 8.28 pmol 5cadC und 109 fmol $[^{15}N_5]$8oxodG. Isotopen-markiertes dT und dC wurden von der *Hartmann Analytic GmbH* und $[^{15}N_5]$8oxodG wurde von *Cambridge Isotope Laboratories* erworben. $[D_3]$5mdC, 5fdC sowie 5cadC wurde nach Vorschriften von *Martin Münzel*,[3-4] und $[D_3,^{15}N_2]$5mdC von *Arne Schröder* synthetisiert. Alle anderen Nukleosid-Derivate wurden im Zuge von Abschnitt 3.4 in dieser Arbeit synthetisiert.[12]

LC-ESI-MS/MS-Analyse

Die quantitative LC-ESI-MS/MS-Analyse der hydrolysierten DNA-Proben erfolgte mit einem 1290 UHPLC-System von *Agilent*, welches mit einem 6490 Triplequadrupol-Massenspektrometer von *Agilent* verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C$_8$-Säule (*Agilent*, 2.7 μm, 2.1 mm × 150 mm) durchgeführt. Für chromatographische Details siehe Experimentalteil von Abschnitt 3.6.[13] Massenspektrometrische Parameter und die Auswertung (z.B. Eichgeraden) waren identisch zum Experimentalteil von Abschnitt 3.4.[12] **Tabelle 1** und **Tabelle 2** fassen die Substanz-abhängigen MS/MS-Parameter für die Messung der 5mdC-enthaltenden bzw. der [^{15}N$_2$]-markierten Oligodesoxynukleotide zusammen.

Tabelle 1 | Substanz-abhängige MS/MS-Parameter für die Analyse von Hydrolysaten 5mdC-enthaltender Oligodesoxynukleotiden. Mit einem Stern (*) gekennzeichnete Substanzen wurden als interner Standard für die Quantifizierung eingesetzt. Die Fragmentorspannung war 380 V. Delta EMV wurde auf 300 im positiven Modus und 800 im negativen Modus gesetzt. CE: *collision energy*; CAV: *collision cell accelerator voltage*; EMV: *electron multiplier voltage*.

Substanz	Vorläufer-ion (*m/z*)	MS1 Auflösung	Produkt-ion (*m/z*)	MS2 Auflösung	*Dwell time* [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 1.5–4.0 min								
[^{15}N$_2$]5cadC*	274.08	*enhanced*	158.03	*enhanced*	15	2	5	Positiv
5cadC	272.09	*enhanced*	156.04	*enhanced*	15	2	5	Positiv
[^{15}N$_2$,D$_2$]5hmdC*	262.12	*enhanced*	146.07	*enhanced*	15	20	1	Positiv
5hmdC	258.11	*enhanced*	142.06	*enhanced*	15	20	1	Positiv
[D$_3$]5mdC*	245.13	*enhanced*	129.09	*enhanced*	15	45	1	Positiv
5mdC	242.11	*enhanced*	126.07	*enhanced*	15	45	1	Positiv
[^{13}C$_9$,^{15}N$_3$]dC*	240.1	*enhanced*	119.1	*enhanced*	15	15	5	Positiv
dC	228.1	*enhanced*	112.1	*enhanced*	15	15	5	Positiv
Zeitsegment 4.0–6.0 min								
5hmdU	257.08	*wide*	214.07	*wide*	13	7	5	Negativ
5hmdU	257.08	*wide*	124.04	*wide*	13	7	5	Negativ
5fdU	255.06	*wide*	212.06	*wide*	13	6	5	Negativ
Zeitsegment 6.0–9.0 min								
[^{15}N$_5$]8oxodG*	289.08	*wide*	173.04	*wide*	17	9	7	Positiv
8oxodG	284.1	*wide*	168.05	*wide*	17	9	7	Positiv
[^{15}N$_2$]5fdC*	258.09	*enhanced*	142.04	*enhanced*	17	2	5	Positiv
5fdC	256.09	*enhanced*	140.05	*enhanced*	17	2	5	Positiv
[^{13}C$_{10}$,^{15}N$_2$]dT*	255.12	*enhanced*	134.06	*enhanced*	17	4	3	Positiv
dT	243.1	*enhanced*	127.05	*enhanced*	17	4	3	Positiv

Tabelle 2 | Substanz-abhängige MS/MS-Parameter für die Analyse von Hydrolysaten [^{15}N$_2$]-markierter Oligodesoxynukleotiden. Mit einem Stern (*) gekennzeichnete Substanzen wurden als interner Standard für die Quantifizierung eingesetzt. Die Fragmentorspannung war 380 V. Delta EMV wurde auf 250 im positiven Modus und 800 im negativen Modus gesetzt. CE: *collision energy*; CAV: *collision cell accelerator voltage*; EMV: *electron multiplier voltage*.

Substanz	Vorläufer-ion (*m/z*)	MS1 Auflösung	Produkt-ion (*m/z*)	MS2 Auflösung	*Dwell time* [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 1.5–4.0 min								
[^{15}N$_2$]5cadC	274.08	*enhanced*	158.03	*enhanced*	30	3	5	Positiv
5cadC*	272.09	*enhanced*	156.04	*enhanced*	30	3	5	Positiv
[^{15}N$_2$,D$_2$]5hmdC*	262.12	*enhanced*	146.07	*enhanced*	30	25	1	Positiv
[^{15}N$_2$]5hmdC	260.1	*enhanced*	144.06	*enhanced*	30	25	1	Positiv
5hmdC	258.11	*enhanced*	142.06	*enhanced*	30	25	1	Positiv
[D$_3$,^{15}N$_2$]5mdC*	247.13	*enhanced*	131.08	*enhanced*	30	45	1	Positiv
[^{15}N$_2$]5mdC	244.11	*enhanced*	128.06	*enhanced*	30	45	1	Positiv
5mdC	242.11	*enhanced*	126.07	*enhanced*	30	45	1	Positiv
[^{13}C$_9$,^{15}N$_3$]dC*	240.1	*enhanced*	119.1	*enhanced*	30	5	5	Positiv
[M+2]dC	230.1	*enhanced*	114.1	*enhanced*	30	5	5	Positiv
Zeitsegment 4.0–6.0 min								
5hmdU	257.08	*wide*	214.07	*wide*	30	7	5	Negativ
5hmdU	257.08	*wide*	124.04	*wide*	30	7	5	Negativ
5fdU	255.06	*wide*	212.06	*wide*	30	6	5	Negativ
Zeitsegment 6.0–9.0 min								
[^{15}N$_5$]8oxodG*	289.08	*wide*	173.04	*wide*	35	9	7	Positiv
8oxodG	284.1	*wide*	168.05	*wide*	35	9	7	Positiv
[^{15}N$_2$]5fdC	258.09	*enhanced*	142.04	*enhanced*	35	2	5	Positiv
5fdC*	256.09	*enhanced*	140.05	*enhanced*	35	2	5	Positiv
[^{13}C$_{10}$,^{15}N$_2$]dT*	255.12	*enhanced*	134.06	*enhanced*	35	4	3	Positiv
dT	243.1	*enhanced*	127.05	*enhanced*	35	4	3	Positiv

4.3 Entwicklung eines neuen Reagenzes zur Detektion und Quantifizierung von abasischen Stellen im epigenetischen Kontext

4.3.1 Einleitung

Seit der Entdeckung von 5fdC und 5cadC als Bestandteil der genomischen DNA embryonaler Stammzellen,[6-8] die durch Tet-vermittelte Oxidation von 5mdC entstehen, wird die Aufklärung der DNA-Demethylierungsmechanismen weltweit intensiv vorangetrieben. Als aktive Demodifizierungsprozesse werden die direkte Deformylierung bzw. Decarboxylierung[3, 11-12, 76, 84, 125, 331, 344] und vor allem reparaturassoziierte Prozesse diskutiert. Bei letzteren Prozessen nimmt man derzeit an, dass Tdg die Hauptrolle spielt, indem es die Basenexzisionsreparatur (BER)[345] durch Deglykosylierung von 5fdC und 5cadC einleitet.[7, 9, 14, 312, 317-318, 346-350] Die Bedeutung von Tdg wird durch dessen absolute Notwendigkeit für die embryonale Entwicklung unterstrichen. Dies ist insofern bemerkenswert, da *Knockouts* jeder anderen DNA-Glykosylase viel geringere phänotypische Effekte zur Folge haben und die jeweiligen Embryos im Gegensatz zu Tdg-*Knockouts* überleben.[312, 320, 350-351]

Da die Exzision von 5fdC und 5cadC durch Tdg aber intermediär abasische Stellen (auch apyrimidinische/apurinische Stellen; kurz AP-Stellen genannt) entstehen lässt, wäre es bezeichnend, wenn gerade dieser Weg der aktiven DNA-Demethylierung selektiert wurde: Bei der Reprogrammierung des Epigenoms könnte mit der Generierung von AP-Stellen ein erhöhtes Risiko für Mutationen[352-356] sowie Einzel- und Doppelstrangbrüche gerade in frühen Entwicklungsstadien in Kauf genommen werden.[84-85, 97, 105, 315, 320, 357]

AP-Stellen können im Zuge des epigenetischen DNA-Metabolismus aber nicht nur bei der Tdg-vermittelten Entfernung von 5fdC und 5cadC entstehen (**Schema 15**). Sofern als alternativer Demethylierungsmechanismus die enzymatische Desaminierung von 5mdC bzw. 5hmdC[VI] nachgewiesen werden kann, resultieren dabei promutagene T:G-[299-300, 302-303, 360-361] bzw. 5hmU:G-Fehlpaarungen,[312-313] die ebenfalls Substrate von Tdg und anderen Glykosylasen sind.[320] Desweiteren gibt es Hinweise auf eine *long-patch* BER-abhängige globale Demethylierung in Mauszygoten, die allerdings durch Desaminierung und anschließender Basenexzision von unmodifiziertem Cytosin ausgelöst wird.[315, 357] Ferner wurde in Abschnitt 3.6 und 4.1 gezeigt, dass Tet-Enzyme auch dT zu 5hmdU[13] und weiter zu 5fdU oxidieren, welche Substrate der Glykosylase Smug1 sind.[13, 312] Des Weiteren kennt

---

[VI] Die enzymatische Desaminierung von 5hmdC kann nach den Ergebnissen in Kapitel 3.6 und den Ergebnissen von *Nabel et al.* und *Rangam et al.* vernachlässigt werden.[358]

man bifunktionelle Glykosylasen, die eine Lyase-Aktivität aufweisen und nach der Deglykosylierung einen β-Eliminierungsschritt katalysieren. Dies führt zu einem DNA-Einzelstrangbruch mit einer α/β-ungesättigten AP-Stelle (uAP) am neu entstandenen 3′-Ende (**Schema 15**).[320, 362] Solche bifunktionellen Glykosylasen, wie die Neil-Enzyme, wurden als spezifische Binder von 5hmdC, 5fdC und 5cadC gefunden (siehe Abschnitt 3.5).[14] Kürzlich konnte gezeigt werden, dass die Neil-Enzyme 1–3 den Verlust von Tdg bei der Tet-vermittelten Reaktivierung eines stillgelegten Reportergens zum Teil kompensieren können.[346]
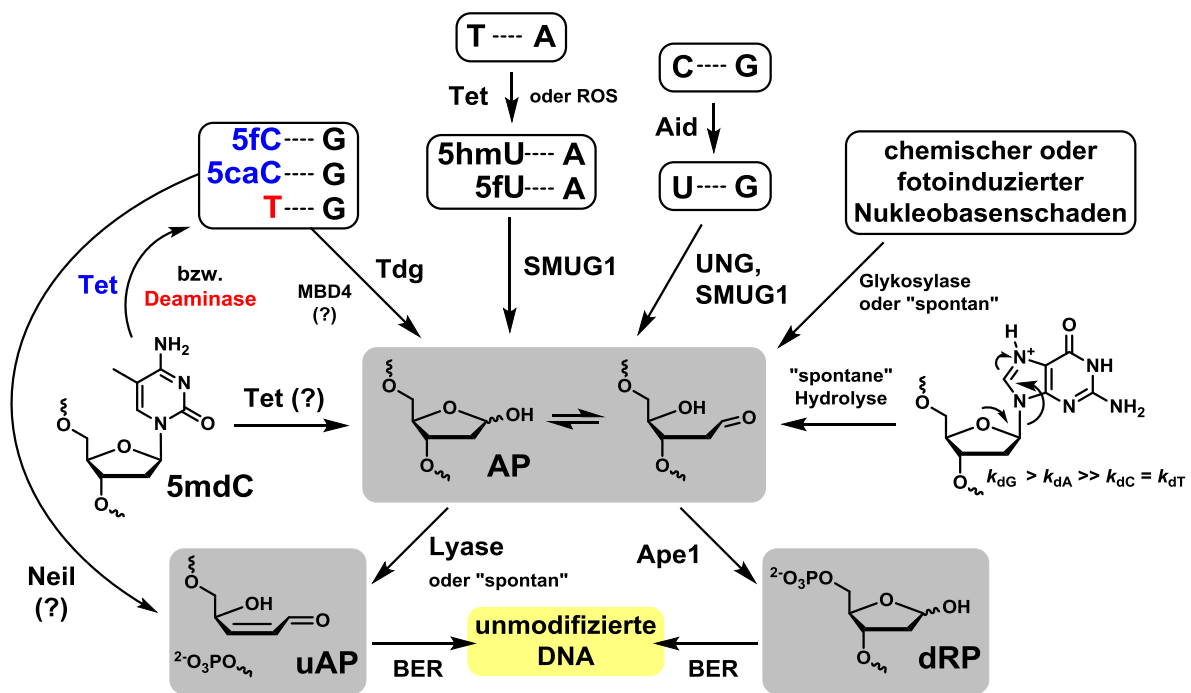
Schema 15 | Entstehungswege von abasischen Stellen durch den epigenetischen DNA-Metabolismus, durch die Reparatur von DNA-Schäden und Basenfehlpaarungen sowie durch die spontane Hydrolyse von intakter oder geschädigter DNA. Abasische Stellen (AP) liegen im Gleichgewicht zwischen ringgeschlossener und offenkettiger Form vor. Lyasen β-eliminieren 3′-ständiges Phosphat wodurch eine α/β-ungesättigte AP-Stelle (uAP) entsteht. Die AP-Endonuklease 1 (Ape1) hydrolysiert AP-Stellen in 5′-Position, wobei ein freies 3′-Hydroxyl- und ein 5′-Desoxyribosephosphat-Ende (dRP) gebildet werden. Typischerweise erfolgt die Reparatur von AP- und uAP-Stellen, die durch spontane Hydrolyse oder eine Glykosylase entstehen, per *Single*-Nukleotid- bzw. *short-patch*-BER.[363]

Neben diesen klassischen Glykosylase-basierten Entstehungswegen könnten AP-Stellen auch durch eine Nebenreaktion der Tet-Enzyme entstehen. Bei der Oxidation von 5mdC, dT und deren Derivate könnte dies, wie in **Schema 16** skizziert, durch einen Zerfall des radikalischen Intermediates im aktiven Zentrum der Tet-Enzyme geschehen.
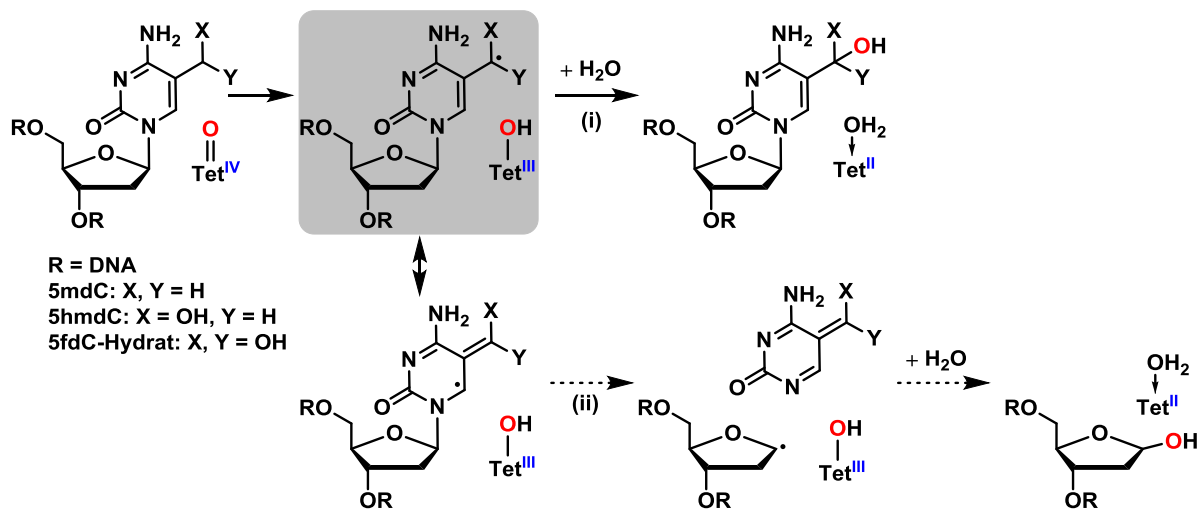
Schema 16 | Möglicher Entstehungsweg von abasischen Stellen bei der Tet-vermittelten Oxidation von 5mdC, 5hmdC oder 5fdC. Das Cytosin-Radikal (grau markiert), welches während des regulären Katalysezyklus entsteht, könnte neben der Oxidation (**i**) auch deglykosylieren (**ii**). Analoge Nebenreaktionen könnten bei der Oxidation von dT, 5hmdU oder 5fdU auftreten. Tet^{IV}=O symbolisiert das Fe(IV)-Oxo-Intermediat von Tet-Enzymen (siehe Schema 8 in Abschnitt 1.4.1).

AP- und uAP-Stellen können somit vielfach bei der epigenetischen Modifizierung von genomischer DNA auftreten. Drei zentrale Fragen ergeben sich hierbei: Erstens, in welchen Mengen treten die AP- und uAP-Stellen im Genom während der frühen Entwicklung z.B. in mES-Zellen überhaupt auf? Zweitens, was sind die Hauptentstehungswege und welchen Anteil haben diejenigen epigenetischen Ursprungs an der Gesamtmenge aller AP-Stellen? Besteht hierbei ferner eine Korrelation zwischen der Bildung von 5fdC, 5cadC und AP-Stellen? Wenn ja, könnte eine Bestimmung der Mengen dieser AP-Stellen einen Rückschluss auf die Effizienz bzw. den Beteiligungsgrad der BER geben.

Um diese Fragen beantworten zu können, wird im Nachfolgenden die Entwicklung eines neuen vielseitigen Derivatisierungsreagenzes präsentiert, das eine exakte Quantifizierung als auch eine Isotopenverfolgung von AP-Stellen im Genom per LC-MS/MS-*Assay* ermöglichen soll.

4.3.2 Ergebnisse und Diskussion

Design eines neuen Reagenzes zur Quantifizierung abasischer Stellen

Eine große Herausforderung bei der Quantifizierung von AP-Stellen besteht darin, diese mit einem geeigneten Reagenz zu derivatisieren, da sie nur so durch spurenanalytische (radioaktive,[364-365] immunologische oder ELISA-ähnliche[366-372] sowie massenspektrometrische[365, 373]) Methoden detektierbar werden. Die Schwierigkeit begründet sich zum einen durch die Labilität der glykosidischen Bindung von Nukleotiden, da spontane Deglykosy-

lierungen[VII] zu den häufigsten DNA-Schäden zählen, [362, 367, 369, 374, 378-380] und zum anderen durch die Labilität von AP-Stellen[362, 374, 381-384] die durch β- und δ-Eliminierung zu Strangbrüchen neigen (**Schema 15**). Diese Prozesse können sich vielfach bei der DNA-Isolation aus der jeweiligen Zellkultur- oder Gewebeprobe und bei der anschließenden Reaktion mit einem Derivatisierungsreagenz ereignen und führen zu verfälschten Messwerten.[371, 380, 385]

Als Reagenzien für die AP-Detektion haben sich Hydroxylamine durchgesetzt,[364, 366-367, 369, 371-373] da die resultierenden Oxime im wässrigen Milieu im Vergleich zu Iminen oder Semicarbazonen stabiler sind.[386-387] Etabliert hat sich bisher ein Biotinhydroxylamin, die sogenannte *Aldehyde Reactive Probe* (ARP, **Schema 17**),[367-368] welches als Kit für einen ELISA-ähnlichen *Assay* kommerziell verkauft wird (*Dojindo Molec. Tech.*) und Detektionslimits von ~2×10^{-7} AP pro dN bei einer Analyse von 3 µg DNA erreicht.[369] Diese Methode hat allerdings mehrere Nachteile, wenn es um die exakte Identifizierung und Quantifizierung von abasischen Stellen geht:

Die Reaktivität und Selektivität der ARP-basierten ELISA-Methode stellen die Zuverlässigkeit der publizierten Messergebnisse in Frage:[368, 385] Die genomische DNA weist unterschiedliche abasische Stellen (AP und uAP sowie diverse oxidierte Spezies) als auch andere Aldehydmodifikationen von Nukleosiden (u.a. 5fdC und 5fdU) auf, die mit ARP unterschiedlich reagieren.[6, 13, 362, 365, 368, 370, 385, 388-390] Ferner können Hydroxylamine auch mit kanonischen Nukleobasen reagieren[368, 391] und zusätzlich können nicht kovalent-gebundene Reagenzmoleküle das ELISA-Ergebnis verfälschen. Aus diesen Gründen ist zur zweifelsfreien Unterscheidung dieses komplexen Gemisches lediglich die kombinierte Anwendung der Umkehrphasenflüssigkeitschromatographie und Tandemmassenspektrometrie (LC-MS/MS) die spurenanalytische Methode der Wahl.[373] *Roberts et al.* entwickelten in diesem Sinne die erste LC-MS/MS-Methode mit dem Reagenz Nitrobenzylhydroxylamin (NBHA, **Schema 17**). Sie erreichten ein Quantifizierungslimit von 3×10^{-7} AP pro dN (100 fmol) bei einem Einsatz von 100 µg DNA pro Analyse.[373]

Dieser hohe DNA-Mengenbedarf, der für biochemische Analysen im Hochdurchsatz ungeeignet ist, sowie mangelnde Reaktivitäten und Selektivitäten zeigen einen Bedarf für die Entwicklung einer verbesserten universalen Methode auf.[364, 366-371, 373, 385, 389] Diese Methode

---

[VII] Die säurekatalysierte Zerfallsreaktion tritt unter physiologischen Bedingungen in einer Zelle etwa 10000 bis 50000mal pro Tag auf und stellt deshalb den häufigsten DNA-Schaden dar, wobei die Purine etwa 20mal häufiger deglykosylieren als die Pyrimidine (**Schema 15**).[374]

soll die Vorzüge einer schnellen und milden Derivatisierungsreaktion mit einer empfindlichen und präzisen LC-MS/MS-basierten Quantifizierungsmethode verbinden.
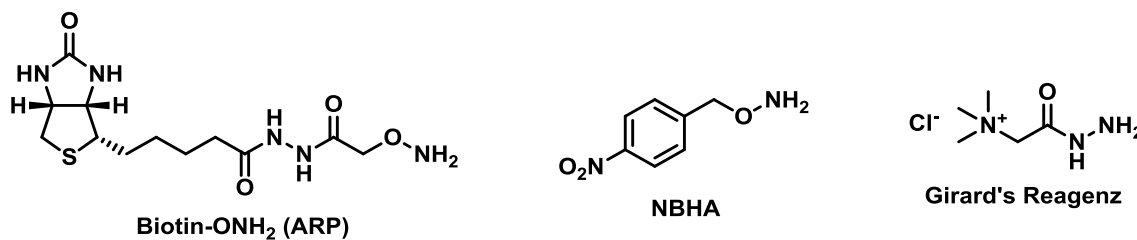
Schema 17 | Reagenzien für die Derivatiserung von Aldehyden in der DNA.

Aus chemischer Sicht soll das Reagenz insbesondere eine hohe Reaktionsgeschwindigkeit gegenüber AP-Stellen aufweisen, auch wenn die *per se* säurekatalysierte Oximligations-reaktion im physiologischen pH-Bereich erfolgen soll, um die Generierung von artifiziellen AP-Stellen zu minimieren. *Kojima et al.* beschäftigten sich erstmals mit einer Struktur-Funktions-basierten Optimierung des ARP-Reagenzes für ELISA-ähnliche *Assays*. Das reaktivste Hydroxylamin-Reagenz war in ihrer Studie aoN-g (Aminooxy-naphthalen-guanidin; **Schema 18A**), dessen Reaktionsrate bei pH 7.5 um eine Größenordnung schneller war. Ihr Reagenzdesign stützte sich auf den Annahmen, über aromatische *Stacking*-Effekte mit flankierenden Nukleobasenpaaren und ionische Wechselwirkungen mit dem negativ geladenen Zucker-Phosphat-Rückgrat eine höhere Affinität und damit eine höhere Reaktivität mit der DNA zu erzielen.[372] Typische Moleküle wie DB1804 (**Schema 18A**) oder Oligoamide, die sich durch extrem hohe DNA-Bindungsaffinitäten auszeichnen, besitzen im Gegensatz zu aoN-g eine länglichere bzw. leicht gebogene Struktur, die sich in die konkave Krümmung der kleinen Furche passgenau einbetten kann.[392] Hydroxylamin-Reagenzien mit solchen Strukturen könnten also die Reaktionsgeschwindigkeit der Oximligation weiter erhöhen.

Aus analytischer Sicht existieren daneben viele Arbeiten, in denen das Prinzip der chemischen Derivatisierung eines Analyten dazu verwendet wird, die Signalintensität und die chromatographische Trennung zu verbessern. Reagenzien, die für die LC-ESI-MS/MS-Analytik gezielt konstruiert wurden, zeichnen sich zudem durch Strukturen aus, die durch Kollisions-induzierte Dissoziation (*collision induced dissociation*; CID) einfach fragmentiert werden können und intensive Tochterionen ergeben.[393] Funktionelle Gruppen, die sich bewährten, sind u.a. Ester, Amide, N–O- sowie N–N-Bindungen und quaternäre Alkylamine.[393] Für die Derivatisierung von Aldehyden wurden so zum Beispiel Reagenzien mit permanent geladenen Gruppen wie *Girard*'s Reagenz T verwendet, mit dem die Detektion

von 5fdU im unteren fmol-Bereich 20fach verbessert werden konnte (**Schema 17**).[394] Dieses kommerzielle Reagenz eignet sich aufgrund seiner Hydrophilie und der oben erwähnten Labilität von Semicarbazonen allerdings nicht für die Umkehrphasen-Analytik von AP-Stellen.
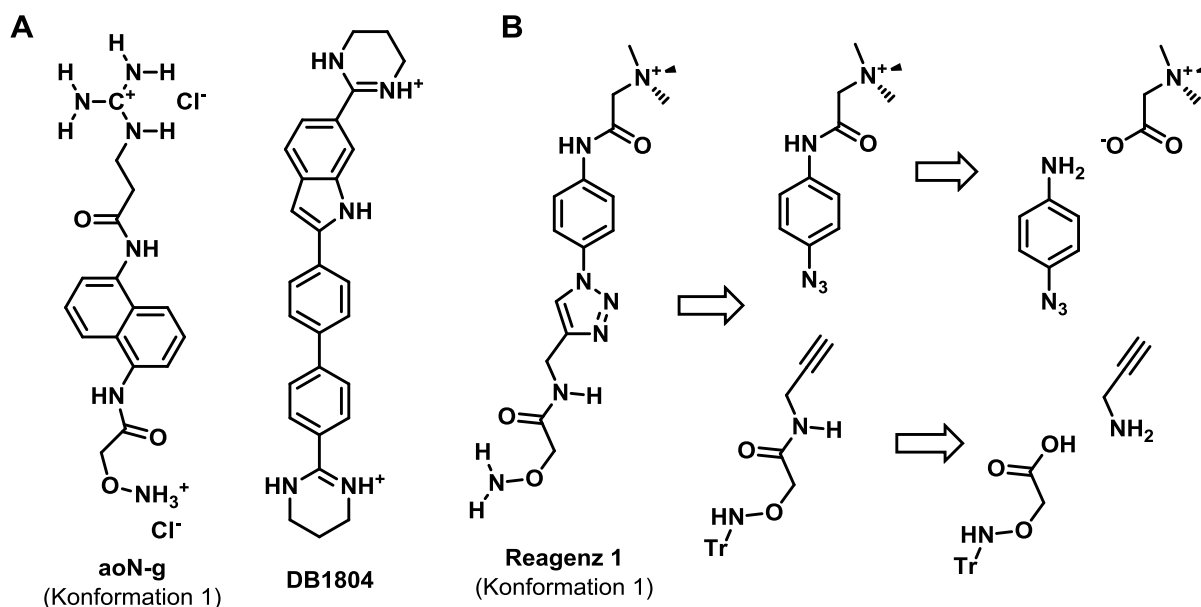
Schema 18 | (**A**) Verbessertes Hydroxylaminreagenz aoN-g zur Derivatisierung von AP-Stellen von *Kojima et al.*[372] und synthetischer DNA-Ligand DB1804.[392] (**B**) Design und Retrosynthese eines neuen Reagenzes für die LC-MS/MS-Detektion von AP-Stellen. Gegenionen sind vereinfachend nicht dargestellt. Weitere mögliche Konformationen von aoN-g und Reagenz **1** sind in Abbildung 18B abgebildet.

Unter Berücksichtigung dieser chemischen und analytischen Aspekte wurde Hydroxylamin **1** entworfen (**Schema 18B**). Es enthält wie *Girard*'s Reagenz eine Betain-Einheit, die aus einem permanent geladenen quaternären Alkylamin besteht. Dieses soll einerseits zu höheren Signalintensitäten in der LC-ESI-MS/MS-Analytik führen und andererseits die Affinität des Reagenzes zum negativ-geladenen Phosphat-Rückgrat der DNA erhöhen. Ferner ist Betain als Isotopen-markierte Variante (D_{11}) käuflich erwerblich und kann synthetisch durch simple Amidkupplung mit dem Reagenz verknüpft werden, um interne Standards für die MS-basierte Quantifizierung von abasischen Stellen zu erhalten.

Die Struktur des entworfenen Reagenzes besteht ferner aus einer Phenyl-Triazol-Kernstruktur. Diese soll zu einer günstigen Retentionszeit bei der Umkehrphasen-Chromatographie führen, um eine simultane Quantifizierung von epigenetisch relevanten DNA-Modifikationen in nur einem analytischen Lauf durchführen zu können. Ferner nimmt das Reagenz durch die Phenyl-Triazol-Kernstruktur eine länglichere Struktur als aoN-g an

und ähnelt daher eher der Struktur des DNA-Liganden DB1804 (**Schema 18A**), welcher eine extrem starke Affinität zur kleinen Furche des DNA-Duplex hat.[392] Dies illustriert das molekulare Docking in **Abbildung 18A**.
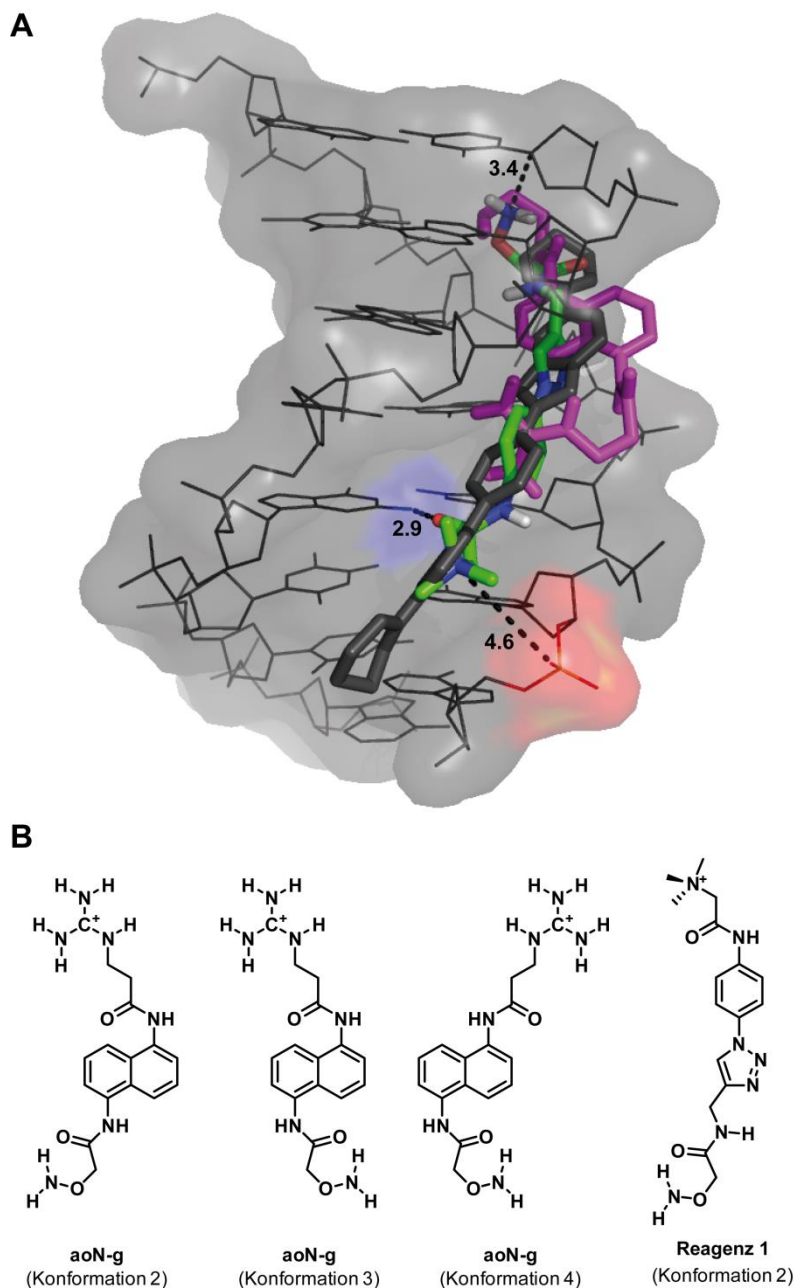
Abbildung 18 | Molekulares Docking zwischen Reagenz und DNA. (**A**) Repräsentatives Docking-Ergebnis mit der DNA-Duplex-Kristallstruktur von *Narayana* und *Weiss* (PDB code: 3BSE).[395] Als Rezeptoren dienten DB1804[392] (grau, Konformation aus Kristallstruktur PDB 3U05), aoN-g[372] (magenta, Konformation 1) und Reagenz **1** (grün, Konformation 1) (siehe Schema 18). Ausgewählte atomare Abstände in Å sind schwarz gepunktet. Das molekulare Docking wurde mit Hilfe von AutoDock Vina durchgeführt.[396] (**B**) Weitere Konformationen von aoN-g und Reagenz **1**; siehe hierzu auch Schema 18.

Wie DB1804 bettet sich das entworfene Reagenz **1** in die konkave Krümmung der kleinen Furche ein. aoN-g nahm eine solche Struktur beim Modelling dagegen nicht an, auch wenn für die Berechnungen unterschiedliche Konformationen des Moleküls eingesetzt wurden (**Abbildung 18B**). Im Modell des molekularen Dockings ist darüber hinaus das permanent geladene Alkylamin von Reagenz **1** zwischen zwei gegenüberliegenden Phosphatgruppen in gutem Abstand positioniert und das Hydroxylamin befindet sich in direkter Nähe zu dem acetalischen C1′-Atom einer potenziellen AP-Stelle (3.4 Å; **Abbildung 18A**).

Zusammenfassend sind in dem entworfenen Reagenz somit mehrere Strukturelemente kombiniert, um chemische und analytische Bedingungen zu optimieren. Ferner besitzt es für die Kollisions-induzierte Dissoziation im Massenspektrometer günstige Fragmentierungs-stellen (siehe oben), um eine hohe Ausbeute von intensiven Tochterionen zu erhalten.[393] Insbesondere können Triazole unter Verlust von molekularem Stickstoff fragmentiert werden.[397]

Synthese der Reagenzien und massenspektrometrischer Referenzverbindungen

Neben der Synthese des Hydroxylamin-Reagenzes **1** erfolgte die Synthese des entsprechenden Oximligationsproduktes von AP-Stellen (**2**) sowie der [D9]-markierten Derivate als Standards für die quantitative LC-MS/MS-Analytik (**Schema 19**).

Die Synthese begann mit der Amidkupplung von *para*-Azidoanilin (**3**) und Betain (**4a**) bzw. [D11]-Betain (**4b**) mittels TBTU als Kupplungsreagenz. Das Ionenpaar aus dem resultierenden Amid und dem Anion von 1-Hydroxybenzotriazol konnte als Zwischenprodukt per Normalphasen-Chromatographie gereinigt werden. Nach HCl-saurer Extraktion konnte das Azid **5a/b** als Chlorid-Salz mit einer Ausbeute von 90% erhalten werden. Unter den Synthese-bedingungen vollzog sich ein vollständiger D/H-Austausch der Methylen-Gruppe, so dass Azid **5b** [D9]-markiert erhalten wurde. Der Schlüsselschritt in der Syntheseroute war der Aufbau der Phenyl-Triazol-Kernstruktur per Cu(I)-katalysierter Cycloadditionsreaktion von Azid **5** mit Alkin **6**, die als "Klick"-Chemie weit verbreitet ist.[398-402] Alkin **6** wurde hierfür durch TBTU-vermittelte Amidkupplung von Propargylamin und *N*-Trityl-geschützter Aminooxy-Essigsäure **7** in 69% Ausbeute erhalten. Die anschließende Klick-Reaktion mit Azid **5** lieferte Triazol **8** mit einer Ausbeute von 77%. Das Produkt ließ sich dank der hydrophoben Trityl-Schutzgruppe trotz der quaternären Alkylamin-Gruppe erneut per Normalphasen-Chromatographie reinigen. Die Entschützung von Triazol **8** erfolgte in einer HCl-sauren Emulsion aus H2O und DCM. Das Ende der Reaktion wurde durch eine

vollständige Phasentrennung sowie Aufklärung der Emulsion angezeigt und massen-spektrometrisch geprüft.
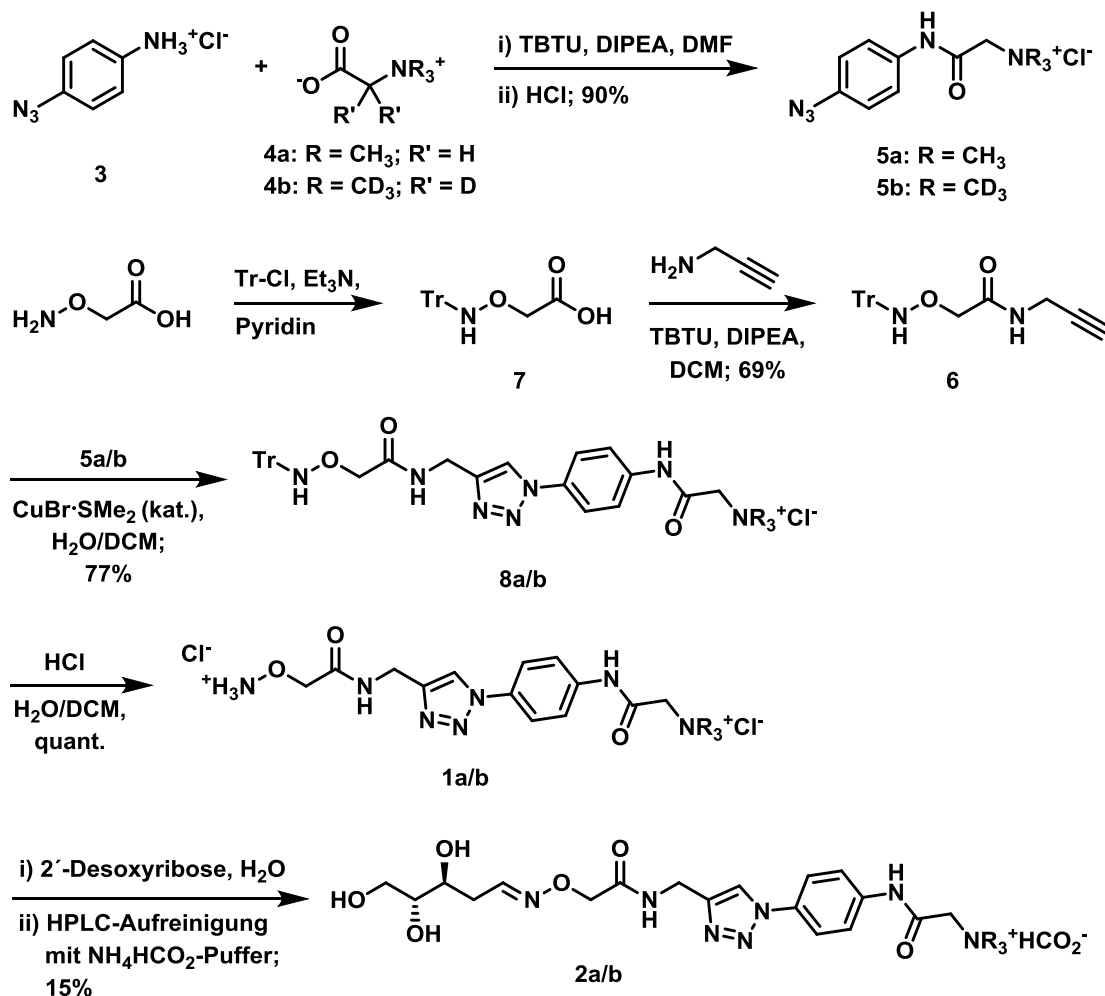
Schema 19 | Synthese des Hydroxylamin-Reagenzes und des [D₉]-markierten Derivates (1a/b) sowie der entsprechenden Oximligationsprodukte mit 2′-Desoxyribose (2a/b) als LC-MS/MS-Referenzverbindungen (Standards).

Die Reinigung gelang durch organische Extraktion der Tritylgruppe. Das Hydroxylamin-Reagenz **1** konnte in quantitativer Ausbeute mit einer Reinheit von 95% erhalten werden (**Abbildung 19A**). Reagenz **1** kondensierte mit 2′-Desoxyribose in einer wässrigen Lösung zu Oxim **2**. Dieses ließ sich nach zweimaliger Reinigung per Umkehrphasen-HPLC mit einer Ausbeute von 15% gewinnen. Nach UHPLC-UV-Analyse betrug die Reinheit 99.4%. Die Verbindung lag in wässriger Lösung als Gemisch von *E*- und *Z*-Isomeren vor. In **Abbildung 19B** ist das entsprechende UHPLC-UV- und -MS/MS-Chromatogramm des *E/Z*-Oxim-Gemisches dargestellt.
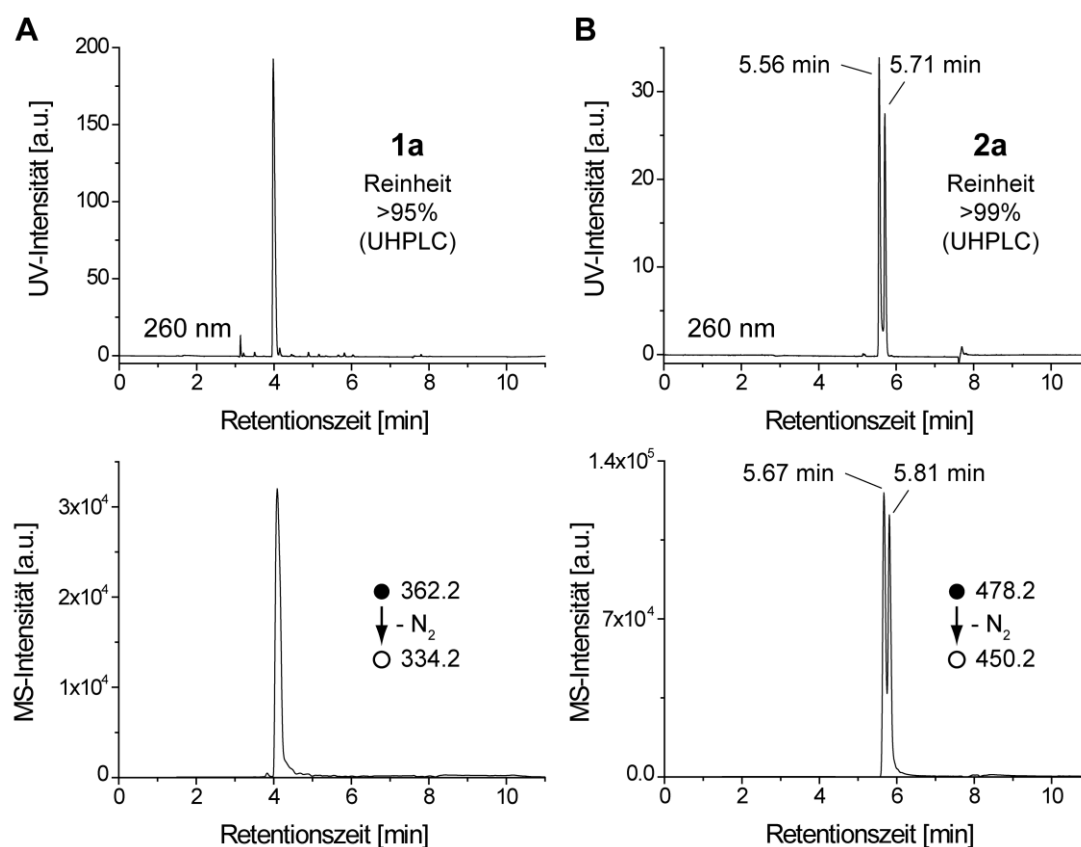
Abbildung 19 | Reinheitsanalyse des Hydroxylaminreagenzes 1a (A) und des Oximligations-produkts 2a (B) per UHPLC-UV-ESI-MS/MS. Dargestellt sind die UV-Chromatogramme (oben) bei 260 nm und die entsprechenden MS/MS-Chromatogramme (unten). Zwischen UV- und MS-Detektor besteht eine Zeitdifferenz von 0.1 min. Die Chromatographie wurde mit einer C_8-Säule durchgeführt. Für eine Beschreibung der Fragmentionen siehe Abbildung 22.

Methodenentwicklung zur Detektion und Quantifizierung abasischer Stellen

Für die Entwicklung der Quantifizierungsmethode von AP-Stellen wurde die Oximligations-reaktion von Reagenz **1** zunächst mit einem synthetischen doppelsträngigen Oligodesoxy-nukleotid charakterisiert. Dies geschah hauptsächlich in Zusammenarbeit mit *Victor Brantl* im Zuge seiner Masterarbeit.[403] Für die Untersuchung wurde eine definierte AP-Stelle per enzymatischer Deglykosylierung eines Uracil-enthaltenden Doppelstrangs durch die Uracil-DNA-Glykosylase (UDG) erhalten (**Abbildung 20A**).[362, 372]

Anschließend wurde die AP-enthaltende DNA mit einem 100fachen Reagenzüberschuss in physiologisch gepufferter Lösung inkubiert.[368, 371-372, 385] Nach 60 min bei 37 °C wurde die Reaktion mit Aceton gestoppt. Es wurde ein vollständiger Umsatz der AP-Stelle per Kapillarzonenelektrophorese erhalten (**Abbildung 20B**). Dabei wurde keine Bildung von Nebenprodukten wie Abbruchfragmente durch *β*-Eliminierung festgestellt.[385] Die MALDI-TOF-Analyse des Reaktionsgemisches bestätigte die Bildung des gewünschten Oxim-ligationsproduktes mit der Detektion der entsprechenden Molekülmasse von

m/z 4261 (berechnet: *m/z* 4262; **Abbildung 20C**). Kinetische Untersuchungen ergaben, dass die Reaktion unter diesen Bedingungen bereits nach 4 min abgeschlossen war ($k = 1.2$ min^{-1}; **Abbildung 21**). Die Reaktionsgeschwindigkeit des Reagenzes halbierte sich in etwa, wenn die Reaktion stattdessen an einzelsträngiger DNA durchgeführt wurde (siehe Masterarbeit *René Rahimoff*).[404] Dies demonstriert die Abhängigkeit der Reaktionsgeschwindigkeit von der Affinität des Reagenzes zur DNA.[372]
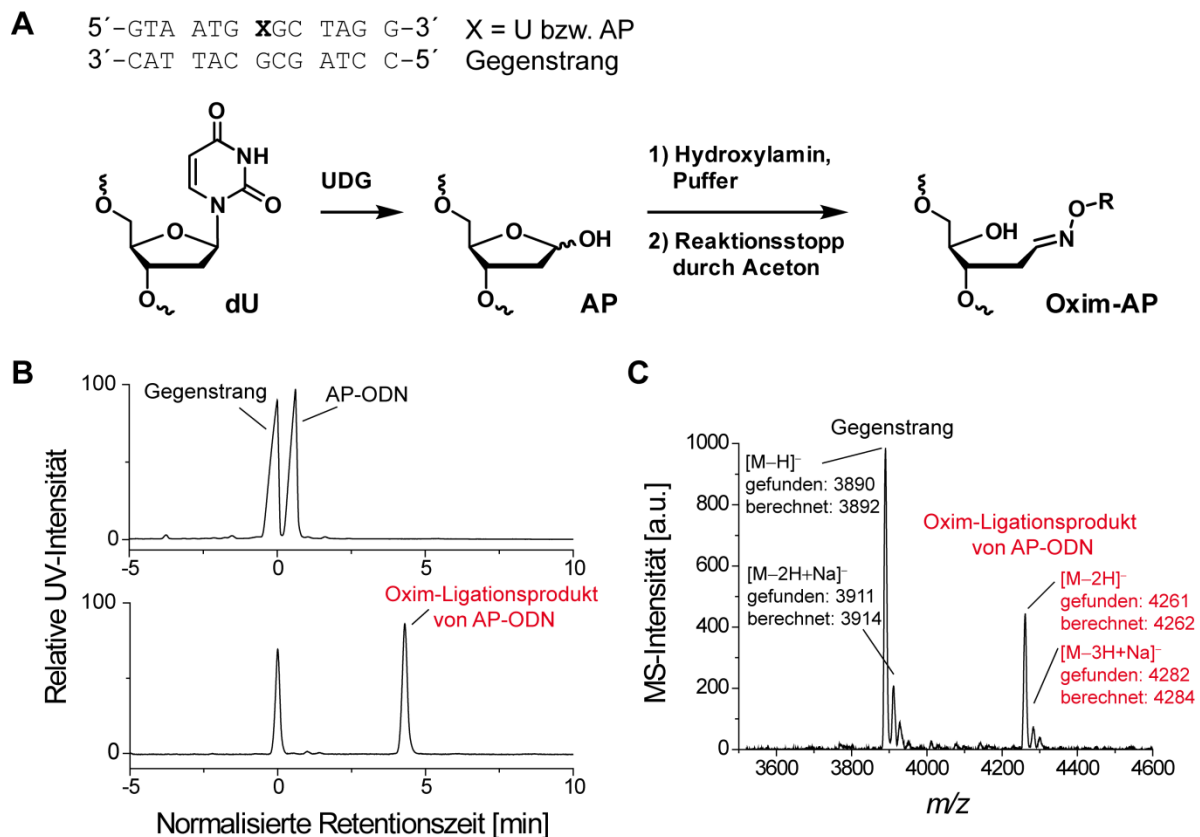
Abbildung 20 | Reaktion von Hydroxylamin 1a mit doppelsträngiger synthetischer DNA, die eine definierte AP-Stelle enthält. (**A**) Schematische Darstellung zur Herstellung von AP-DNA und anschließender Derviatisierung per Oximligation. (**B**) Kapillarzonengelelektrophorese (254 nm) von doppelsträngiger DNA nach UDG-Behandlung (oben) und nach der Derivatisierungsreaktion (unten; 1.5 mM Reagenz, PBS-Puffer pH 7.4, 37 °C, 60 min). ODN = Oligodesoxynukleotid (**C**) MALDI-TOF-Spektrum nach der Derivatisierungsreaktion dargestellt in B.

Nachdem die gute Derivatisierbarkeit von abasischen Stellen in synthetischer DNA mit dem neuen Reagenz gezeigt werden konnte, erfolgte im nächsten Schritt die Implementierung der Detektion der derivatisierten AP-Stelle in die bestehende UHPLC-ESI-MS/MS-Methode (siehe Abschnitt 3.6). Diese Methodenerweiterung sollte die simultane Detektion bzw. Quantifizierung der derivatisierten AP-Stellen und der epigenetischen DNA-Modifikationen ermöglichen. Nach enzymatischer Hydrolyse der DNA-Probe durch Nukleasen, Phosphatasen

und Phosphodiesterasen, soll neben den resultierenden, freien Nukleosiden die derivatisierte AP-Stelle als Oximligationsprodukt **2a** für die LC-MS/MS-Analytik vorliegen.
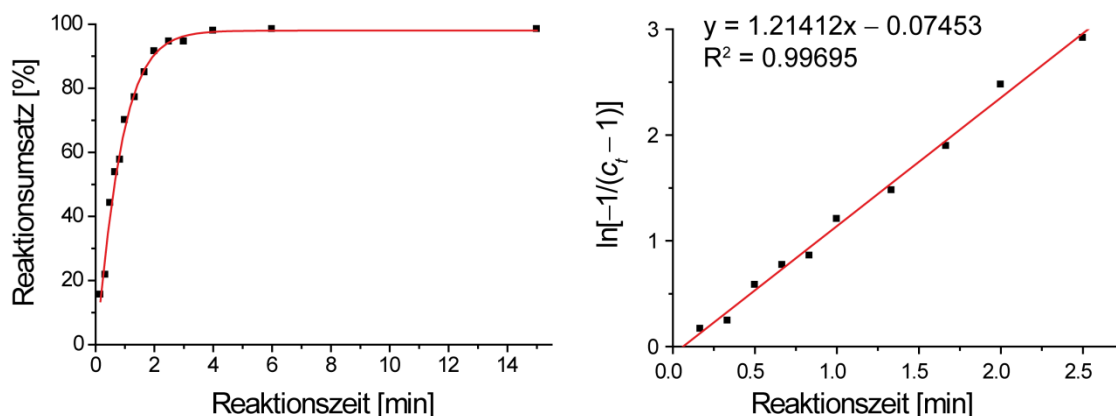
Abbildung 21 | Reaktionskinetik der Derivatisierungsreaktion ist pseudo-erster Ordnung. Die Geschwindigkeitskonstante (*k*) entspricht der Steigung der linearen Regression nach logarithmischer Auftragung des Reaktionsumsatzes gegen die Zeit (ln[−1/(c_t− 1)]; c_t = Reaktionsumsatz zur Zeit *t*).

Für die LC-MS/MS-Detektion dieses Analyten wurde zunächst das Fragmentierungsmuster des synthetisch hergestellten Oximligationsproduktes **2a** in einem Triplequadrupol-Massenspektrometer bestimmt. Die direkte Injektion einer wässrigen Lösung von **2a** lieferte als intensivstes Mutterion das erwartete Molekülkation mit der monoisotopischen Masse *m/z* 478.1 (berechnet: *m/z* 478.2). **Abbildung 22A** zeigt das Produktionen-Spektrum der Kollisions-induzierten Dissoziation des Mutterions. Bei mittlerer Kollisionsenergie wurden zahlreiche Tochterionen erhalten. Die sieben intensivsten MS-Übergänge wurden anschließend für eine Optimierung der Kollisionsparameter verwendet (**Abbildung 22B**). Der intensivste MS-Übergang war *m/z* 478/450, der mit dem Verlust eines Neutralteilchens von 28 amu auf die Fragmentierung des Triazols unter Freisetzung von molekularem Stickstoff zurückzuführen ist (**Schema 20**).[397] Diese Fragmentierung ereignete sich interessanterweise im Vergleich zu den anderen funktionellen Gruppen mit der geringsten Kollisionsenergie (**Abbildung 22B**),[393] weswegen die Triazol-Einheit als zusätzliches Motiv bei der Konstruktion von Reagenzien für die LC-ESI-MS/MS-Analytik in Betracht gezogen werden könnte.[397] Ein weiterer, etwas weniger intensiver MS-Übergang wurde bei höheren Kollisionsenergien mit *m/z* 478/190 erhalten. Dieser Übergang diente im Nachfolgenden als Qualifizierer, um die Selektivität der Methode bei der Detektion der derivatisierten AP-Stellen zu gewährleisten.
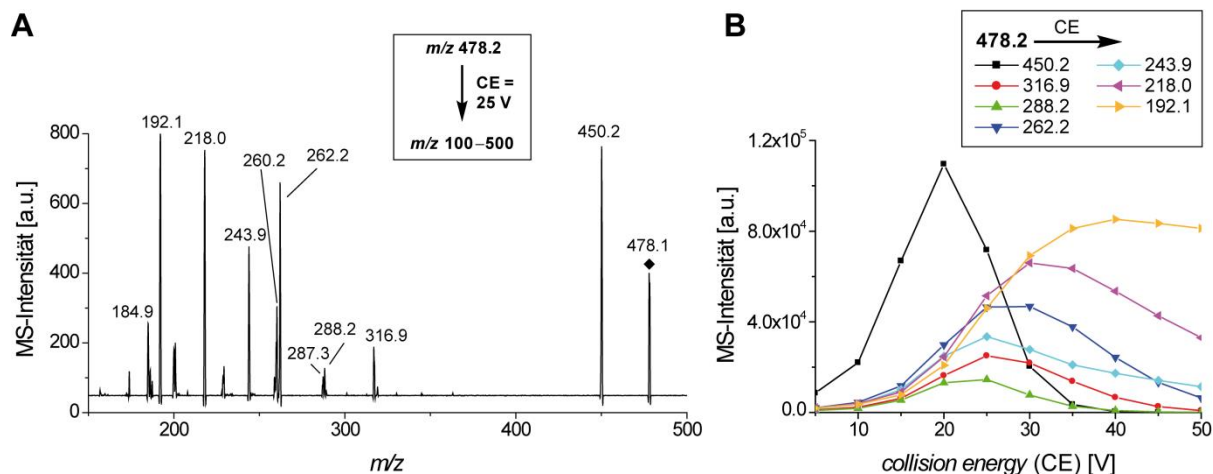
Abbildung 22 | MS/MS-Charakterisierung und Optimierung der Kollisionsparameter von Oximligationsprodukt 2a. (**A**) Produktionen-Spektrum von **2a** bei einer Kollisionsenergie (*collision energy*; CE) von 25 V im positiven Ionenmodus. Das MS-Signal von unfragmentierten Mutterionen ist mit einer Raute symbolisiert. (**B**) Optimierung der CE für ausgewählte Fragmentionen. Dargestellt ist die Signalintensität in Abhängigkeit der CE.

berechnet: *m/z* 478.2

$- N_2$

berechnet: *m/z* 450.2

berechnet: *m/z* 317.2

berechnet: *m/z* 288.2

berechnet: *m/z* 260.2

berechnet: *m/z* 192.1
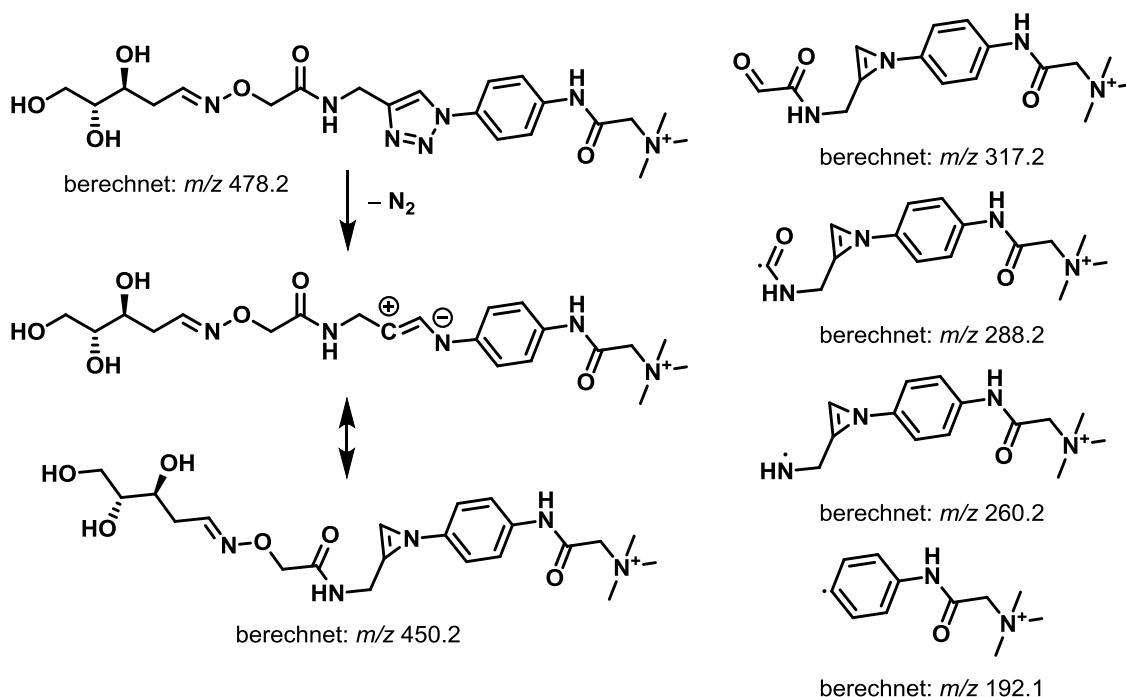
Schema 20 | Mögliches Fragmentierungsmuster bei der Kollisions-induzierten Dissoziation von Oximligationsprodukt 2a.

Nach der Optimierung der Fragmentierungsparameter erfolgte die Analyse der derivatisierten AP-Stellen in synthetischer DNA per LC-UV-ESI-MS/MS. Hierfür musste ein eigenes Reinigungsprotokoll entwickelt werden, da sich das überschüssige Reagenz von der DNA zunächst nicht abtrennen ließ und die Analyse stark beeinträchtigte. Die Reaktion des Reagenzes wurde, wie oben beschrieben, mit der Zugabe von Aceton beendet. Klassische Methoden, wie die ethanolische DNA-Fällung, waren trotz mehrmaliger Waschschritte nicht

geeignet, das resultierende Oximligationsprodukt abzutrennen. Sogar die Dialyse der DNA-Probe mit Hilfe eines semipermeablen Membranfilters (0.025 µm), die normalerweise für die Entsalzung von DNA-Proben verwendet wird, führte nicht zum Erfolg. Dies verdeutlicht die hohe Affinität des Reagenzes zur DNA im Vergleich zu den bereits publizierten Reagenzien (ARP, aoN-g), da sich letztere von der DNA durch mehrmaliges Waschen entfernen ließen.[368, 372]

Mit Hilfe eines aromatischen Aldehyds (1-Naphthylaldehyd, **Abbildung 23A**) konnte das Reagenz bzw. das entsprechende Oximligationsprodukt final abgetrennt werden. Der aromatische Aldehyd erhöhte die Hydrophobizität des Produktes, welches deshalb bei der DNA-Präzipitation mit *iso*-Propanol in Lösung blieb. Die Ausarbeitung des Protokolls zur Derivatisierung und anschließenden Reinigung von DNA erfolgte in Zusammenarbeit mit *Victor Brantl* im Zuge seiner Masterarbeit. Weitere Details zur Protokollentwicklung sind in seiner Masterarbeit nachzulesen.[403] Die Prozedur verlief zusammengefasst wie folgt: Die DNA-Proben wurden zur Beendigung der Derivatisierungsreaktion mit 1-Naphthylaldehyd (1.5 M) bei 37 °C für 10 min in einem *iso*-Propanol/Wasser-Gemisch inkubiert. Dabei war 1-Naphthylaldehyd bei 75 Volumenprozent *iso*-Propanol vollständig gelöst. Anschließend erfolgte die Fällung der DNA nach Zugabe von NaOAc und weiterem *iso*-Propanol durch Zentrifugation. Das DNA-Pellet wurde nach zwei Waschschritten mit *iso*-Propanol (75%) und Ethanol (75%) in Wasser gelöst und abschließend für die LC-UV-ESI-MS/MS-Analytik enzymatisch hydrolysiert. **Abbildung 23B** zeigt die UV-Spur der Umkehrphasen-Chromatographie des enzymatischen Hydrolysats der synthetischen DNA, deren AP-Stelle mit dem Reagenz umgesetzt und nach diesem Protokoll gereinigt wurde. Neben den kanonischen Nukleosiden (dC, dG, dT und dA) eluierte das Oximligationsprodukt der AP-Stelle mit einer höheren Retention trotz der permanent geladenen, quaternären Alkylamingruppe. Identisch zu dem synthetisch hergestellten Oximligationsprodukt **2a** (**Abbildung 19B**) eluierte die derivatisierte AP-Stelle per UV-Detektion als Doppelpeak (9.30 und 9.33 min), welcher sich durch Diffusions-abhängige Peak-Verbreiterung vereinigte und bei 9.42 min den MS-Detektor erreichte (**Abbildung 23C**). Die Identifizierung des Oximligationsproduktes erfolgte durch die charakteristischen MS-Übergänge *m/z* 478/450 und 478/192.
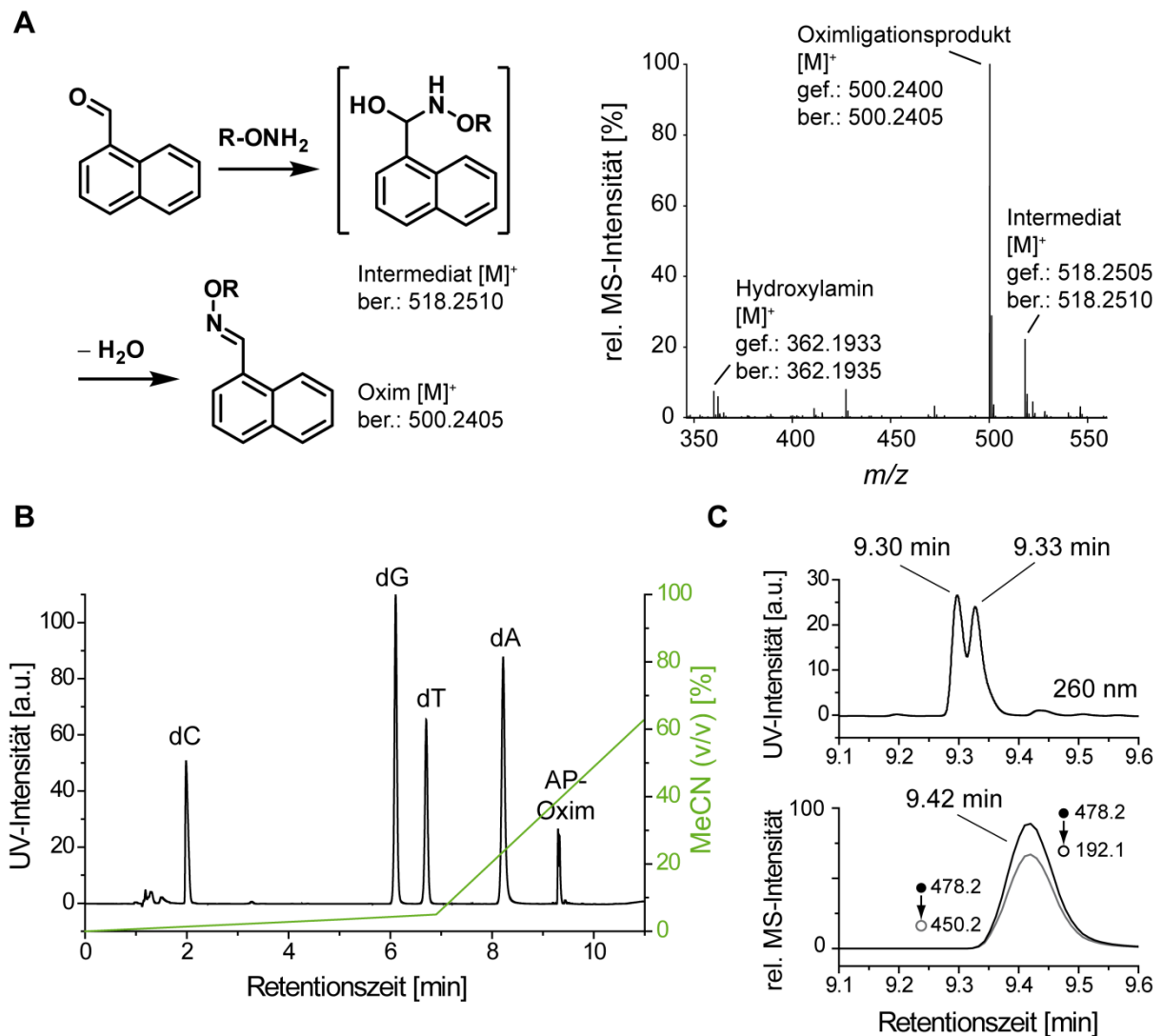
Abbildung 23 | Repräsentative UHPLC-UV-ESI-MS/MS-Analytik nach Derivatisierung der abasischen Stellen in synthetischer DNA durch das Reagenz 1a. (**A**) Beendingung der Reaktion zwischen dem Hydroxylamin-Reagenz und der DNA mit Hilfe von 1-Naphthylaldehyd; dargestellt ist ein Schema der Abfangreaktion überschüssigen Hydroxylamins mit zugehörigem hochauflösendem Massenspektrum (ESI-FT-ICR-Analytik). (**B+C**) UHPLC-UV-ESI-MS/MS-Analytik des enzymatischen Verdaus des AP-ODNs aus Abbildung 20 nach quantitativem Umsatz der AP-Stellen mit Reagenz **1a**. Die Abtrennung des überschüssigen Reagenzes erfolgte mit 1-Naphthylaldehyd. (**B**) UV-Chromatogramm des DNA-Verdaus bei 260 nm in schwarz; Gradient des MeCN-Anteils in grün. (**C**) Das resultierende Oxim der AP-Stellen eluiert mit ca. 40% (v/v) MeCN bei 9.30 bzw. 9.33 min per UV-Detektion (260 nm; oben) und erreicht bei 9.42 min den MS-Detektor. Die Identifizierung des Oximligationsproduktes erfolgte durch die MS/MS-Chromatogramme der charakteristischen Fragment-ionen (*m/z* 478/450: CE = 13 V; *m/z* 478/192: CE = 25 V).

Abasische Stellen in genomischer DNA von mES-Zellen

Mit dem oben beschriebenen Protokoll wurde nachfolgend die Detektion und Quantifizierung von AP-Stellen in genomischer DNA von mES-Zellen per LC-UV-ESI-MS/MS durchgeführt. Hierfür genügte die Verwendung von lediglich 5 µg DNA. Die Reaktion mit dem Hydroxylamin-Reagenz wurde nach 15 min mit 1-Naphthylaldehyd beendet. Diese

Reaktionszeit sollte aufgrund der zuvor ermittelten Reaktionskinetik des Oligodesoxynukleotids ausreichend sein (**Abbildung 21**). Das Reinigungsprotokoll lieferte hochreine DNA: Neben den kanonischen Nukleosiden sind im UV-Chromatogramm des enzymatischen DNA-Hydrolysats keine weiteren Substanzen, wie überschüssiges Reagenz, in größeren Mengen zu erkennen (**Abbildung 24A**). Die Identifizierung und Quantifizierung von AP-Stellen gelang mit der MS-basierten Isotopenverdünnungsmethode unter Verwendung des [D$_9$]-markierten Oximligationsproduktes **2b** als internen MS-Standard. Die AP-Stellen und der interne Standard eluierten mit identischer Retention und identischem Fragmentierungsmuster (**Abbildung 24B**,**C**).
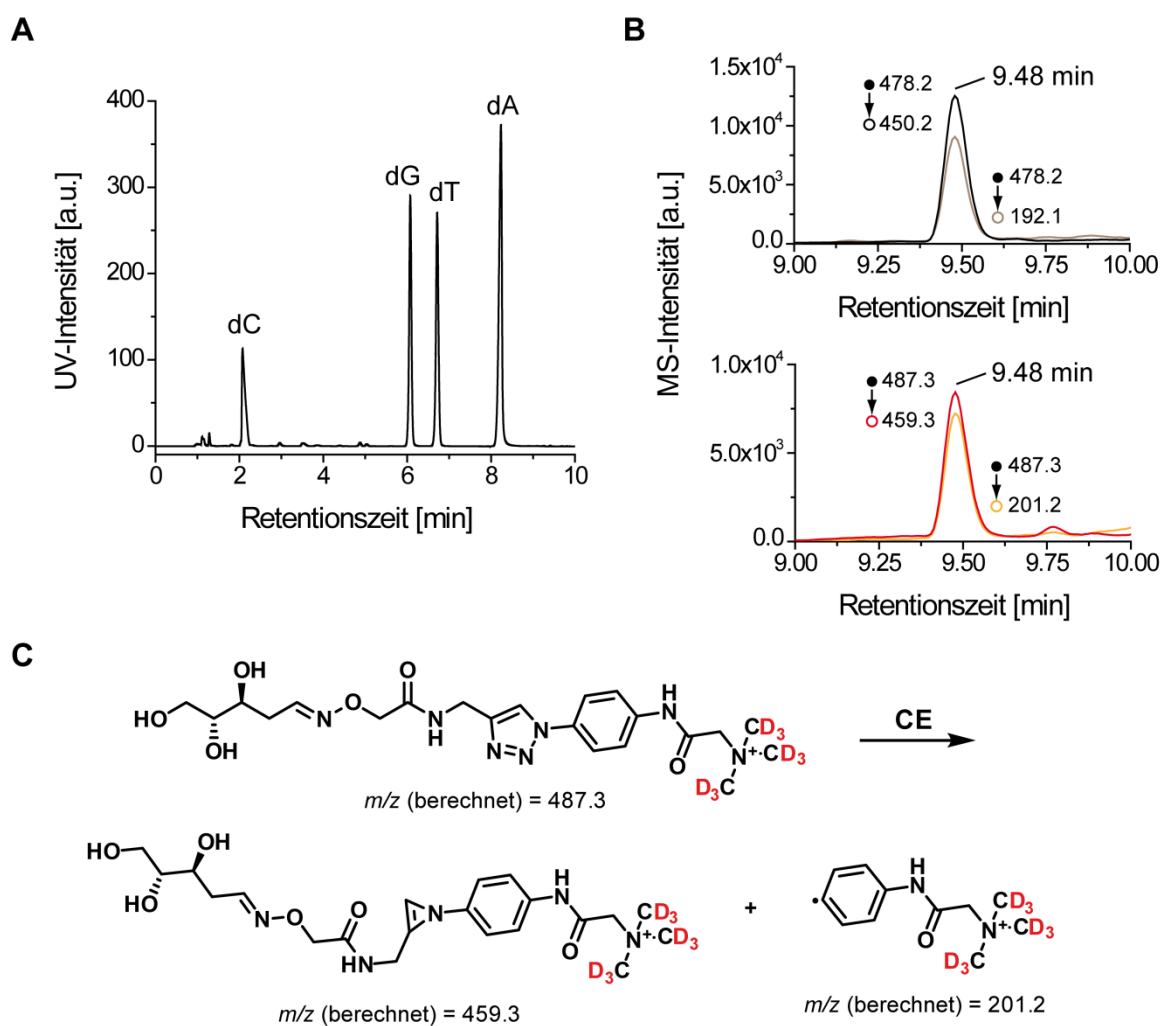
Abbildung 24 | Detektion und Quantifizierung von AP-Stellen in genomischer DNA von mES-Zellen per UHPLC-UV-ESI-MS/MS. (**A**) Repräsentatives UV-Chromatogramm (260 nm) nach Derivatisierung von abasischen Stellen in mES-DNA durch das Hydroxylamin-Reagenz **1a**, anschließender Reinigung mit dem Naphthylaldehyd-Protokoll und enzymatischer DNA-Hydrolyse. (**B**) Zugehörige MS/MS-Chromatogramme der derivatisierten AP-Stellen (oben) und des [D$_9$]-markierten, internen Standards (unten). *m/z* 478/450: CE = 19 V; *m/z* 478/192: CE = 40 V. (**C**) Entsprechendes Fragmentierungsschema des [D$_9$]-markierten, internen Standards (Oximligationsprodukt **2b**).

Der AP-Gehalt dieser repräsentativen DNA-Probe war innerhalb des linearen Quantifizierungsbereiches der etablierten LC-MS/MS-Methode und lag mit 27 fmol deutlich über dem Quantifizierungslimit von 2.1 fmol bzw. über dem Detektionslimit von 0.2 fmol (**Abbildung 25**). Diese Methode ist mindestens 50mal empfindlicher als die LC-MS/MS-Methode von *Roberts et al.* und benötigt 20mal weniger DNA.[373] Darüber hinaus ermöglichte die neue Methode die simultane Quantifizierung der DNA-Modifikationen 5mdC, 5hmdC, 5fdC sowie 5hmdU, da sie sich in die bestehende LC-MS/MS-Methode aus Abschnitt 3.6 integrieren ließ. Während 5fdC unter den Reaktionsbedingungen kaum mit dem Hydroxylamin-Reagenz reagierte, wies 5fdU eine höhere Reaktivität auf, wodurch 5fdU nicht co-quantifiziert werden kann.[403]
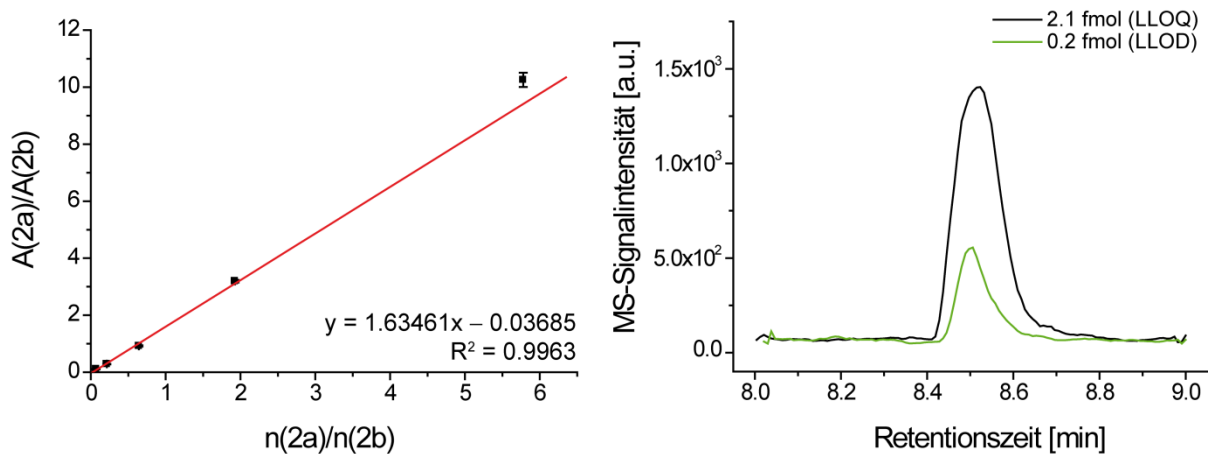
Abbildung 25 | Eichgerade und repräsentative MS/MS-Chromatogramme von 2a und 2b. Dargestellt sind die Mittelwerte und die Standardabweichung von fünf technischen Replikaten. Die Linearität war innerhalb eines Stoffmengenbereiches von 2.13−173 fmol **2a** mit einer Genauigkeit von 88−113% gegeben. Die Stoffmenge (n) des internen Standards **2b** betrug 30 fmol. A = Integral des MS/MS-Peaks. n = LLOQ = *lower limit of quantification* (unteres Quantifizierungslimit); LLOD = *lower limit of detection* (unteres Detektionslimit; Intensität viermal über dem Hintergrundsignal).

Um die Anwendbarkeit der Methode zu demonstrieren, wurden säure-induzierte AP-Stellen in genomischer DNA quantifiziert. Während der Inkubation der DNA bei pH 3.8 und 70 °C konnte eine starke Zunahme von AP-Stellen innerhalb weniger Minuten festgestellt werden (**Abbildung 26**). Als weitere Positivkontrolle, um die Detektion von endogenen AP-Level mit dieser Methode unter Beweis zu stellen, wurden die AP-Stellen in genomischer DNA von mES-Zellen quantifiziert, deren Ape1-Enzymlevel (*Apurinic/Apyrimidinic endonuclease 1*) per *Knockdown* (KD) reduziert wurden. Das Ape1-Enzym hydrolysiert den Zucker-Phosphat-Ester von AP-Stellen und erzeugt somit einen Strangbruch für die Folgeprozesse der Basenexzisionsreparatur (**Schema 15**).[363, 374, 405] Durch *Knockdown* dieses Enzyms akkumulieren folglich AP-Stellen in genomischer DNA,[406] wie übereinstimmend mit dieser

135

Quantifizierungsmethode festgestellt werden konnte (**Abbildung 27**). Durch den Ape1-*Knockdown* wurden in mES-Zellen vierfach höhere Level von AP-Stellen im Vergleich zu Wildtyp-Zellen bestimmt (1.0 bzw. 3.9 AP pro 10^{-6} dN in Wildtyp- bzw. Ape1 KD-Zellen).
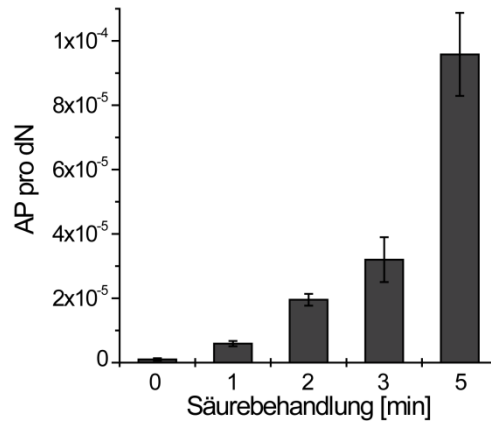
Abbildung 26 | Quantifizierung der Säure-induzierten Bildung von AP-Stellen in genomischer DNA. Die Generierung erfolgte durch Inkubation von genomischer DNA (aus dem Hoden des Herings) bei pH 3.8 und 70 °C. Dargestellt sind die Mittelwerte von drei unabhängigen Replikaten; Fehlerbalken repräsentieren die Standardabweichung.

Eine Korrelation von abasischen Stellen mit 5fdC oder 5cadC, die Substrate von Tdg bzw. der Basenexzisionsreparatur sind,[7, 9, 312, 346-347, 350] konnte jedoch nicht festgestellt werden: Der Gehalt von AP-Stellen unterschied sich nicht, wenn Dnmt3a/3b oder Tdg molekularbiologisch entfernt wurden (**Abbildung 27**). Allenfalls eine Reduktion der Enzymmengen von Tet1 oder Tet2 korrelierte mit der Reduktion von AP-Stellen. Ein ähnlicher Zusammenhang wurde in Abschnitt 3.6 bzgl. 5hmdU gefunden.[13] Folglich korrelieren die AP-Stellen positiv mit Tet-Expressions- und 5hmdU-Level. Schlussfolgernd besteht die Möglichkeit, dass Tet-Enzyme direkt für die Bildung von AP-Stellen durch einen Nebenreaktion verantwortlich sein könnten oder indirekt durch Basenexzisionsreparatur von 5hmdU[13] (siehe **Schema 15** und **Schema 16**). Dieser Zusammenhang könnte auch bei Vergleich mit adulten neuralen Stammzellen für deren geringere AP-Level gelten (3.3×10^{-7} pro dN). In diesen Zellen sind Tet-Enzyme offensichtlich kaum aktiv, wie an den vergleichsweise geringen Modifikationslevel von 5hmdC, 5fdC und 5hmdU abzuleiten ist.

Da *in vivo* bereits gezeigt wurde, dass 5fdC und 5cadC Substrate von Tdg bzw. der Basenexzisionsreparatur sind,[7, 9, 312, 346-347, 350] stellen die unbeeinflussten AP-Level bei einem *Knockout* von Dnmt3a/3b oder *Knockdown* von Tdg ein unerwartetes Ergebnis dar (**Abbildung 27**). Ein Grund könnte sein, dass die epigenetisch relevanten AP-Stellen bei den globalen LC-MS/MS-Messungen von AP-Stellen anderen Ursprungs überdeckt werden.

Hierfür kommen zum einen die bereits angesprochenen anderen Bildungswege von AP-Stellen in Frage (siehe **Schema 15**) und vor allem die Generierung von artifiziellen, exogenen AP-Stellen bei der DNA-Isolation oder der Derivatisierungsreaktion mit dem Hydroxylamin-Reagenz.
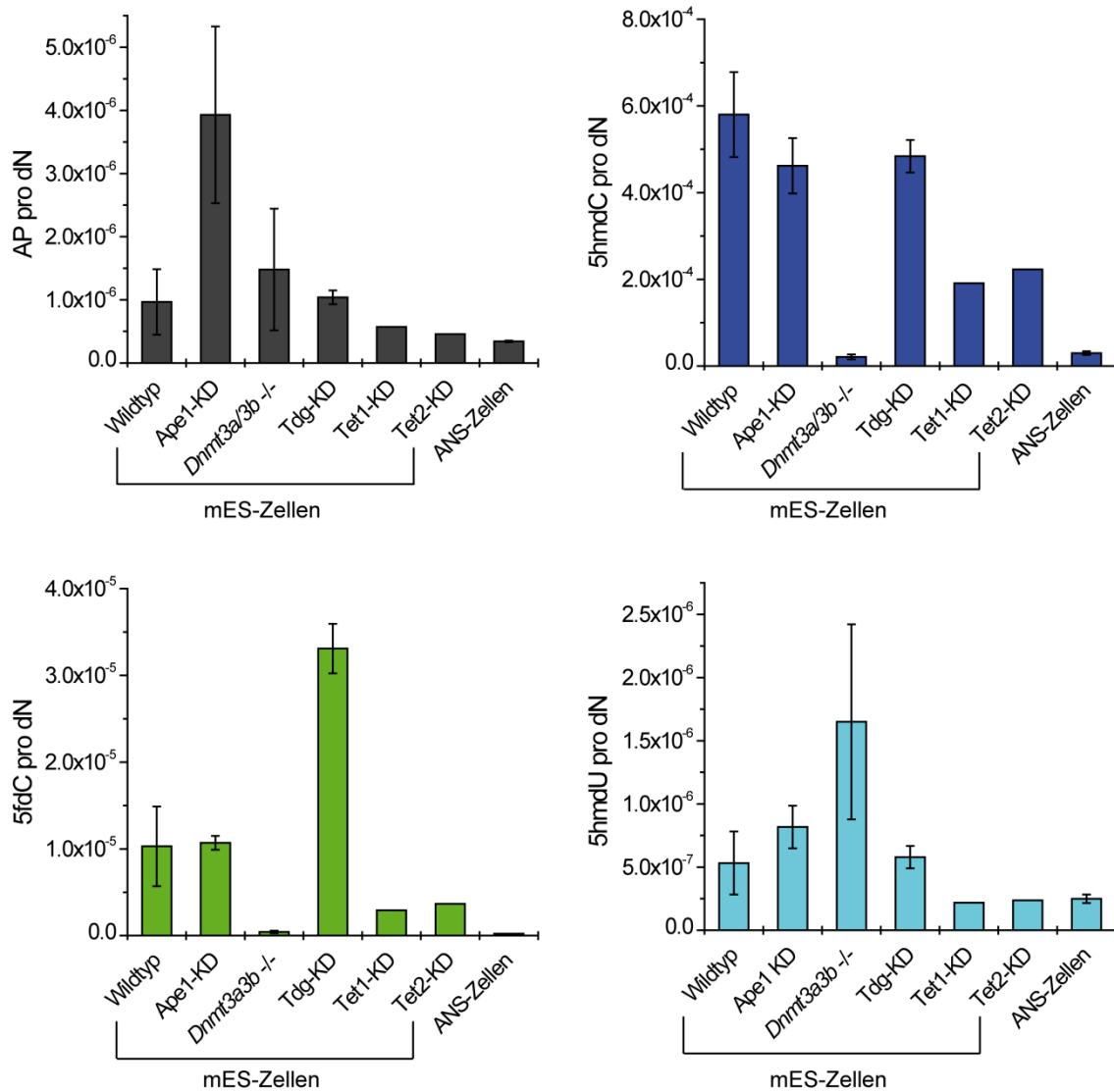
Abbildung 27 | Co-Quantifizierung von AP, 5hmdC, 5fdC und 5hmdU in embryonalen und adulten neuralen Stammzellen der Maus. Dargestellt sind die Ergebnisse nach Hydroxylamin-Behandlung. Unter den Reaktionsbedingungen (1.5 mM Reagenz **1a**, PBS-Puffer, 15 min, 37 °C) wurde keine Reaktion mit 5fdC beobachtet. Für Details und Modifikationslevel von 5mdC, 5fdU, 8oxodG siehe Masterarbeit von *Victor Brantl*[403] und Abschnitt 3.6 im Vergleich. Dargestellt sind die Mittelwerte biologischer Replikate ±Standardabweichung: Wildtyp-mES-Zellen (n = 6 biologische Replikate), Ape1 KD (2), Dnmt3ab$^{-/-}$ KO (3), Tdg KD (4), Tet1 KD (1), Tet2 KD (1), ANS-Zellen (4).

Um die Generierung von artifiziellen AP-Stellen bei der Inkubation der DNA mit dem Reagenz zu untersuchen, wurde eine Reaktionskinetik mit genomischer DNA durchgeführt. Dabei stellte sich heraus, dass die Reaktion im Gegensatz zu der Oligodesoxynukleotid-

Kinetik (**Abbildung 21**) im Spurenbereich keinen Sättigungsbereich erreichte (**Abbildung 28**). Als Grund wurde die geringe Pufferkapazität des verwendeten PBS-Puffers (pH 7.4)[367, 369-372] identifiziert, der die Azidität des Hydrochlorid-Salzes des Reagenzes nicht puffern konnte. Der pH-Wert der Reaktionsmischung lag daher im sauren Bereich.
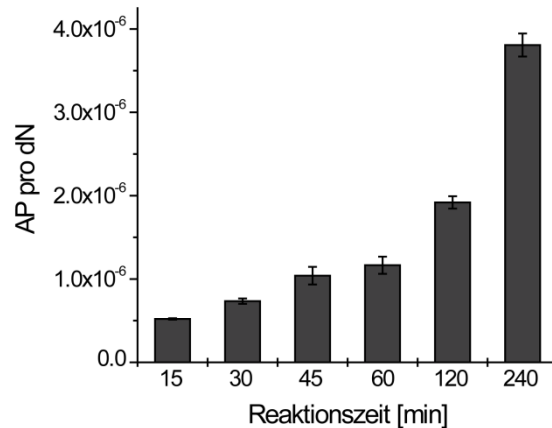
Abbildung 28 | AP-Stellen in genomischer DNA in Abhängigkeit der Reaktionsdauer mit Hydroxylamin 1a in PBS-Puffer (37 °C). Die verwendete DNA stammte aus dem Hoden des Herings.

Dieses Problem wird aktuell im Arbeitskreis *Carell* durch Verwendung eines Hepes-Puffers überwunden.[404] Anschließend soll die Untersuchung der epigenetischen AP-Stellen endogenen Ursprungs fortgesetzt werden. Insbesondere sollen die Methoden der Isotopenverfolgung zur Anwendung kommen. Wie in Abschnitt 3.6 gezeigt, kann die genomische DNA von mES-Zellen durch Zugabe Isotopen-markierter Nukleoside spezifisch über den Zellmetabolismus markiert werden.[13] So soll eine Isotopenmarkierung in der Riboseeinheit von dC bzw. dT eine Unterscheidungsmöglichkeit der globalen AP-Stellen bieten. Auf diese Weise können Veränderungen im Gehalt von AP-Stellen direkt auf ihren Modifikationsursprung zurückgeführt werden. Ferner hydrolysieren Pyrimidine mit einer 20mal geringeren Rate als Purine,[374] weswegen durch die Isotopen-Markierung von dC und dT die Unsicherheit gegenüber artifiziellen AP-Stellen minimiert werden könnte. Durch Zugabe von Isotopen-markierter DNA zum Zellaufschluss könnte ferner die Entstehung von artifiziellen abasischen Stellen bei der DNA-Isolation direkt verfolgt und evaluiert werden.

Weitere Anwendungen des Reagenzes

Wie bei allen Literatur-bekannten Methoden, ist auch bei der hier entwickelten Methode die Höhe der quantifizierten AP-Level von der Derivatisierungsreaktion abhängig. Artifizielle AP-Stellen können bei der Derivatisierungsreaktion und bei der vorhergehenden DNA-

Isolation entstehen.[371, 380, 385] Um letzteres zu umgehen, führten *Atamna et al.* AP-Markierungen mit ARP in lebenden Zellen oder intakten Zellkernen durch, bevor sie die DNA isolierten.[371] *In vivo* muss hierbei jedoch berücksichtigt werden, dass die Oximligationsprodukte die DNA-Reparaturprozesse beeinflussen, ferner die Hydroxylamine mit anderen Aldehyden und Keto-Gruppen im inneren der Zelle reagieren, und deshalb toxisch sein können. Ein analoger Einsatz des Reagenzes **1a** kann allerdings nicht direkt erfolgen, da es wahrscheinlich die Zellmembran nicht passiert.[403] Ein alternativer Ansatz wäre die *in vivo*-Applikation des retrosynthetischen Alkyl-Hydroxylamins (**Schema 18B**). In einer anschließenden "Klick"-Reaktion der isolierten DNA mit Azid **5a** könnte so ebenfalls das Oximligationsprodukt für die oben beschriebene LC-ESI-MS/MS-Quantifizierung erhalten werden.

Neben der massenspektrometrischen Detektion von AP-Stellen und deren Derivate, könnte das Reagenz **1a** auch für die Derivatisierung von anderen Aldehyden der DNA, wie 5fdC und 5fdU verwendet werden, um eine Empfindlichkeitssteigerung in Massenspektrometern zu erzielen. Die Derivatisierung von 5fdC ist bei einer Erhöhung der Konzentration des Reagenzes und bei verlängerten Inkubationszeiten im Vergleich zu AP-Stellen möglich.[403]

Die Derivatisierung von 5fdC könnte auch zur Generierung von hochspezifischen Antikörpern genutzt werden.[407] Erste Ergebnisse wurden im Arbeitskreis *Carell* bereits erzielt. Polyklonale Antikörper gegen das Oxim waren in ersten Untersuchungen spezifisch für 5fdC. Dieses Projekt wird von *Barbara Steigenberger* im Zuge ihrer Dissertation durchgeführt.

4.3.3 Zusammenfassung

Zusammenfassend konnte eine selektive Derivatisierungsmethode zur präzisen Quantifizierung von AP-Stellen etabliert werden, die sich durch eine schnelle Probenvorbereitung auszeichnet und die zum ersten Mal die Co-Quantifizierung von DNA-Modifikationen per Umkehrphasen-LC-MS/MS in einem analytischen Untersuchungslauf ermöglicht. Die Nachweisgrenze übertraf mit 0.2 fmol bisherige Methoden.[368, 373] Ferner genügte der Einsatz von lediglich 5 µg DNA pro Analyse. Der Gehalt von AP-Stellen betrug in dieser Machbarkeitsstudie in adulten neuronalen Stammzellen 3.3×10^{-7} pro dN und in mES-Zellen 1.0×10^{-6} pro dN. Die Tet-Enzyme sind in mES-Zellen möglicherweise für erhöhte Level von AP-Stellen verantwortlich. Da die bisherigen Bedingungen der Derivatisierungsreaktion allerdings zur artifiziellen Bildung von AP-Stellen im Spurenbereich

führte, müssen kontrollierte Reaktionsbedingungen im pH-neutralen Bereich geschaffen werden, bevor dies näher untersucht werden kann. Um ferner epigenetisch-relevante AP-Stellen direkt untersuchen zu können, müssen Isotopenverfolgungsexperimente durchgeführt werden. Die charakteristischen, intensiven MS-Übergänge des Reagenzes eignen sich darüber hinaus für die Untersuchung von natürlichen Derivaten der AP-Stellen, wie das β-Eliminierungsprodukt (siehe **Schema 15**). Dieses könnte durch bifunktionelle Glykosylasen wie den Neil-Enzymen beim Abbau von 5fdC oder 5cadC oder bei der Desaminierung von 5mdC gebildet werden.[14, 320, 346]

4.3.4 Projektbeiträge in diesem Abschnitt

Ich entwickelte das Hydroxylamin-Reagenz und etablierte die Methode zur Quantifizierung von abasischen Stellen. Die Ausarbeitung des Protokolls zur Derivatisierung und anschließenden Reinigung von DNA sowie die Vermessung der mES-Zellen erfolgte in Zusammenarbeit mit *Victor Brantl* im Zuge seiner Masterarbeit. *Victor Brantl* etablierte ferner die Zellkultur der adulten neuralen Stammzellen.[403] *Matthias Truss* (Charité Berlin, Labor für Pädiatrische Molekularbiologie) führte die molekularbiologischen Arbeiten der mES-Zellkultur durch. *Olesea Kosmatchev* und *René Rahimoff* führen das Projekt fort.

4.3.5 Materialien und Methoden

2-((Tritylamino)oxy)essigsäure (7)
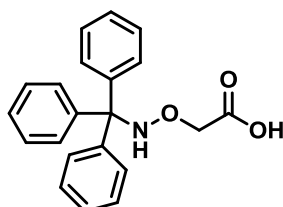
2-((Tritylamino)oxy)essigsäure wurde nach einer Vorschrift von *Kojima et al.* synthetisiert.[372]

***N*-(Prop-2-in-1-yl)-2-((tritylamino)oxy)acetamid (6)**
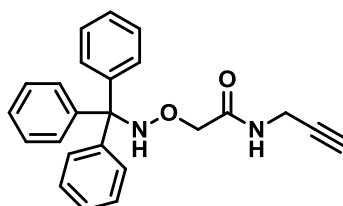
In einem trockenen Rundkolben wurde **7** (2.50 g, 7.50 mmol, 1.0 Äq.) zusammen mit TBTU (2.89 g, 9.00 mmol, 1.2 Äq.), DIPEA (1.6 mL, 9.19 mmol, 1.2 Äq.) und Propargylamin

(1.4 mL, 22.6 mmol, 3.0 Äq.) in DCM (40 mL) bei Raumtemperatur gerührt. Die anfängliche Suspension klarte auf und wurde für 16 h weitergerührt. Das Reaktionsgemisch wurde mit EtOAc (300 mL) verdünnt und mit gesättigter NH$_4$Cl-Lösung (1× 300 mL) und gesättigter NaHCO$_3$-Lösung (1× 300 mL) extrahiert. Die organische Phase wurde über Na$_2$SO$_4$ getrocknet und *in vacuo* bis zur Trockne eingeengt. Das Rohprodukt wurde anschließend per Säulenchromatographie (SiO$_2$; *i*Hex/EtOAc 10:1 → 2:1) gereinigt und Produkt **6** (1.93 g, 5.21 mmol, 69%) als farbloser Feststoff erhalten.

¹H-NMR (300 MHz, CDCl$_3$, ppm): δ = 7.37–7.22 (m, 15H, (C$_6$**H**$_5$)$_3$C), 6.59 (s, 1H, (C$_6$H$_5$)$_3$C-N**H**-O), 5.81 (bs, 1H, O=C-N**H**), 4.25 (s, 2H, O-C**H**$_2$-C=O), 3.85 (dd, ^3J=5.5, ^4J=2.6 Hz, 2H, HN-C**H**$_2$), 2.15 (t, ^4J = 2.6 Hz, 1H, C≡C-**H**). **¹³C-NMR** (75 MHz, CDCl$_3$, ppm): δ = 169.1 (C=O), 143.9 (3C, 3x O-NH-C-**C**), 129.0 (6C, **C**$_{Ar}$-H), 128.2 (6C, **C**$_{Ar}$-H), 127.4 (3C, **C**$_{para}$-H), 79.3 (C≡**C**-H), 74.6 (**C**(C$_6$H$_5$)$_3$), 73.4 (O-**C**H$_2$), 71.7 (**C**≡C-H), 28.8 (NH-**C**H$_2$). **HRMS** (ESI+): Berechnet für C$_{24}$H$_{22}$N$_2$NaO$_2$$^+$ [M+Na]$^+$: 393.1573; gefunden: 393.1571. **IR** (ATR): $\tilde{v}$ (cm^{-1}) = 3288 (w), 3222 (w), 3056 (w), 2913 (w), 2359 (w), 2339 (w), 1635 (m), 1542 (m), 1489 (m), 1065 (m), 996 (m), 763 (m), 747 (m), 707 (s), 697 (s), 685 (s), 627 (s). **Schmelz-bereich**: 157–158 °C. **Elementaranalytik** (%): berechnet für C$_{24}$H$_{22}$N$_2$O$_2$: N: 7.56, C: 77.81, H: 5.99; gefunden: N: 7.65, C: 77.16, H: 5.96.

2-((4-Azidophenyl)amino)-*N,N,N*-trimethyl-2-oxoethanaminium-Chlorid (5a)

Betain (0.30 g, 2.56 mmol, 1.0 Äq.) wurde *in vacuo* für 20 min bei 180 °C getrocknet. DMF (25 mL), DIPEA (1.1 mL, 6.15 mmol, 2.4 Äq.), TBTU (0.99 g, 3.07 mmol, 1.2 Äq.) und 4-Azidoanilin-Hydrochlorid (0.54 g, 3.18 mmol, 1.2 Äq.) wurden hinzugegeben und das Gemisch wurde für 15 h bei Raumtemperatur gerührt. Anschließend wurde das Gemisch *in vacuo* (60 °C) eingeengt und der zähflüssige Rückstand in einer Lösung aus DCM/MeOH/H$_2$O/NH$_3$ (90:10:0.6:0.6; 10 mL) gelöst und auf Kieselgel für die Säulenchromatographie aufgetragen. Die Reinigung per Normalphasen-Säulen-chromatographie (SiO$_2$ 7.5 cm × 4.5 cm (Länge × Breite); DCM/MeOH/H$_2$O/NH$_3$ 90:10:0.6:0.6) war durch das Ausnutzen der Ionenpaarung des Produktkations und Benzotriazol-1-olat als Nebenprodukt von TBTU möglich. Das resultierende Rohprodukt wurde in H$_2$O (50 mL) gelöst und der pH-Wert auf 1 mit HCl (2 M) gebracht. Die wässrige Phase wurde mit Et$_2$O extrahiert, bis die organische Phase per Dünnschichtchromatographie

keine UV-Absorption mehr aufwies. Die wässrige Phase wurde daraufhin mit konzentriertem Ammoniak neutralisiert und *in vacuo* (60 °C) bis zur Trockne eingeengt. Das Produkt **5a** (0.62 g, 2.30 mmol, 90%) wurde in Form eines braunen Feststoffs erhalten.

^{1}H-NMR (300 MHz, [D$_6$]-DMSO, ppm): δ = 11.22 (s, 1H, N**H**), 7.69 (d, ^3J=8.9 Hz, 2H, C**H**=C-NH), 7.12 (d, ^3J=8.4 Hz, 2H, **C**H=C-N$_3$), 4.42 (s, 2H, C**H$_2$**), 3.30 (s, 9H, N(C**H$_3$**)$_3$). **^{13}C-NMR** (101 MHz, [D$_6$]-DMSO, ppm): δ = 162.0 (**C**=O), 135.0 (NH-**C**=CH), 134.9 (N$_3$-**C**=CH), 121.2 (2C, NH-C=**C**H), 119.5 (2C, N$_3$-C=**C**H), 64.3 (**C**H$_2$), 53.4 (3C, N(**C**H$_3$)$_3$). **HRMS** (ESI+): Berechnet für C$_{11}$H$_{16}$N$_5$O$^+$ [M]$^+$: 234.1349; gefunden: 234.1348. **IR** (ATR): $\tilde{v}$ (cm^{-1}) = 3348 (w), 2983 (w), 2118 (s), 2083 (m), 1692 (s), 1676 (m), 1615 (m), 1549 (m), 1508 (s), 1287 (s), 1256 (m), 1050 (s), 1038 (s), 922 (s), 833 (s). **Schmelzbereich**: 144–146 °C.

N,N,N-Trimethyl-2-oxo-2-((4-(4-((2-(((tritylamino)oxy)acetamido)methyl)-1*H*-1,2,3-triazol-1-yl)phenyl)amino)ethanaminium-Chlorid (8a)

In einem trockenen Schlenk-Kolben wurde Alkin **6** (0.24 g, 0.65 mmol, 1.0 Äq.), Azid **5a** (0.18 g, 0.67 mmol, 1.0 Äq.) und CuBr•SMe$_2$ (40 mg, 0.20 mmol, 0.3 Äq.) in entgastem DCM (5 mL) und H$_2$O (5 mL) in einer Argonatmosphäre bei Raumtemperatur suspendiert. Während der fortschreitenden Reaktion bildete sich eine bräunliche Emulsion. Die Reaktion wurde nach 24 h beendet, das Gemisch *in vacuo* eingeengt und per Normalphasen-Chromatographie gereinigt (SiO$_2$ 5.0 cm × 4.5 cm (Breite × Länge); DCM/MeOH/H$_2$O/NH$_3$ 90:10:0.6:0.6). Das Triazol **8a** (0.32 g, 0.50 mmol, 77%) wurde als schwach gelber Feststoff erhalten.

^{1}H-NMR (400 MHz, [D$_6$]-DMSO, ppm): δ = 11.67 (s, 1H, N**H**-C$_6$H$_4$), 8.55 (s, 1H, CH$_2$-C=C**H**-N), 8.34 (s, 1H, Ph$_3$C-N**H**), 8.32 (t, ^3J=5.8 Hz, 1H, O=C-N**H**-CH$_2$), 7.91–7.84 (m, 4H, C$_6$**H**$_4$), 7.34–7.19 (m, 15H, C(C$_6$**H$_5$**)$_3$), 4.53 (s, 2H, (C**H$_2$**-N(CH$_3$)$_3$), 4.45 (d, ^3J=5.8, 2H, NH-C**H$_2$**), 3.85 (s, 2H, N-O-C**H$_2$**), 3.33 (s, 9H, N(C**H$_3$**)$_3$). **^{13}C-NMR** (101MHz, [D$_6$]-DMSO, ppm): δ = 169.7 (O=**C**-NH-CH$_2$), 162.4 (O=**C**-CH$_2$-N), 146.0 (CH$_2$-**C**=C), 144.1 (3C, O-NH-C-**C**), 138.1 (N-**C**=CH-CH), 132.6 (N-**C**=CH-CH), 128.9 (6C, **C$_{Ar}$**-H), 127.6 (6C, **C$_{Ar}$**-H), 126.7 (3C, **C$_{para}$**-H), 121.0 (CH$_2$-C=**C**H-N), 120.5 (4C, N-C=**C**H-**C**H=C-N), 73.7 (**C**(C$_6$H$_5$)$_3$),

73.2 (O-$\underline{C}$H$_2$), 64.4 ($\underline{C}$H$_2$-N(CH$_3$)$_3$), 53.4 (N($\underline{C}$H$_3$)$_3$), 33.8 (NH-$\underline{C}$H$_2$). **HRMS** (ESI+): Berechnet für C$_{35}$H$_{38}$N$_7$O$_3^+$ [M]$^+$: 604.3031; Gefunden: 604.3026. **IR** (ATR): $\tilde{v}$ (cm^{-1}) = 3387 (w), 3054 (w), 2923 (w), 1685 (m), 1613 (m), 1558 (m), 1519 (s), 1490 (m), 1446 (m), 1413 (m), 1312 (m), 1265 (m), 1224 (m), 1192 (m), 1085 (m), 1045 (m), 1002 (m), 990 (m), 948 (m), 922 (m), 876 (m), 838 (m), 757 (s), 698 (s), 627 (s). **Schmelzbereich**: 142–152 °C.

2-((4-(4-((2-(Ammoniooxy)acetamido)methyl)-1*H*-1,2,3-triazol-1-yl)phenyl)amino)-*N,N,N*-trimethyl-2-oxoethanaminium-Chlorid (1a)

Azid **8a** (129 mg, 0.20 mmol) wurde in DCM (3 mL) und wässriger HCl (3 mL, 2 M) gelöst. Die Emulsion wurde bei Raumtemperatur über Nacht bis zur Phasentrennung und vollständigem Aufklaren der Lösung gerührt. Das Ende der Reaktion wurde per ESI-MS geprüft. Die wässrige Phase wurde anschließend mit DCM extrahiert, bis die organische Phase per Dünnschichtchromatographie keine UV-Absorption mehr aufwies. Die wässrige Phase wurde anschließend *in vacuo* bis zur Trockne eingeengt und **1a** (87 mg, 0.20 mmol, 100%) als schwach rötlicher, hygroskopischer Feststoff erhalten. Die Reinheit wurde per UHPLC-UV geprüft und war 95% (Säule: Poroshell 120 SB-C$_8$ von *Agilent*; 2.7 µm, 2.1 mm × 150 mm; Gradient aus Wasser und MeCN mit 0.01% (v/v) Ameisensäure: 0→30% (v/v) MeCN in 6.0 min bei einem Fluss von 0.35 mL min^{-1}).

^{1}H-NMR (400 MHz, D$_2$O, ppm): δ = 8.40 (s, 1H, CH$_2$-C=C$\underline{H}$-N), 7.78 (d, ^3J = 8.4 Hz, 2H, CH-C$\underline{H}$=C-N$_3$), 7.72 (d, ^3J = 8.4 Hz, 2H, CH-C$\underline{H}$=C-NH), 4.75 (s, 2H, N-O-C$\underline{H}_2$), 4.66 (s, 2H, NH-C$\underline{H}_2$), 4.37 (s, 2H, **C$\underline{H}_2$**-N(CH$_3$)$_3$), 3.43 (s, 9H, N(C$\underline{H}_3$)$_3$). **^{13}C-NMR** (101 MHz, D$_2$O, ppm): δ = 168.8 (O=$\underline{C}$-NH-CH$_2$), 162.7 (O=$\underline{C}$-CH$_2$-N), 144.6 (CH$_2$-$\underline{C}$=C), 136.9 (N-$\underline{C}$=CH-CH), 133.5 (N-$\underline{C}$=CH-CH), 122.5 (2C, $\underline{C}$H=C-N$_3$), 122.4 (CH$_2$-C=$\underline{C}$H-N), 121.8 (2C, $\underline{C}$H=C-NH), 71.8 (O-$\underline{C}$H$_2$), 65.2 ($\underline{C}$H$_2$-N(CH$_3$)$_3$), 54.3 (N($\underline{C}$H$_3$)$_3$), 34.1 (NH-$\underline{C}$H$_2$). **HRMS** (ESI+): Berechnet für C$_{16}$H$_{24}$N$_7$O$_3^+$ [M]$^+$: 362.1935; gefunden: 362.1935. **IR** (ATR): $\tilde{v}$ (cm^{-1}) = 3130 (m), 3037 (s), 2807 (m), 2649 (m), 2363 (w), 1684 (s), 1610 (m), 1556 (s), 1517 (s), 1487 (m), 1475 (m), 1442 (m), 1403 (s), 1312 (m), 1262 (m), 1193 (m), 1128 (w), 1083 (w), 1048 (m), 991 (m), 967 (w), 921 (s), 837 (s).

N,N,N-Trimethyl-2-oxo-2-((4-(4-((2-((((3S,4R)-3,4,5-trihydroxypentyliden)-amino)oxy)acet-amido)methyl)-1H-1,2,3-triazol-1-yl)phenyl)amino)ethanaminium-Formiat (2a)

Hydroxylamin **1a** (50.0 mg, 0.115 mmol, 1.0 Äq.) und 2′-Desoxyribose (182 mg, 1.36 mmol, 11.8 Äq.) wurden in H_2O (2.7 mL) gelöst und in einem Thermomixer (*Eppendorf comfort*) bei 1400 rpm und 30 °C über Nacht inkubiert. Das Reaktionsende wurde per ESI-MS geprüft. Das Gemisch wurde anschließend filtriert (0.2 µm Spritzenfiltermembran) und per Umkehrphasen-HPLC gereinigt (Säule: VP 250/10 Nucleosil 100-7 C_{18} von *Macherey-Nagel*; Gradient aus Wasser und MeCN mit 2.5 mM Ammoniumformiat pH 4.3: 0→15% (v/v) MeCN in 45 min bei einem Fluss von 5 mL min^{-1}). Nach zweimaliger HPLC-Reinigung wurden die Fraktionen auf Reinheit (>99%) per UHPLC-UV-ESI-MS/MS geprüft (Säule: Poroshell 120 SB-C8 from *Agilent*; 2.7 µm, 2.1 mm × 150 mm; Gradient aus H_2O und MeCN mit 0.01% (v/v) Ameisensäure: 0→17% (v/v) MeCN in 6.0 min bei einem Fluss von 0.35 mL min^{-1}). Vereinigte Fraktion wurden lyophilisiert und Produkt **2a** (9.1 mg, 0.017 mmol, 15%) wurde als farbloser Schaum erhalten. Nach UHPLC-UV-ESI-MS/MS und NMR-Analyse lag die Verbindung in wässriger Lösung als Gemisch von *E*- und *Z*-Isomeren vor. Das Verhältnis war nach UHPLC-UV-Analyse 57/43. Eine Zuordnung der *E*/*Z*-Isomeren war nicht möglich. Im Nachfolgenden werden die Verbindungen deshalb A und B genannt.

^{1}H-NMR (600 MHz, D_2O, ppm): δ = 8.46 (s, 1H, **H**COO), 8.34 (s, 1H, CH_2-C=C**H**-N), 7.79 (d, *J*=9.0 Hz, 2H, CH-C**H**=C-N_3), 7.74–7.71 (m, 2.6H, CH-C**H**=C-NH, C1′-**H**^{A}), 7.08 (t, ^3J = 5.4 Hz, 0.4H, C1′-**H**^{B}), 4.67 (s, 0.8H, N-O-C**H$_2$**^{B}), 4.63 (s, 2H, NH-C**H$_2$**), 4.62 (s, 1.2H, N-O-C**H$_2$**^{A}), 4.35 (s, 2H, **CH$_2$**-N(CH$_3$)$_3$), 3.92–3.87 (m, 0.4H, C3′-**H**^{B}), 3.85–3.80 (m, 0.6H, C3′-**H**^{A}), 3.78–3.69 (m, 1H, 1x C5′-**H$_2$**), 3.66–3.53 (m, 2H, 1x C5′-**H$_2$**, C4′-**H**), 3.42 (s, 9H, N(C**H$_3$**)$_3$), 2.79–2.69 (m, 0.8H, C2′-**H$_2$**^{B}), 2.58–2.54 (m, 0.6H, 1x C2′-**H$_2$**^{A}), 2.41–2.35 (m, 0.6H, 1x C2′-**H$_2$**^{A}). **^{13}C-NMR** (150 MHz, D_2O, ppm): δ = 172.4 (O=**C**-NH-CH$_2$), 170.9 (H**C**OO), 162.7 (O=**C**-CH$_2$-N), 153.5 (C1′^{A}), 153.1 (C1′^{B}), 145.1 (CH$_2$-**C**=C), 136.8 (N-**C**=CH-CH), 133.5 (N-**C**=CH-CH), 122.5 (2C, **C**H=C-NH), 122.3 (CH$_2$-C=**C**H-N), 121.9 (2C, **C**H=C-N_3), 74.2 (C4′), 74.0 (C4′), 71.7 (N-O-**C**H$_2$^{B}), 71.5 (N-O-**C**H$_2$^{A}), 69.0 (C3′^{A}), 68.8 (C3′^{B}), 65.1 (**C**H$_2$-N(CH$_3$)$_3$), 62.3 (C5′), 54.3 (N(**C**H$_3$)$_3$), 34.1 (NH-**C**H$_2$), 32.4 (C2′^{A}), 29.2 (C2′^{B}). **HRMS** (ESI+): Berechnet für $C_{21}H_{32}N_7O_6^+$ [M]$^+$: 478.2409; gefunden: 478.2404.

2-((4-Azidophenyl)amino)-*N*,*N*,*N*-tris(methyl-*d₃*)-2-oxoethan-1-aminium-Chlorid (5b)

Die Synthese wurde analog zur Synthese von **5a** unter Verwendung von [D₁₁]-Betain (C₅D₁₁NO₂, 98% D; *EURISO-TOP GmbH*) durchgeführt. Die Isotopenmarkierung der Methylen-Gruppe war unter den Bedingungen nicht stabil und es vollzog sich ein vollständiger D/H-Austausch, wodurch das entsprechende [D₉]-markierte Produkt **5b** erhalten wurde.

¹H-NMR (600 MHz, D₂O, ppm): δ = 7.39 (d, ³*J*=8.6, 2H, CH-C**H**=C-NH), 7.04 (d, ³*J*=8.5, 2H, CH-**C**H=C-N₃), 4.18 (s, 2H, C**H₂**). **¹³C-NMR** (150 MHz, D₂O, ppm): δ = 162.7 (**C**=O), 137.5 (NH-**C**=CH), 132.5 (N₃-**C**=CH), 123.5 (2C, NH-C=**C**H), 119.6 (2C, N₃-C=**C**H), 65.0 (**C**H₂), N(**C**D₃)₃ nicht beobachtet. **HRMS** (ESI+): Berechnet für C₁₁H₇D₉N₅O⁺ [M]⁺: 243.1914; gefunden: 243.1916.

***N*,*N*,*N*-Tris(methyl-*d₃*)-2-oxo-2-((4-(4-((2-(((((3*S*,4*R*)-3,4,5-trihydroxypentyliden)-amino)oxy)acetamido)methyl)-1*H*-1,2,3-triazol-1-yl)phenyl)amino)ethan-1-aminium-Diformiat (2b)**

Die Synthese und Analytik wurde analog zur unmarkierten Verbindung **2a** unter Verwendung von Azid **5b** durchgeführt. Das Produkt **2b** wurde wie Oxim **2a** als Gemisch von *E*- und *Z*-Isomeren erhalten. Das Verhältnis war nach UHPLC-UV-Analyse 57/43 (siehe **Abbildung 29** unten). Eine Zuordnung der *E/Z*-Isomeren war nicht möglich. Im Nachfolgenden werden die Verbindungen deshalb A und B genannt.

¹H-NMR (600 MHz, D₂O, ppm): δ = 8.46 (s, 2H, **H**COO), 8.35 (s, 1H, CH₂-C=C**H**-N), 7.81 (d, ³*J* = 7.5 Hz, 2H, CH-C**H**=C-N₃), 7.75–7.73 (m, 2.6H, CH-C**H**=C-NH, C1´-**H**^A), 7.08 (t, ³*J* = 5.4 Hz, 0.4H, C1´-**H**^B), 4.67 (s, 0.8H, N-O-C**H₂**^B), 4.64 (s, 2H, NH-C**H₂**), 4.62 (s, 1.2H, N-O-C**H₂**^A), 4.34 (s, 2H, C**H₂**-N(CH₃)₃), 3.93–3.87 (m, 0.4H, C3´-**H**^B), 3.86–3.79 (m, 0.6H, C3´-**H**^A), 3.78–3.69 (m, 1H, 1x C5´-**H₂**), 3.67–3.53 (m, 2H, 1x C5´-**H₂,** C4´-**H**), 2.80–2.68 (m, 0.8H, C2´-**H₂**^B), 2.59–2.54 (m, 0.6H, 1x C2´-**H₂**^A), 2.43–2.34 (m, 0.6H, 1x C2´-**H₂**^A). **¹³C-NMR** (150 MHz, D₂O, ppm): δ = 172.4 (O=**C**-NH-CH₂), 170.9 (H**C**OO), 162.8 (O=**C**-

CH$_2$-N), 153.5 (C1$'^A$), 153.1 (C1$'^B$), 145.1 (CH$_2$-**C**=C), 136.8 (N-**C**=CH-CH), 133.6 (N-**C**=CH-CH), 122.6 (2C, **C**H=C-NH), 122.4 (CH$_2$-C=**C**H-N), 122.0 (2C, **C**H=C-N$_3$), 74.2 (C4$'$), 74.0 (C4$'$), 71.7 (N-O-**C**H$_2$^{B}), 71.5 (N-O-**C**H$_2$^{A}), 69.0 (C3$'^A$), 68.8 (C3$'^B$), 64.9 (**C**H$_2$-N(CD$_3$)$_3$), 62.3 (C5$'$), 53.3 (N(**C**D$_3$)$_3$; per HMBC-NMR) 34.1 (NH-**C**H$_2$), 32.4 (C2$'^A$), 29.2 (C2$'^B$). **HRMS** (ESI+): Berechnet für C$_{21}$H$_{23}$D$_9$N$_7$O$_6$$^+$ [M]$^+$: 487.2973; gefunden: 487.2967.
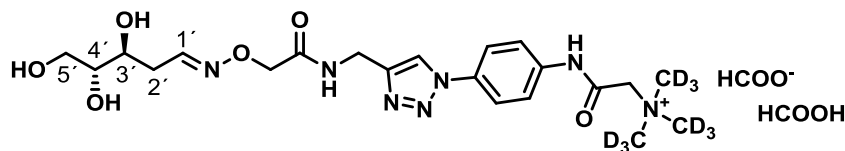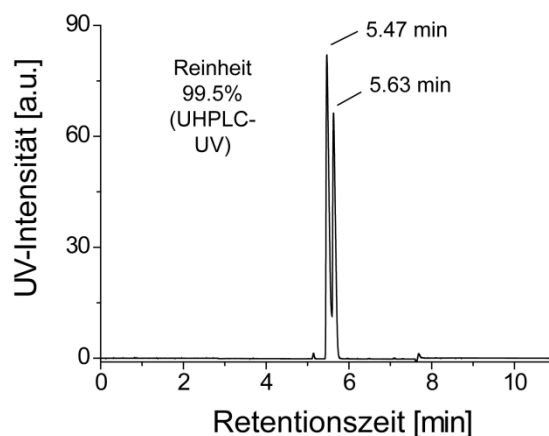
Abbildung 29 | Reinheitsanalyse des Oximligationsprodukts 2b per UHPLC-UV (260 nm). Für Chromatographiebedingungen siehe Beschreibung von Oximligationsprodukt 2a.

Bestimmung des Extinktionskoeffizienten

Da Hydroxylamin **1a** und die Oximligationsprodukte **2a/b** als hygroskopischer Feststoff bzw. Schaum erhalten wurden, war die Bestimmung des Extinktionskoeffizienten über die Einwaage der Verbindungen nicht möglich. Stattdessen wurde eine NMR-basierte Methode nach *Hong et al.* gewählt.[408] **1a** (Reinheit 95% bei UHPLC-UV) wurde mit einer bekannten Menge von Thymidin in D$_2$O gemischt. Anschließend wurde ein ^{1}H-NMR-Spektrum aufgenommen und die Menge/Konzentration von **1a** über die Verhältnisse der Signalflächen bestimmt. Insgesamt wurden für die Auswertung drei Signale von **1a** (δ = 8.40, 4.37, 3.43 ppm) und fünf Signale von Thymidin (δ = 6.32, 4.09–4.00, 3.90–3.75, 2.46–2.36, 1.87 ppm) verwendet. Ferner wurde der Versuch mit drei unterschiedlichen Mischungs-verhältnissen der Verbindungen wiederholt und alle Ergebnisse gemittelt. Nach der Bestimmung der UV-Absorption des Reagenzes mit einem *NanoDrop*-Spektrophotometer (ND-1000) wurde der Extinktionskoeffizient mit Hilfe des Lambert-Beerschen Gesetzes errechnet. Bei λ_{max} = 268 nm betrug dieser bei Raumtemperatur ε = 18800 mol^{-1} L cm^{-1} mit einer relativen Standardabweichung von ±0.8% und diente der Konzentrationsbestimmung von **1a**, **2a** und **2b**.

Herstellung von AP-enthaltenden Oligodesoxynukleotiden

Zwei komplementäre Oligodesoxynukleotide (je 10 nmol) mit der Sequenz 5′-GTA ATG U̲GC TAG G-3′ und 5′-CCT AGC GCA TTA C-3′, hergestellt von *Metabion*, wurden in einem UDG-Reaktionspuffer (100 µL; 20 mM Tris/HCl pH 8, 1 mM DTT, 1 mM EDTA; *New England Biolabs*) hybridisiert (3 min bei 90 °C, dann Temperaturgradient von 90 °C → 4 °C, 1 °C min^{-1}). Anschließend wurde UDG (6 µL, 30 U; *New England Biolabs*) hinzugegeben und die Mischung bei 37 °C für 2 h inkubiert. Die DNA wurde per Phenol/Chloroform-Extraktion und NaOAc/EtOH-Präzipitation gereinigt. Das DNA-Pellet wurde in H$_2$O aufgenommen und bei −20 °C gelagert.

Derivatisierung von AP-enthaltenden Oligodesoxynukleotiden mit Reagenz 1a

Das AP-enthaltende Oligodesoxynukleotid (300 pmol) wurde mit PBS-Puffer, Na$_2$-EDTA (0.1 mM) in H$_2$O versetzt und die Reaktion mit der Zugabe von Reagenz **1a** (1.5 mM) bei 37 °C/800 rpm in einem Thermomixer (*Eppendorf comfort*) gestartet. Das Reaktionsvolumen betrug 20 µL. Die Reaktion wurde mit der Zugabe von Aceton gestoppt, überschüssiges Aceton wurde an einer *Speedvac* (RVC-2-33 IR, *Christ*) entfernt. Für MALDI-TOF-Analysen und zur Kapillarzonengelelektrophorese wurde die Probe über 30 min auf einer 0.025 µm-Membran (VSWP, *Millipore*) gegen H$_2$O dialysiert.

MALDI-TOF-Analyse von Oligodesoxynukleotiden

MALDI-TOF Spektren wurden mit einem *Bruker*-Autoflex II-Massenspektrometer und einem AnchorChip™-Target (*Bruker MTP* var/384) generiert. Entsalzte Oligodesoxynukleotide wurden auf einer HPA-Matrix co-kristallisiert. Zusammensetzung der Matrix: 3-Hydroxypicolin-Säure (50 mg), 15-Krone-5 (10 µL), Ammonium-Hydrogencitrat (10 mg) in H$_2$O (500 µL) und MeCN (500 µL).

Kapillarzonengelelektrophorese

Die Kapillarzonengelelektrophorese von Oligodesoxynukleotiden wurde mit einem *P/ACE MDQ DNA System* mit UV-Detektor der Firma *Beckman Coulter* durchgeführt. Die verwendete Glaskapillare mit Polyacrylamid-Innenbeschichtung von *Beckman Coulter* hatte einen Innendurchmesser von 100 µm, eine Gesamtlänge von 30 cm. Es wurde ein lyophilisiertes 6%iges Polyacrylamidgel von *Beckman Coulter* verwendet, welches mit einem denaturierenden Gelpuffer (0.1 M Tris/Borat, 2 mM EDTA, 7 M Harnstoff) rehydratisiert wurde. Als Elektrodenpuffer wurde derselbe Puffer verwendet. Der Elektrodenpuffer und benötigtes H$_2$O wurden vorher durch Spritzenfilteraufsätze (0.2 µm *GHP Acrodisc 13*)

gefiltert und im Ultraschallbad entgast. Die Kapillare wurde mit dem Gel bei 30 psi über einen Zeitraum von 15 min beladen. Die DNA-Proben wurden vor der Analyse entsalzt (siehe oben) und auf eine Konzentration von 5 µM gebracht. Das Gel und die Proben wurden durch Zentrifugation (2 min, 6000 rpm) vor den Messungen entgast. Die Probeninjektion erfolgte elektrokinetisch mit einer Injektionsspannung von 10 kV für 2 s. Die Trennung wurde bei 11 kV, 25 °C und einem Druck von 20 psi über einen Zeitraum von 45 min durchgeführt. Oligodesoxynukleotide wurden durch Absorption bei 254 nm detektiert.

Erzeugung von AP-Stellen in genomischer DNA durch Säurebehandlung

Genomische DNA (aus dem Hoden des Herings) wurde von *Sigma* käuflich erworben. Die bereits vorhandenen AP-Stellen wurden mit einer *O*-Methylhydroxylamin-Lösung (5 mM in PBS-Puffer und 0.1 mM EDTA) zunächst maskiert.[369] Für die zeitabhängige Generierung von AP-Stellen wurde jeweils eine 5 µg-DNA-Probe in einem sauren Natriumcitrat-Puffer (20 µL, 10 mM Natriumcitrat, 10 mM NaH_2PO_4 und 10 mM NaCl, pH 3.8) bei 70 °C inkubiert.[369] Die Proben wurden mit NaOH neutralisiert und per NaOAc/EtOH-Fällung gereinigt.

Derivatisierung von genomischer DNA mit Reagenz 1a

Genomische DNA (5 µg) wurde mit PBS-Puffer, EDTA (0.1 mM) und Hydroxylamin **1a** (1.5 mM) in H_2O auf ein Volumen von 20 µL gebracht und für 15 min bei 37 °C und 800 rpm in einem Thermomixer (*Eppendorf comfort*) inkubiert. Die Reaktion wurde mit der Zugabe von 66.7 µL einer 2 M-Lösung von 1-Naphthylaldehyd in absolutem *iso*-PrOH und weiteren 10 min bei 37 °C und 1400 rpm beendet. Es folgte die Zugabe einer 3 M-Lösung von NaOAc (3.3 µL), einer Inkubation von 5 min bei 37 °C und 1400 rpm und einer weiteren Zugabe von absolutem *iso*-PrOH (66.7 µL). Anschließend wurde die DNA bei 15000 rpm und 10 °C für 30 min mit einer *Eppendorf*-Zentrifuge (5427 R) pelletiert, einmal mit *iso*-PrOH (75% v/v, 200 µL) und zweimal mit EtOH (75% v/v, 200 µL) gewaschen (15000 rpm, 30 min, 4 °C). Nach Entfernen des Restalkohols mit einer Pipette wurde das DNA-Pellet ohne Lufttrocknung sofort in H_2O (25 µL) aufgenommen, durch starkes mechanisches Schütteln mit einem *Qiagen-Tissue-Lyser* in Lösung gebracht (30 Hz, 1 min) und bei −20 °C gelagert.

DNA-Verdau für die LC-UV-ESI-MS/MS-Analyse

Die gereinigten DNA-Proben (5 µg in 25 µL) wurden nach dem Protokoll im Experimentalteil von Abschnitt 3.6 enzymatisch hydrolysiert.[13] Für die Quantifizierung der genomischen DNA-Modifikationen und der derivatisierten AP-Stellen wurden folgende interne Standards zu der DNA-Verdau-Lösung gegeben: 51.0 pmol [D_3]5mdC, 7.66 pmol [D_2,$^{15}N_2$]5hmdC,

45.6 fmol [^{15}N$_2$]5fdC, 43.0 fmol [^{15}N$_2$]5cadC, 109 fmol [^{15}N$_5$]8oxodG, 160 fmol [D$_2$]5hmdU, 180 fmol [^{15}N$_2$]5fdU und 29.9 fmol **2b**. [^{15}N$_5$]8oxodG wurde von *Cambridge Isotope Laboratories* erworben. [D$_3$]5mdC wurde von *Münzel et al.* synthetisiert.[4] Alle anderen Nukleosid-Derivate wurden im Zuge von Abschnitt 3.4 in dieser Arbeit synthetisiert.[12] Die verdauten DNA-Proben wurden vor der LC-UV-MS-Analyse mit einer AcroPrep™ Advance 96 Filterplatte (0.2 µm Supor®, *Pall Life Sciences*) filtriert.

LC-UV-ESI-MS/MS-Analyse

Die LC-UV-ESI-MS/MS-Analyse der Reaktionsprodukte **1a**, **2a** und **2b** sowie der enzymatisch hydrolysierten DNA-Proben erfolgte mit einem 1290 UHPLC-System mit UV-Detektor von *Agilent*, welches mit einem 6490 Triplequadrupol-Massenspektrometer von *Agilent* verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C$_8$-Säule (*Agilent*, 2.7 µm, 2.1 mm × 150 mm) durchgeführt. Die Trennung erfolgte durch einen Gradienten aus H$_2$O und MeCN mit jeweils 0.0075% (v/v) Ameisensäure. Der Fluss betrug dabei 0.35 mL min^{-1}. Für die Trennung von enzymatisch hydrolysierten DNA-Proben wurde folgender Gradient verwendet: 0 → 5 min; 0 → 3.5% MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 12.2 min; 5 → 80% MeCN; 12.2 → 13.8 min; 80% MeCN; 13.8 → 14.3 min; 80 → 0% MeCN; 14.3 → 16 min; 0% MeCN. Die Gradienten für die Produktanalysen von **1a**, **2a** und **2b** sind unter den Synthesevorschriften aufgeführt. Das Eluat bis 1.5 min und nach 12.2 min wurde dem Massenspektrometer per Ventilschaltung nicht zugeführt. Der *Autosampler* wurde auf 10 °C temperiert und das Injektionsvolumen betrug 29 µL. Die massenspektrometrischen Ionenquell-Parameter und die Auswertung der zu quantifizierenden Nukleoside waren identisch zum Experimentalteil von Abschnitt 3.6.[13] **Tabelle 3** und **Tabelle 4** fassen die Substanz-abhängigen Parameter für die Messung von genomischen bzw. synthetischen DNA-Proben zusammen.

Tabelle 3 | Substanz-abhängige MS/MS-Parameter für die Analyse von Hydrolysaten genomischer DNA. Die Fragmentorspannung war 380 V. Delta EMV wurde auf 500 (positiver Modus) und 800 (negativer Modus) gesetzt. CE: *collision energy*; CAV: *collision cell accelerator voltage*; EMV: *electron multiplier voltage*.

Substanz	Vorläufer-ion (*m/z*)	MS1 Auflösung	Produkt-ion (*m/z*)	MS2 Auflösung	*Dwell time* [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 1.5–4.0 min								
[^{15}N$_2$]5cadC	274.08	*wide*	158.03	*wide*	170	5	5	Positiv
5cadC	272.09	*wide*	156.04	*wide*	170	5	5	Positiv
[^{15}N$_2$,D$_2$]5hmdC	262.12	*enhanced*	146.07	*enhanced*	40	27	1	Positiv
5hmdC	258.11	*enhanced*	142.06	*enhanced*	40	27	1	Positiv
[D$_3$]5mdC	245.13	*enhanced*	129.09	*enhanced*	30	60	1	Positiv

Substanz	Vorläufer-ion (*m/z*)	MS1 Auflösung	Produkt-ion (*m/z*)	MS2 Auflösung	*Dwell time* [ms]	CE (V)	CAV (V)	Polarität
5mdC	242.11	*enhanced*	126.07	*enhanced*	30	60	1	Positiv
Zeitsegment 4.0–6.0 min								
[D₂]5hmdU	259.09	*wide*	216.08	*wide*	48	7	5	Negativ
5hmdU	257.08	*wide*	214.07	*wide*	48	7	5	Negativ
[¹⁵N₂]5fdU	257.06	*wide*	213.05	*wide*	48	6	5	Negativ
5fdU	255.06	*wide*	212.06	*wide*	48	6	5	Negativ
Zeitsegment 6.0–8.1 min								
[¹⁵N₅]8oxodG	289.08	*wide*	173.04	*wide*	120	9	7	Positiv
8oxodG	284.1	*wide*	168.05	*wide*	120	9	7	Positiv
[¹⁵N₂]5fdC	258.09	*wide*	142.04	*wide*	120	5	5	Positiv
5fdC	256.09	*wide*	140.05	*wide*	120	5	5	Positiv
Zeitsegment 8.1–12.2 min								
1-Naphthyl-Oxim	500.24	*enhanced*	472.23	*enhanced*	50	19	5	Positiv
2b	487.3	*wide*	459.29	*wide*	50	19	5	Positiv
2b	487.3	*wide*	201.18	*wide*	50	40	5	Positiv
2a	478.24	*wide*	450.23	*wide*	50	19	5	Positiv
2a	478.24	*wide*	192.13	*wide*	50	40	5	Positiv
1b	371.25	*wide*	343.24	*wide*	50	19	5	Positiv
1a	362.19	*wide*	334.19	*wide*	50	19	5	Positiv

Tabelle 4 | Substanz-abhängige MS/MS-Parameter für die Analyse von Hydrolysaten synthetischer DNA. Die Parameter für die Nukleosid-Verbindungen waren identisch zu Tabelle 3. Unterschiede betrafen die Delta EMV (300 im positiven, 800 im negativen Modus) sowie die Fragmentierungsparameter der Verbindungen **1a/b** und **2a/2b**.

Substanz	Vorläufer-ion (*m/z*)	MS1 Auflösung	Produkt-ion (*m/z*)	MS2 Auflösung	*Dwell time* [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 8.1–12.2 min								
1-Naphthyl-Oxim	500.24	*enhanced*	472.23	*enhanced*	50	19	5	Positiv
2b	487.3	*enhanced*	459.29	*enhanced*	50	13	2	Positiv
2b	487.3	*enhanced*	201.18	*enhanced*	50	25	2	Positiv
2a	478.24	*enhanced*	450.23	*enhanced*	50	13	2	Positiv
2a	478.24	*enhanced*	192.13	*enhanced*	50	25	2	Positiv
1b	371.25	*enhanced*	343.24	*enhanced*	50	10	1	Positiv
1a	362.19	*enhanced*	334.19	*enhanced*	50	10	1	Positiv

5 Schlussfolgerungen und Ausblick

Seit 1948 kannte man mit 5mdC nur eine DNA-Modifikation in Säugetieren.[54, 73] Erst über 60 Jahre später entdeckte man in den Jahren 2009 und 2010, dass 5hmdC als sechste Base des Genoms ein weiterer genereller Bestandteil von Säugetieren ist.[2-5] Durch die vorliegende und weitere Arbeiten folgte schließlich die Entdeckung, dass 5fdC, 5cadC und 5hmdU, als weitere natürliche Bestandteile der genomischen DNA vorkommen, die durch Tet-vermittelte Oxidation von 5mdC und dT entstehen (siehe 3.1, 3.6 und Referenzen).[6-8, 13, 409] Diese neue DNA-Modifikationsvielfalt hatte eine hohe Durchschlagskraft in der epigenetischen Forschungswelt, da man sich durch ein detailliertes Verständnis des epigenetischen DNA-Metabolismus neue Möglichkeiten in der regenerativen Medizin und der Krebstherapie verspricht.[84, 107, 187, 410]

Während man bei 5hmdC mittlerweile von einer größtenteils stabilen Modifikation mit epigenetischer Funktion in der Genregulation sprechen kann, ist dies bei 5fdC, 5cadC und 5hmdU noch völlig unklar.[14, 84, 203, 215-216] Allein ein Vergleich der Häufigkeiten der Modifikationen in genomischer DNA lässt vermuten, dass letztere wahrscheinlich dynamische bzw. transiente Funktionen erfüllen. In der Regel treten diese Modifikationen in Mengen auf, die im Vergleich zu 5hmdC global gesehen um zwei bis vier Größenordnungen geringer sind (siehe 3.1, 3.6 und Referenzen).[6-8, 13, 409] Diese Unterschiede können zum einen von der Affinität und Selektivität der Tet-Enzyme zu den Substraten herrühren. Aus ersten Studien ist bekannt, dass Tet1CD und Tet2CD 5mdC *in vitro* mit höherer Reaktionsgeschwindigkeit als 5hmdC und 5fdC oxidieren (siehe 4.2 und Referenz),[8] und dass sie eine deutlich höhere Selektivität für 5mdC gegenüber dT aufweisen (siehe 4.1). Weiterführende Arbeiten müssen zeigen, inwiefern sich die sequenzabhängige Substrat-affinität, Selektivität und Reaktionsrate der verschiedenen Tet-Enzyme im Detail unterscheiden. Im Gegensatz zu 5hmdC sind 5fdC, 5cadC und 5hmdU Substrate von Glykosylasen und werden deshalb durch die Basenexzisionsreparatur entfernt.[7, 9, 13, 312-313, 320] Kontrollierte Bildungs- und Abbauraten sind somit für das geringe Vorkommen dieser Modifikationen verantwortlich und machen dynamische bzw. transiente Funktionen wahrscheinlich. Solche Funktionen können die Aktivierung und "Feinjustierung" der Transkription,[261, 282, 346, 411] die Chromatin-Remodellierung[13-14, 208] oder das mRNA-*Splicing* sein.[84, 411] Hochaufgelöste Sequenzierdaten zu 5hmdC,[66, 167, 170, 174] 5fdC,[412-413] 5cadC[414]

151

und 5hmdU könnten in Zukunft weitere wichtige Hinweise für die Aufklärung der biologischen Funktionen liefern.

Die Proteom-Analyse in Abschnitt 3.5 zeigte, dass 5hmdC im Vergleich deutlich mehr Transkriptionsregulatoren bindet, während 5fdC und 5cadC mehr Reparaturproteine rekrutieren. Schlussfolgernd könnte die Oxidation von 5hmdC zu 5fdC ein Signal für die Reparatur-basierte (Tdg-vermittelte) DNA-Demethylierung sein, die unmodifiziertes dC wiederherstellt. In diesem Zusammenhang sei Uhrf2 hervorgehoben, welches in dieser Studie als 5hmdC-Binder identifiziert wurde, der im Modellsystem die Aktivität von Tet1 bei der sukzessiven Oxidation zu 5fdC und 5cadC steigerte.[14] Andere Studien demonstrierten, dass Vitamin C einen enormen aktivitätssteigernden Effekt auf Tet-Enzyme hat.[78, 321-323] Unklar ist bislang, wie die katalytische Aktivität von Tet-Enzymen reguliert wird bzw. ob und wie die sukzessiven Oxidationsreaktionen auf der jeweiligen Oxidationsstufe angehalten werden können. Hilfreich ist hierfür sicherlich, dass 5hmdC und 5fdC, wie oben erwähnt, *in vitro* langsamer oxidiert werden, als 5mdC.[8, 129] Ferner müssten DNA-bindende Proteine vorhanden sein, die vor Oxidation schützen, d.h. für die Erhaltung der jeweiligen Modifikationsstufe verantwortlich sind. Die Proteine MeCP2-MBD[210] oder Stella[94, 181] schützen z.B. 5mdC vor Oxidation und Mbd3 könnte ein möglicher Kandidat für 5hmdC sein.[190] In analoger Weise müsste dies für 5fdC, 5cadC und 5hmdU gegeben sein, falls diese für längerfristige Funktionen vor Oxidation oder Reparatur geschützt sein sollen. Insbesondere stellt sich bei 5fdC und 5cadC die Frage, ob diese wie 5mdC und 5hmdC bei der Zellteilung vererbt werden können,[203, 215-216, 415] andernfalls werden sie passiv verdünnt und dienen primär als Intermediate der DNA-Demethylierung.[92, 95, 416] Bei den Vorgängen der DNA-Oxidation, der aktiven und passiven Demethylierung sowie der Transkription kann hierbei ferner ein komplexes Zusammenspiel zwischen Histon- und DNA-Modifikationen erwartet werden, dessen Diversität die Wissenschaft vor neue Herausforderungen stellt.[84]

Die Komplexität biologischer Prozesse ist vermutlich für die Vielfalt von DNA-Modifikationen und Mechanismen in höheren Säugetieren verantwortlich, oder anders ausgedrückt, ohne diese wahrscheinlich nicht möglich. So ist es denkbar, dass die DNA-Modifikationen je nach Organismus, Zeitpunkt der Individualentwicklung oder je nach Zelltyp mit unterschiedlichen Proteinen interagieren (siehe z.B. auch 3.5), unterschiedlich gebildet und abgebaut werden.

Gerade für die gezielte Reprogrammierung von Zellen in der regenerativen Medizin oder in der Krebstherapie ist eine detaillierte Aufklärung der Demethylierungs- bzw. Demodi-

fizierungsprozesse von fundamentaler Bedeutung. Mit der Entdeckung der stufenweisen Oxidation von 5mdC und der Entfernung von 5fdC und 5cadC durch Tdg aus dem Genom, konnte zum ersten Mal eine vollständige enzymatische Kaskade beschrieben werden, mit der 5mdC in Zellen aktiv in dC umgewandelt werden kann.[5-9, 346] Andere Mechanismen, bei denen 5mdC und 5hmdC desaminiert werden, könnten ebenfalls eine Rolle spielen. Nach den in dieser Arbeit durchgeführten MS-basierten Isotopenverfolgungsexperimenten (siehe 3.6), ist die Desaminierung von 5hmdC in mES-Zellen jedoch vernachlässigbar.[13] Die etablierte UHPLC-ESI-MS/MS-Methode war empfindlich genug, um die unterschiedlichen Isotopen-markierten Spezies der Modifikationen im Spurenbereich zu detektieren. Aktuell wird dieser Ansatz im Arbeitskreis *Carell* für die Untersuchung der Desaminierung von 5mdC[298-304] weiter genutzt.

Als attraktive Alternative zu den Reparatur-basierten DNA-Demethylierungsprozessen wurden in dieser Arbeit C–C-Bindungsbruchreaktionen von 5hmdC, 5fdC und 5cadC untersucht, die unmodifiziertes dC auf direktem Wege generieren würden (siehe 3.3, 3.4 und 4.2). Mit Hilfe von Isotopen-markierten Oligonukleotiden konnte in mES-Zellkernextrakten in der Tat eine schwache Decarboxylierungsaktivität von 5cadC festgestellt werden (siehe 3.3).[11] Ferner konnte in chemischen Modellstudien gezeigt werden, dass ein möglicher Mechanismus über eine Thiol-vermittelte transiente Absättigung der C(5)=C(6)-Doppelbindung ablaufen kann (siehe 3.4).[12] Dabei nahm die Effizienz der C–C-Bindungsbruchreaktion in der Reihe 5cadC > 5fdC > 5hmdC ab. Trotz der in Abschnitt 3.3 gefundenen decarboxylierenden Aktivität von mES-Zellkernextrakten, fehlen entsprechende Nachweise, ob solche Reaktionen *in vivo* in einer beträchtlichen Menge auftreten und ob ein entsprechendes Enzym existiert (siehe hierzu auch 4.2). Ein experimenteller Ansatz, der die Untersuchung dieses Sachverhaltes in Zukunft ermöglichen könnte, ist in **Schema 21A** dargestellt. Durch die zelluläre Aufnahme und den Einbau von synthetisch hergestelltem [D$_2$,^{15}N$_2$]-dC in genomische DNA, welches neben der zweifachen ^{15}N-Markierung der N(1)- und N(3)-Position auch eine Deuterierung der C(5)- und C(6)-Position aufweist, könnte über den epigenetischen DNA-Metabolismus ein D/H-Austausch in C(5)-Position erfolgen. Sofern in Kontrollversuchen ein unspezifischer Verlust der Deuterierung ausgeschlossen werden kann,[417-419] wäre in der genomischen DNA ein D/H-Austausch nur dann denkbar, wenn über Methylierung, Oxidation und C–C-Bindungsbruch unmodifiziertes dC wiederhergestellt wird. Durch MS/MS-basierte Quantifizierung der [M+3] und [M+4]-Isotopologen könnte aufgrund der geringen natürlichen Häufigkeit entsprechender natürlicher Isotopologe der C–C-

Bindungsbruch in der genomischen DNA im Spurenbereich verfolgt werden. Mit diesem Ansatz könnten verschiedene Zellarten und verschiedene Differenzierungszustände von Stammzellen untersucht werden und die zugrundeliegenden Enzyme per *Knockout* identifiziert werden. Entscheidend für diesen Ansatz ist jedoch eine hohe synthetische Isotopenanreicherung der Reporterverbindung (**Schema 21B**). Diese könnte über einen Pt(0)-katalysierten H/D-Austausch von [^{15}N$_2$]-Uracil in D$_2$-Atmosphäre[420-421] analog zur Synthese von [^{15}N$_2$]-dC (siehe 3.4) zugänglich sein.
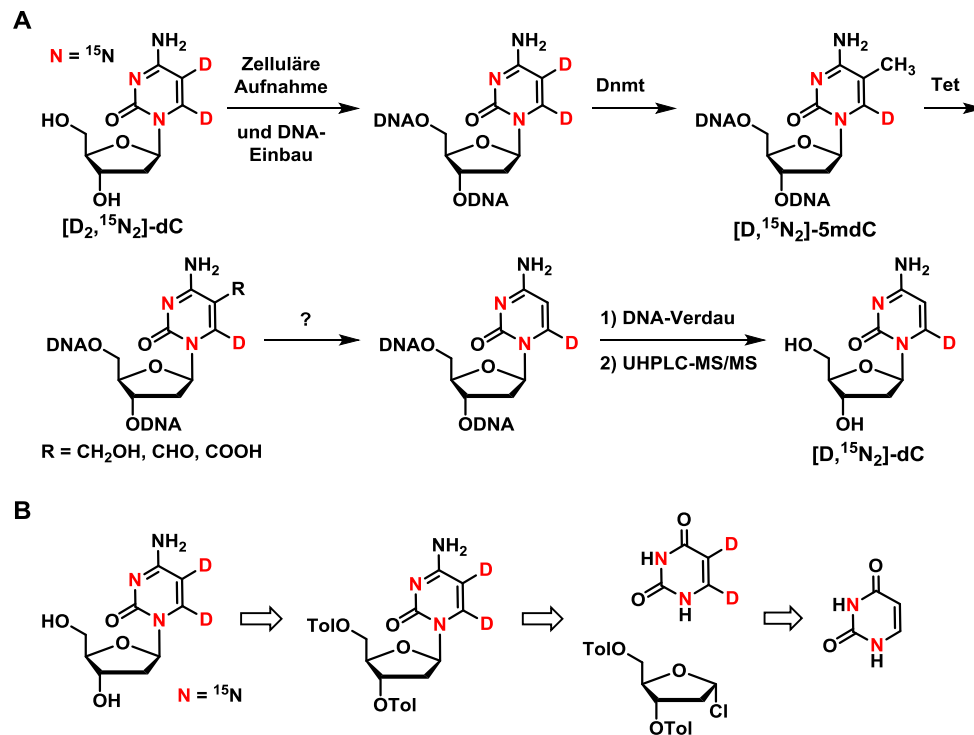
Schema 21 | Entwurf eines Isotopenverfolgungsexperiments zur Untersuchung von C–C-Bindungsbruchaktivitäten *in vivo*. (**A**) Arbeitsmodell. (**B**) Retrosynthese der Isotopen-markierten Reporterverbindung [D$_2$,^{15}N$_2$]-dC.

Neben der Reparatur-basierten oxidativen DNA-Demethylierung von 5mdC wird wahrscheinlich auch ein großer Teil von 5fdC und 5cadC bei der Zellteilung passiv verdünnt.[92, 95, 416] Unklar ist dabei, wie groß der Anteil des jeweiligen Prozesses in mES-Zellen ist. Durch die Quantifizierung von AP-Stellen mit Hilfe des in dieser Arbeit entwickelten Hydroxylamin-Reagenzes (siehe 4.3) könnten wichtige Rückschlüsse über den Beteiligungsgrad der Basenexzisionsreparatur gewonnen werden und herausgefunden werden, welche Prozesse zu welchem Anteil für die Bildung von AP- und uAP-Stellen verantwortlich sind. Um dabei die epigenetisch-relevanten AP und uAP-Stellen zu detektieren, können erneut Isotopen-Markierungsexperimente, wie in **Schema 22** dargestellt, zum Einsatz kommen. Unterschiedliche Markierungen in der Ribose-Einheit von synthetischen dC- und dT-

Isotopologen könnten eine massenspektrometrische Unterscheidung von AP-Stellen ermöglichen.
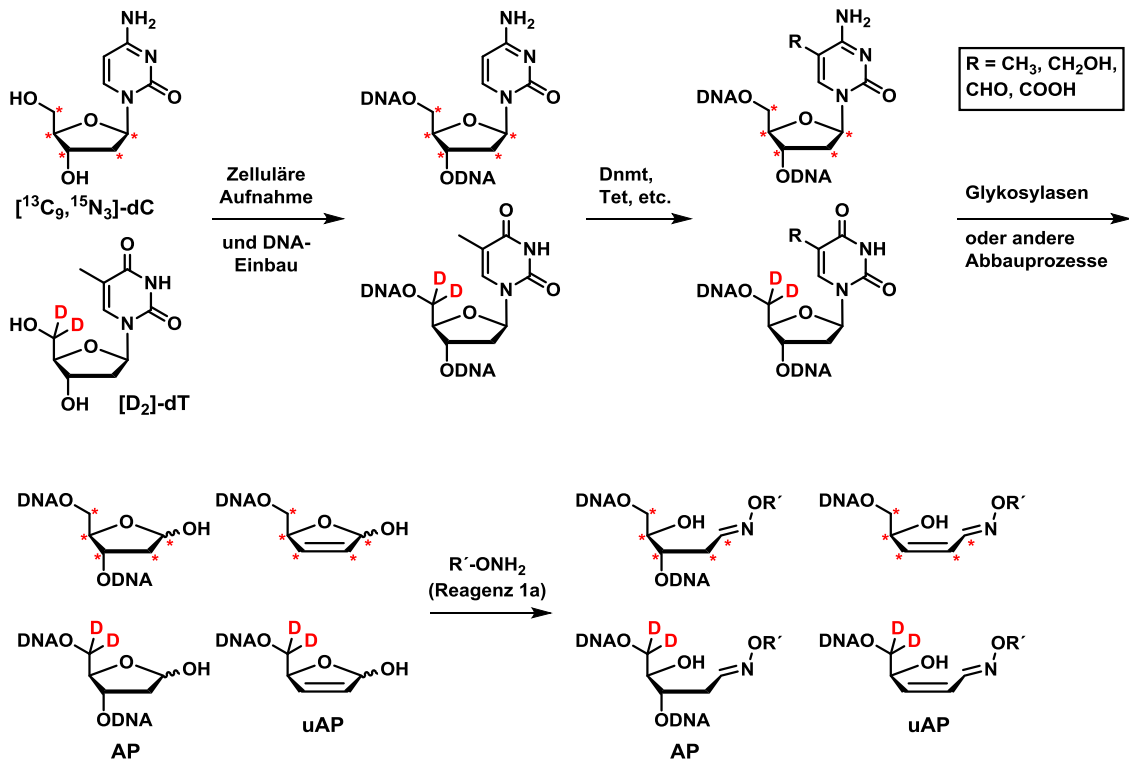
Schema 22 | Entwurf eines MS/MS-basierten Isotopenverfolgungsexperiments zur Untersuchung von AP-und uAP-Stellen epigenetischen Ursprungs (siehe Abschnitt 4.3). Die synthetischen dC- und dT-Isotopologen sind käuflich erwerblich ([^{13}C$_9$,^{15}N$_3$]-dC von *Hartmann-Analytic*, [D$_2$]-dT von *Cambridge Isotope Laboratories*). Aus Gründen der Übersichtlichkeit sind nur die ^{13}C-Atome der Riboseeinheit von [^{13}C$_9$,^{15}N$_3$]-dC mit einem roten Stern symbolisiert.

6 Abkürzungsverzeichnis

-/-	Knockout beider Allele eines Enzyms
5caC	5-Carboxycytosin
5cadC	5-Carboxy-2′-desoxycytidin
5cadU	5-Carboxy-2′-desoxyuridin
5caU	5-Carboxyuracil
5fC	5-Formylcytosin
5fdC	5-Formyl-2′-desoxycytidin
5fdU	5-Formyl-2′-desoxyuridin
5fU	5-Formyluracil
5hmC	5-Hydroxymethylcytosin
5hmdC	5-Hydroxymethyl-2′-desoxycytidin
5hmdU	5-Hydroxymethyl-2′-desoxyuridin
5hmU	5-Hydroxymethyluracil
5mC	5-Methylcytosin
5mdC	5-Methyl-2′-desoxycytidin
8oxodG	8-Oxo-2′-desoxyguanosin
8oxoG	8-Oxoguanin
A	Adenin
Ac	Acetyl
Aid	*activation-induced deaminase*
amu	*atomic mass unit*
ANS-Zellen	adulte neurale Stammzellen
Ape1	AP-Endonuklease 1
Apobec	*apolipoprotein B pre-mRNA editing enzyme, catalytic polypeptide*
AP-Stelle	apurinische/apyrimidinische Stelle
ARP	*aldehyde reactive probe*
ATP	Adenosintriphosphat
ATR	*attenuated total reflection*
BER	Basenexzisionsreparatur
C	Cytosin
CE	*collision energy*
CID	*collision induced dissociation*
CpG	dC-dG-Dinukleotid
dA	2′-Desoxyadenosin
dC	2'-Desoxycytidin
DC	Dünnschichtchromatographie
DCM	Dichlormethan

dG	2′-Desoxyguanosin
DIPEA	*N,N*-Di*iso*propylethylamin
DMF	Dimethylformamid
dN	2′-Desoxynukleosid
DNA	*deoxyribonucleic acid;* Desoxyribonukleinsäure
Dnmt	DNA-Methyltransferase
dRP	5′-Desoxyribosephosphat-Ende
DSBH	doppelsträngige β-Helix-Region
dsDNA	doppelsträngige DNA
dT	2′-Desoxythymidin
dTMP	2′-Desoxythymidin-Monophosphat
dTTP	2′-Desoxythymidin-Triphosphat
dU	2′-Desoxyuridin
EDTA	*ethylenediaminetetraacetic acid*
ESI	Elektrospray-Ionisierung
ES-Zelle	embryonale Stammzelle
Et	Ethyl
EtOAc	Ethylacetat
EtOH	Ethanol
FACS	*fluorescence activated cell sorting*
FAD	Flavin-Adenin-Dinukleotid
FT-ICR	*Fourier-Transformation-Ion-Cyclotron*
G	Guanin
GFP	Grün fluoreszierendes Protein
H3K14ac	Acetyliertes Lysin 14 in Histon H3
H3K27me3	Trimethyliertes Lysin 27 in Histon H3
H3K9ac	Acetyliertes Lysin 9 in Histon H3
H3K9me3	Trimethyliertes Lysin 9 in Histon H3
HAT	Histon-Acetyltransferase
HDAC	Histon-Deacetylase
HEK	*human embryoid kidney*
Hepes	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure
HMBC	*heteronuclear multiple bond correlation*
HMT	Histon-Methyltransferase
HP1	Heterochromatin-Protein 1
HPLC	Hochdruckflüssigkeitschromatographie
HRMS	hochauflösene (*high resolution*) Massenspekrometrie
Idc	Isoorotatdecarboxylase
Idh	Isocitrat-Dehydrogenase

iHex	Isohexan
iPrOH	Isopropanol
iPS-Zellen	induzierte pluripotente Stammzellen
IR	Infrarot
Isoorotat	5-Carboxyuracil (5caU)
IZM	innere Zellmasse
J	5-(β-D-Glukopyranosyloxymethyl)-2′-desoxyuridin
Jbp	J-bindendes Protein
KD	*Knockdown*
KO	*Knockout*
LIF	*leukemia inhibitory factor*
MALDI-TOF	*matrix assisted laser desorption ionization time of flight*
Me	Methyl
MeCN	Acetonitril
MeCP2	*methyl-CpG binding protein 2*
MeOH	*Methanol*
mES-Zelle	embryonale Stammzelle der Maus
MLL	*Mixed Lineage Leukemia*
MS	Massenspektrometrie, Massenspektrometer
MS/MS	Tandemmassenspektrometrie, Triplequadrupolmassenspektrometer
mTet	Tet-Enzym der Maus
Nanog	Transkriptionsfaktor; der Name stammt von *Tír na nóg* (Land ewiger Jugend)
Neil	*endonuclease VIII-like glycosylase*
NMR	*nucleic magnetic resonance*
NOG	*N*-Oxalylglycin
Oct4	Octamer-bindender Transkriptionsfaktor 4
ODN	Oligodesoxynukleotid
OG	2-Oxoglutarat
Ogt	*O-linked* β-D-*N*-Acetylglucosamin-Transferase
P	Signifikanzwert berechnet durch einen *t*-Test
PAGE	Polyacrylamidgelelektrophorese
PBS	*phosphate buffered saline*
PDB	*protein data bank*
ROS	*reactive oxygen species*
rpm	*rounds per minute*
SAH	*S*-Adenosylhomocystein
SAM	*S*-Adenosylmethionin
SD	Standardabweichung; *standard deviation*
Smug1	*single-stranded-selective monofunctional uracil-DNA glycosylase 1*

ssDNA	einzelsträngige (*single-stranded*) DNA
T	Thymin
T7h	Thymin-7-Hydroxylase
TBTU	*O*-(Benzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium-Tetrafluoroborat
Tdg	Thymin-DNA-Glykosylase
Tet	*Ten-eleven translocation* Protein
Tet1CD	rekombinante katalytische (*catalytic*) Domäne von Tet1
Tet1FL	rekombinantes Volllängen-Tet1
Tet3FL	rekombinantes Volllängen-Tet3
U	Uracil
uAP	α/β-ungesättigte AP-Stelle
UDG	Uracil-DNA-Glykosylase
UHPLC	Ultrahochdruckflüssigkeitschromatographie
UHPLC-UV-ESI-MS/MS	Ultrahochdruckflüssigkeitschromatographie gekoppelt mit UV-Detektion, Elektrospray-Ionisierung und Tandemmassenspektrometrie
Uhrf	*ubiquitin-like, containing PHD and RING finger domains*
UV	ultraviolettes Licht
v/v	Volumenanteil
WT	Wildtyp

7 Literaturverzeichnis

[1]   Z. D. Smith, A. Meissner, *Nat. Rev. Genet.* **2013**, *14*, 204-220. *DNA methylation: roles in mammalian development*.

[2]   S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930. *The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain*.

[3]   D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367. *Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates*.

[4]   M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375-5377. *Quantification of the sixth DNA base hydroxymethylcytosine in the brain*.

[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. *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*.

[6]   T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012. *The discovery of 5-formylcytosine in embryonic stem cell DNA*.

[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. *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA*.

[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. *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine*.

[9]   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*.

[10]   M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chemistry* **2011**, *17*, 13782-13788. *Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine*.

[11]   S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516-6520. *Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing*.

[12]   S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schroder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599. *Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine*.

[13]   T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574–581. *Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA*.

[14]   C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159. *Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives*.

[15]   O. T. Avery, C. M. Macleod, M. McCarty, *J. Exp. Med.* **1944**, *79*, 137-158. *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii*.

[16]   A. D. Hershey, M. Chase, *J. Gen. Physiol.* **1952**, *36*, 39-56. *Independent functions of viral protein and nucleic acid in growth of bacteriophage*.

[17]   R. Holliday, *Biol. Rev. Cambridge Philos. Soc.* **1990**, *65*, 431-471. *Mechanisms for the control of gene activity during development*.

[18]   G. Egger, G. Liang, A. Aparicio, P. A. Jones, *Nature* **2004**, *429*, 457-463. *Epigenetics in human disease and prospects for epigenetic therapy*.

[19]   A. Bird, *Nature* **2007**, *447*, 396-398. *Perceptions of epigenetics*.

[20]   O. J. Rando, K. J. Verstrepen, *Cell* **2007**, *128*, 655-668. *Timescales of genetic and epigenetic inheritance*.

[21]   T. Bartke, M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann, T. Kouzarides, *Cell* **2010**, *143*, 470-484. *Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation*.

[22]   R. Bonasio, S. J. Tu, D. Reinberg, *Science* **2010**, *330*, 612-616. *Molecular Signals of Epigenetic States*.

[23]   H. Cedar, Y. Bergman, *Nat. Rev. Genet.* **2009**, *10*, 295-304. *Linking DNA methylation and histone modification: patterns and paradigms*.

[24]   A. Portela, M. Esteller, *Nat. Biotechnol.* **2010**, *28*, 1057-1068. *Epigenetic modifications and human disease*.

[25]   D. C. Kraushaar, K. Zhao, *Int. J. Biol. Sci.* **2013**, *9*, 1134-1144. *The epigenomics of embryonic stem cell differentiation*.

[26]   K. Takahashi, S. Yamanaka, *Cell* **2006**, *126*, 663-676. *Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors*.

[27]   S. H. Orkin, K. Hochedlinger, *Cell* **2011**, *145*, 835-850. *Chromatin connections to pluripotency and cellular reprogramming*.

[28]   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*.

[29]   H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, M. P. Scott, A. Bretscher, H. Ploegh, P. Matsudiara, *Molecular cell biology*, 6 ed., W. H. Freeman, New York, **2007**.

[30]   D. J. Tremethick, *Cell* **2007**, *128*, 651-654. *Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber*.

[31]   M. A. Dawson, T. Kouzarides, *Cell* **2012**, *150*, 12-27. *Cancer epigenetics: from mechanism to therapy*.

[32]   A. Barski, S. Cuddapah, K. Cui, T.-Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, K. Zhao, *Cell* **2007**, *129*, 823-837. *High-Resolution Profiling of Histone Methylations in the Human Genome*.

[33]   T. Jenuwein, C. D. Allis, *Science* **2001**, *293*, 1074-1080. *Translating the Histone Code*.

[34]   T. Kouzarides, *Cell* **2007**, *128*, 693-705. *Chromatin Modifications and Their Function*.

[35]   B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41-45. *The language of covalent histone modifications*.

[36]   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*.

[37]   Z. Wang, C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, W. Peng, M. Q. Zhang, K. Zhao, *Nat. Genet.* **2008**, *40*, 897-903. *Combinatorial patterns of histone acetylations and methylations in the human genome*.

[38]   N. D. Heintzman, R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, B. Ren, *Nat. Genet.* **2007**, *39*, 311-318. *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome*.

[39]   E. Meshorer, T. Misteli, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 540-546. *Chromatin in pluripotent embryonic stem cells and differentiation*.

[40]   T. K. Barth, A. Imhof, *Trends Biochem. Sci.* **2010**, *35*, 618-626. *Fast signals and slow marks: the dynamics of histone modifications*.

[41]   K. H. Hansen, A. P. Bracken, D. Pasini, N. Dietrich, S. S. Gehani, A. Monrad, J. Rappsilber, M. Lerdrup, K. Helin, *Nat. Cell Biol.* **2008**, *10*, 1291-1300. *A model for transmission of the H3K27me3 epigenetic mark*.

[42]   P. A. Levene, *J. Biol. Chem.* **1919**, *40*, 415-424. *The Structure of Yeast Nucleic Acid*.

[43]   J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737-738. *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*.

[44]   E. Chargaff, S. Zamenhof, C. Green, *Nature* **1950**, *165*, 756-757. *Composition of human desoxypentose nucleic acid*.

[45]   R. E. Franklin, R. G. Gosling, *Nature* **1953**, *171*, 740-741. *Molecular configuration in sodium thymonucleate*.

[46]   J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 964-967. *Genetical implications of the structure of deoxyribonucleic acid*.

[47]   M. H. Wilkins, A. R. Stokes, H. R. Wilson, *Nature* **1953**, *171*, 738-740. *Molecular structure of deoxypentose nucleic acids*.

[48]   M. Meselson, F. W. Stahl, *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 671-682. *The Replication of DNA in Escherichia Coli*.

[49]   E. T. Kool, *Curr. Opin. Chem. Biol.* **2000**, *4*, 602-608. *Synthetically modified DNAs as substrates for polymerases*.

[50]   J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry*, 6 ed., W. H. Freeman, New York, **2006**.

[51]   F. H. Crick, *Symp. Soc. Exp. Biol.* **1958**, *12*, 138-163. *On protein synthesis*.

[52]   F. H. Crick, *Nature* **1970**, *227*, 561-563. *Central dogma of molecular biology*.

[53]   R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Nature* **1980**, *287*, 755-758. *Crystal structure analysis of a complete turn of B-DNA*.

161

[54] R. D. Hotchkiss, *J. Biol. Chem.* **1948**, *175*, 315-332. *The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography*.

[55] P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068-1070. *The role of DNA methylation in mammalian epigenetics*.

[56] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204-220. *Establishing, maintaining and modifying DNA methylation patterns in plants and animals*.

[57] A. Jeltsch, *ChemBioChem* **2002**, *3*, 274-293. *Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases*.

[58] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *ChemBioChem* **2011**, *12*, 206-222. *Structure and function of mammalian DNA methyltransferases*.

[59] G. Liang, M. F. Chan, Y. Tomigahara, Y. C. Tsai, F. A. Gonzales, E. Li, P. W. Laird, P. A. Jones, *Mol. Cell. Biol.* **2002**, *22*, 480-491. *Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements*.

[60] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257. *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development*.

[61] S. Bheemanaik, Y. V. R. Reddy, D. N. Rao, *Biochem. J.* **2006**, *399*, 177-190. *Structure, function and mechanism of exocyclic DNA methyltransferases*.

[62] 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*.

[63] M. Ehrlich, M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, C. Gehrke, *Nucleic Acids Res.* **1982**, *10*, 2709-2721. *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells*.

[64] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315-322. *Human DNA methylomes at base resolution show widespread epigenomic differences*.

[65] J. U. Guo, Y. Su, J. H. Shin, J. Shin, H. Li, B. Xie, C. Zhong, S. Hu, T. Le, G. Fan, H. Zhu, Q. Chang, Y. Gao, G.-l. Ming, H. Song, *Nat. Neurosci.* **2014**, *17*, 215-222. *Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain*.

[66] R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, *Science* **2013**, *341*, 1237905. *Global epigenomic reconfiguration during mammalian brain development*.

[67] B. K. Duncan, J. H. Miller, *Nature* **1980**, *287*, 560-561. *Mutagenic deamination of cytosine residues in DNA*.

[68] A. M. Deaton, A. Bird, *Genes Dev.* **2011**, *25*, 1010-1022. *CpG islands and the regulation of transcription*.

[69] F. Antequera, A. Bird, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 11995-11999. *Number of CpG Islands and Genes in Human and Mouse*.

[70] M. W. Simmen, *Genomics* **2008**, *92*, 33-40. *Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals*.

[71] M. Gardiner-Garden, M. Frommer, *J. Mol. Biol.* **1987**, *196*, 261-282. *CpG Islands in vertebrate genomes*.

[72] J. M. Craig, W. A. Bickmore, *Nat. Genet.* **1994**, *7*, 376-382. *The Distribution of CPG Islands in Mammalian Chromosomes*.

[73] P. A. Jones, *Nat. Rev. Genet.* **2012**, *13*, 484-492. *Functions of DNA methylation: islands, start sites, gene bodies and beyond*.

[74] J. Nichols, A. Smith, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, DOI: 10.1101/cshperspect.a008128. *Pluripotency in the embryo and in culture*.

[75] A. Meissner, *Nat. Biotechnol.* **2010**, *28*, 1079-1088. *Epigenetic modifications in pluripotent and differentiated cells*.

[76] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607-620. *Active DNA demethylation: many roads lead to Rome*.

[77] G. Ficz, T. A. Hore, F. Santos, H. J. Lee, W. Dean, J. Arand, F. Krueger, D. Oxley, Y. L. Paul, J. Walter, S. J. Cook, S. Andrews, M. R. Branco, W. Reik, *Cell Stem Cell* **2013**, *13*, 351-359. *FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency*.

[78] K. Blaschke, K. T. Ebata, M. M. Karimi, J. A. Zepeda-Martinez, P. Goyal, S. Mahapatra, A. Tam, D. J. Laird, M. Hirst, A. Rao, M. C. Lorincz, M. Ramalho-Santos, *Nature* **2013**, *500*, 222-226. *Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells*.

162

[79] E. Habibi, A. B. Brinkman, J. Arand, L. I. Kroeze, H. H. Kerstens, F. Matarese, K. Lepikhov, M. Gut, I. Brun-Heath, N. C. Hubner, R. Benedetti, L. Altucci, J. H. Jansen, J. Walter, I. G. Gut, H. Marks, H. G. Stunnenberg, *Cell Stem Cell* **2013**, *13*, 360-369. *Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells*.

[80] K. Hayashi, H. Ohta, K. Kurimoto, S. Aramaki, M. Saitou, *Cell* **2011**, *146*, 519-532. *Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells*.

[81] H. G. Leitch, K. R. McEwen, A. Turp, V. Encheva, T. Carroll, N. Grabole, W. Mansfield, B. Nashun, J. G. Knezovich, A. Smith, M. A. Surani, P. Hajkova, *Nat. Struct. Mol. Biol.* **2013**, *20*, 311-316. *Naive pluripotency is associated with global DNA hypomethylation*.

[82] A. Maitra, D. E. Arking, N. Shivapurkar, M. Ikeda, V. Stastny, K. Kassauei, G. Sui, D. J. Cutler, Y. Liu, S. N. Brimble, K. Noaksson, J. Hyllner, T. C. Schulz, X. Zeng, W. J. Freed, J. Crook, S. Abraham, A. Colman, P. Sartipy, S. Matsui, M. Carpenter, A. F. Gazdar, M. Rao, A. Chakravarti, *Nat. Genet.* **2005**, *37*, 1099-1103. *Genomic alterations in cultured human embryonic stem cells*.

[83] Q. L. Ying, M. Stavridis, D. Griffiths, M. Li, A. Smith, *Nat. Biotechnol.* **2003**, *21*, 183-186. *Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture*.

[84] H. Wu, Y. Zhang, *Cell* **2014**, *156*, 45-68. *Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions*.

[85] S. H. Feng, S. E. Jacobsen, W. Reik, *Science* **2010**, *330*, 622-627. *Epigenetic Reprogramming in Plant and Animal Development*.

[86] N. Rougier, D. Bourc'his, D. M. Gomes, A. Niveleau, M. Plachot, A. Paldi, E. Viegas-Pequignot, *Genes Dev.* **1998**, *12*, 2108-2113. *Chromosome methylation patterns during mammalian preimplantation development*.

[87] F. Santos, B. Hendrich, W. Reik, W. Dean, *Dev. Biol.* **2002**, *241*, 172-182. *Dynamic reprogramming of DNA methylation in the early mouse embryo*.

[88] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501-502. *Demethylation of the zygotic paternal genome*.

[89] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, *Curr. Biol.* **2000**, *10*, 475-478. *Active demethylation of the paternal genome in the mouse zygote*.

[90] H. Fulka, M. Mrazek, O. Tepla, J. Fulka, Jr., *Reproduction* **2004**, *128*, 703-708. *DNA methylation pattern in human zygotes and developing embryos*.

[91] T. P. Gu, F. Guo, H. Yang, H. P. Wu, G. F. Xu, W. Liu, Z. G. Xie, L. Shi, X. He, S. G. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabo, G. P. Pfeifer, J. Li, G. L. Xu, *Nature* **2011**, *477*, 606-610. *The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes*.

[92] A. Inoue, Y. Zhang, *Science* **2011**, *334*, 194. *Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos*.

[93] K. Iqbal, S.-G. Jin, G. P. Pfeifer, P. E. Szabó, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3642-3647. *Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine*.

[94] 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. *5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming*.

[95] F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T. P. Gu, B. Hu, C. P. Walsh, J. Li, F. Tang, G. L. Xu, *Cell Stem Cell* **2014**, *15*, 447-458. *Active and passive demethylation of male and female pronuclear DNA in the Mammalian zygote*.

[96] W. Dean, F. Santos, M. Stojkovic, V. Zakhartchenko, J. Walter, E. Wolf, W. Reik, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 13734-13738. *Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos*.

[97] M. Gehring, W. Reik, S. Henikoff, *Trends Genet.* **2009**, *25*, 82-90. *DNA demethylation by DNA repair*.

[98] M. A. Surani, K. Hayashi, P. Hajkova, *Cell* **2007**, *128*, 747-762. *Genetic and epigenetic regulators of pluripotency*.

[99] T. C. Petrossian, S. G. Clarke, *Mol. Cell. Proteomics* **2011**, *10*, DOI: 10.1074/mcp.M1110.000976. *Uncovering the human methyltransferasome*.

[100] M. Bostick, J. K. Kim, P.-O. Estève, A. Clark, S. Pradhan, S. E. Jacobsen, *Science* **2007**, *317*, 1760-1764. *UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells*.

[101] G. Pichler, P. Wolf, C. S. Schmidt, D. Meilinger, K. Schneider, C. Frauer, K. Fellinger, A. Rottach, H. Leonhardt, *J. Cell. Biochem.* **2011**, *112*, 2585-2593. *Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways*.

[102] J. Zhang, Q. Gao, P. Li, X. Liu, Y. Jia, W. Wu, J. Li, S. Dong, H. Koseki, J. Wong, *Cell Res.* **2011**, *21*, 1723-1739. *S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance*.

[103]   P. A. Defossez, I. Stancheva, *Prog. Mol. Biol. Transl. Sci.* **2011**, *101*, 377-398. *Biological functions of methyl-CpG-binding proteins.*

[104]   P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter, M. A. Surani, *Mech. Dev.* **2002**, *117*, 15-23. *Epigenetic reprogramming in mouse primordial germ cells.*

[105]   P. Hajkova, S. J. Jeffries, C. Lee, N. Miller, S. P. Jackson, M. A. Surani, *Science* **2010**, *329*, 78-82. *Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway.*

[106]   H. Sasaki, Y. Matsui, *Nat. Rev. Genet.* **2008**, *9*, 129-140. *Epigenetic events in mammalian germ-cell development: reprogramming and beyond.*

[107]   M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460-6468. *5-Hydroxymethylcytosine, the sixth base of the genome.*

[108]   C. Loenarz, C. J. Schofield, *Nat. Chem. Biol.* **2008**, *4*, 152-156. *Expanding chemical biology of 2-oxoglutarate oxygenases.*

[109]   J. M. Simmons, T. A. Muller, R. P. Hausinger, *Dalton Trans.* **2008**, 5132-5142. *Fe-II/alpha-ketoglutarate hydroxylases involved in nucleobase, nucleoside, nucleotide, and chromatin metabolism.*

[110]   C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580-583. *Oxygenase Catalyzed 5-Methylcytosine Hydroxylation.*

[111]   C. Loenarz, C. J. Schofield, *Trends Biochem. Sci.* **2011**, *36*, 7-18. *Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases.*

[112]   N. R. Rose, M. A. McDonough, O. N. King, A. Kawamura, C. J. Schofield, *Chem. Soc. Rev.* **2011**, *40*, 4364-4397. *Inhibition of 2-oxoglutarate dependent oxygenases.*

[113]   P. Borst, R. Sabatini, *Annu. Rev. Microbiol.* **2008**, *62*, 235-251. *Base J: discovery, biosynthesis, and possible functions.*

[114]   Z. Yu, P. A. Genest, B. ter Riet, K. Sweeney, C. DiPaolo, R. Kieft, E. Christodoulou, A. Perrakis, J. M. Simmons, R. P. Hausinger, H. G. van Luenen, D. J. Rigden, R. Sabatini, P. Borst, *Nucleic Acids Res.* **2007**, *35*, 2107-2115. *The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase.*

[115]   L. J. Cliffe, G. Hirsch, J. Wang, D. Ekanayake, W. Bullard, M. Hu, Y. Wang, R. Sabatini, *J. Biol. Chem.* **2012**, *287*, 19886-19895. *JBP1 and JBP2 proteins are Fe2+/2-oxoglutarate-dependent dioxygenases regulating hydroxylation of thymidine residues in trypanosome DNA.*

[116]   T. Heidebrecht, E. Christodoulou, M. J. Chalmers, S. Jan, B. ter Riet, R. K. Grover, R. P. Joosten, D. Littler, H. van Luenen, P. R. Griffin, P. Wentworth, P. Borst, A. Perrakis, *Nucleic Acids Res.* **2011**, *39*, 5715-5728. *The structural basis for recognition of base J containing DNA by a novel DNA binding domain in JBP1.*

[117]   S. Vainio, P. A. Genest, B. ter Riet, H. van Luenen, P. Borst, *Mol. Biochem. Parasitol.* **2009**, *164*, 157-161. *Evidence that J-binding protein 2 is a thymidine hydroxylase catalyzing the first step in the biosynthesis of DNA base J.*

[118]   L. M. Iyer, V. Anantharaman, M. Y. Wolf, L. Aravind, *Int. J. Parasitol.* **2008**, *38*, 1-31. *Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes.*

[119]   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.*

[120]   B. Yu, W. C. Edstrom, J. Benach, Y. Hamuro, P. C. Weber, B. R. Gibney, J. F. Hunt, *Nature* **2006**, *439*, 879-884. *Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB.*

[121]   T. W. Roy, A. S. Bhagwat, *Nucleic Acids Res.* **2007**, *35*, e147. *Kinetic studies of Escherichia coli AlkB using a new fluorescence-based assay for DNA demethylation.*

[122]   P. Koivisto, T. Duncan, T. Lindahl, B. Sedgwick, *J. Biol. Chem.* **2003**, *278*, 44348-44354. *Minimal methylated substrate and extended substrate range of Escherichia coli AlkB protein, a 1-methyladenine-DNA dioxygenase.*

[123]   R. D. Palmatier, R. P. McCroskey, M. T. Abbott, *J. Biol. Chem.* **1970**, *245*, 6706-6710. *Enzymatic Conversion of Uracil 5-Carboxylic Acid to Uracil and Carbon Dioxide.*

[124]   J. A. Smiley, J. M. Angelot, R. C. Cannon, E. M. Marshall, D. K. Asch, *Anal. Biochem.* **1999**, *266*, 85-92. *Radioactivity-based and spectrophotometric assays for isoorotate decarboxylase: Identification of the thymidine salvage pathway in lower eukaryotes.*

[125]   S. Xu, W. Li, J. Zhu, R. Wang, Z. Li, G.-L. Xu, J. Ding, *Cell Res.* **2013**, *23*, 1296-1309. *Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase.*

[126]   R. Ono, T. Taki, T. Taketani, M. Taniwaki, H. Kobayashi, Y. Hayashi, *Cancer Res.* **2002**, *62*, 4075-4080. *LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23).*

[127]   R. B. Lorsbach, J. Moore, S. Mathew, S. C. Raimondi, S. T. Mukatira, J. R. Downing, *Leukemia* **2003**, *17*, 637-641. *TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23).*

[128] L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang, Y. Xu, *Cell* **2013**, *155*, 1545-1555. *Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation.*

[129] H. Hashimoto, J. E. Pais, X. Zhang, L. Saleh, Z. Q. Fu, N. Dai, I. R. Correa, Y. Zheng, X. Cheng, *Nature* **2014**, *506*, 391-395. *Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA.*

[130] L. M. Iyer, S. Abhiman, L. Aravind, *Prog. Mol. Biol. Transl. Sci.* **2011**, *101*, 25-104. *Natural history of eukaryotic DNA methylation systems.*

[131] L. Tan, Y. G. Shi, *Development* **2012**, *139*, 1895-1902. *Tet family proteins and 5-hydroxymethylcytosine in development and disease.*

[132] R. P. Hausinger, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21-68. *FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes.*

[133] H. M. Hanauske-Abel, V. Günzler, *J. Theor. Biol.* **1982**, *94*, 421-455. *A stereochemical concept for the catalytic mechanism of prolylhydroxylase: applicability to classification and design of inhibitors.*

[134] Y. Fu, C. He, *Curr. Opin. Chem. Biol.* **2012**, *16*, 516-524. *Nucleic acid modifications with epigenetic significance.*

[135] A. L. Feig, S. J. Lippard, *Chem. Rev. (Washington, DC, U. S.)* **1994**, *94*, 759-805. *Reactions of Non-Heme Iron(II) Centers with Dioxygen in Biology and Chemistry.*

[136] A. K. Upadhyay, J. R. Horton, X. Zhang, X. Cheng, *Curr. Opin. Struct. Biol.* **2011**, *21*, 750-760. *Coordinated methyl-lysine erasure: structural and functional linkage of a Jumonji demethylase domain and a reader domain.*

[137] M. D. Allen, C. G. Grummitt, C. Hilcenko, S. Y. Min, L. M. Tonkin, C. M. Johnson, S. M. Freund, M. Bycroft, A. J. Warren, *EMBO J.* **2006**, *25*, 4503-4512. *Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase.*

[138] C. Frauer, A. Rottach, D. Meilinger, S. Bultmann, K. Fellinger, S. Hasenoder, M. Wang, W. Qin, J. Soding, F. Spada, H. Leonhardt, *PLoS One* **2011**, *6*, e16627. *Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1.*

[139] J. P. Thomson, P. J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A. R. Kerr, A. Deaton, R. Andrews, K. D. James, D. J. Turner, R. Illingworth, A. Bird, *Nature* **2010**, *464*, 1082-1086. *CpG islands influence chromatin structure via the CpG-binding protein Cfp1.*

[140] M. Ko, J. An, H. S. Bandukwala, L. Chavez, T. Aijo, W. A. Pastor, M. F. Segal, H. Li, K. P. Koh, H. Lahdesmaki, P. G. Hogan, L. Aravind, A. Rao, *Nature* **2013**, *497*, 122-126. *Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX.*

[141] K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343-348. *TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.*

[142] 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. *Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells.*

[143] Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Mol. Cell* **2011**, *42*, 451-464. *Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells.*

[144] H. Zhang, X. Zhang, E. Clark, M. Mulcahey, S. Huang, Y. G. Shi, *Cell Res.* **2010**, *20*, 1390-1393. *TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine.*

[145] Y. Xu, C. Xu, A. Kato, W. Tempel, J. G. Abreu, C. Bian, Y. Hu, D. Hu, B. Zhao, T. Cerovina, J. Diao, F. Wu, H. H. He, Q. Cui, E. Clark, C. Ma, A. Barbara, G. J. Veenstra, G. Xu, U. B. Kaiser, X. S. Liu, S. P. Sugrue, X. He, J. Min, Y. Kato, Y. G. Shi, *Cell* **2012**, *151*, 1200-1213. *Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development.*

[146] A. J. J. Straathof, J. L. L. Rakels, J. J. Heijnen, *Biocatalysis* **1992**, *7*, 13-27. *Kinetics of the Enzymatic Resolution of Racemic Compounds in Bi-Bi Reactions.*

[147] E. Holme, *Biochim. Biophys. Acta* **1982**, *707*, 259-266. *Determination of Intramolecular and Intermolecular Tritium Isotope Effects in the Reaction of Thymine 7-Hydroxylase.*

[148] E. Holme, S. Lindstedt, *Biochim. Biophys. Acta* **1982**, *704*, 278-283. *Studies on the Partial Reaction of Thymine 7-Hydroxylase in the Presence of 5-Fluorouracil.*

[149] L. D. Thornburg, M. T. Lai, J. S. Wishnok, J. Stubbe, *Biochemistry* **1993**, *32*, 14023-14033. *A Nonheme Iron Protein With Heme Tendencies - an Investigation of the Substrate-Specificity of Thymine Hydroxylase.*

[150] Z. Zhang, J. Ren, D. K. Stammers, J. E. Baldwin, K. Harlos, C. J. Schofield, *Nat. Struct. Biol.* **2000**, *7*, 127-133. *Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase.*

[151] J. Zhou, W. L. Kelly, B. O. Bachmann, M. Gunsior, C. A. Townsend, E. I. Solomon, *J. Am. Chem. Soc.* **2001**, *123*, 7388-7398. *Spectroscopic studies of substrate interactions with clavaminate synthase 2, a*

multifunctional alpha-KG-dependent non-heme iron enzyme: correlation with mechanisms and reactivities.

[152] C. Krebs, D. Galonic Fujimori, C. T. Walsh, J. M. Bollinger, Jr., *Acc. Chem. Res.* **2007**, *40*, 484-492. *Non-heme Fe(IV)-oxo intermediates.*

[153] J. H. Gommers-Ampt, P. Borst, *FASEB J.* **1995**, *9*, 1034-1042. *Hypermodified bases in DNA.*

[154] H. Hayatsu, M. Shiragami, *Biochemistry* **1979**, *18*, 632-637. *Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs.*

[155] R. L. Sinsheimer, *Science* **1954**, *120*, 551-553. *Nucleotides from T2r+ Bacteriophage.*

[156] E. Volkin, *J. Am. Chem. Soc.* **1954**, *76*, 5892-5893. *The Linkage of Glucose in Coliphage Nucleic Acids.*

[157] R. A. J. Warren, *Annu. Rev. Microbiol.* **1980**, *34*, 137-158. *Modified Bases in Bacteriophage DNAs.*

[158] G. R. Wyatt, S. S. Cohen, *Nature* **1952**, *170*, 1072-1073. *A New Pyrimidine Base from Bacteriophage Nucleic Acids.*

[159] G. R. Wyatt, S. S. Cohen, *Biochem. J.* **1953**, *55*, 774-782. *The Bases of the Nucleic Acids of Some Bacterial and Animal Viruses - the Occurence of 5-Hydroxymethylcytosine.*

[160] 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.*

[161] N. W. Penn, *Biochem. J.* **1976**, *155*, 709-712. *Modification of brain deoxyribonucleic acid base content with maturation in normal and malnourished rats.*

[162] R. M. Kothari, V. Shankar, *J. Mol. Evol.* **1976**, *7*, 325-329. *5-Methylcytosine Content in Vertebrate Deoxyribonucleic Acids - Species Specificity.*

[163] C. Bienvenu, J. R. Wagner, J. Cadet, *J. Am. Chem. Soc.* **1996**, *118*, 11406-11411. *Photosensitized oxidation of 5-methyl-2'-deoxycytidine by 2-methyl-1,4-naphthoquinone: Characterization of 5-(hydroperoxymethyl)-2'-deoxycytidine and stable methyl group oxidation products.*

[164] H. Cao, Y. Wang, *Nucleic Acids Res.* **2007**, *35*, 4833-4844. *Quantification of oxidative single-base and intrastrand cross-link lesions in unmethylated and CpG-methylated DNA induced by Fenton-type reagents.*

[165] M. Münzel, Doktorarbeit, Ludwig-Maximilians-Universität (München), **2011**. *Synthese natürlicher DNA Modifikationen und deren Detektion in DNA - Identifikation und Totalsynthese eines neuartigen DNA Photoschadens.*

[166] T. F. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Münzel, M. Müller, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schuller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, *131*, 1577-1590. *Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors.*

[167] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934-937. *Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution.*

[168] G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398-402. *Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation.*

[169] W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394-397. *Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells.*

[170] C. X. Song, T. A. Clark, X. Y. Lu, A. Kislyuk, Q. Dai, S. W. Turner, C. He, J. Korlach, *Nat. Methods* **2012**, *9*, 75-77. *Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine.*

[171] 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. *Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.*

[172] C.-X. Song, Y. Sun, Q. Dai, X.-Y. Lu, M. Yu, C.-G. Yang, C. He, *ChemBioChem* **2011**, *12*, 1682-1685. *Detection of 5-Hydroxymethylcytosine in DNA by Transferring a Keto-Glucose by Using T4 Phage β-Glucosyltransferase.*

[173] H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *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.*

[174] M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368-1380. *Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome.*

[175] Q. Chen, Y. Chen, C. Bian, R. Fujiki, X. Yu, *Nature* **2013**, *493*, 561-564. *TET2 promotes histone O-GlcNAcylation during gene transcription.*

[176] R. Deplus, B. Delatte, M. K. Schwinn, M. Defrance, J. Mendez, N. Murphy, M. A. Dawson, M. Volkmar, P. Putmans, E. Calonne, A. H. Shih, R. L. Levine, O. Bernard, T. Mercher, E. Solary, M. Urh, D. L. Daniels, F. Fuks, *EMBO J.* **2013**, *32*, 645-655. *TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS.*

[177] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129-1133. *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.*

[178] K. P. Koh, A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley, A. Rao, *Cell Stem Cell* **2011**, *8*, 200-213. *Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells.*

[179] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472-479. *TET enzymes, TDG and the dynamics of DNA demethylation.*

[180] Y. Wang, Y. Zhang, *Cell Rep.* **2014**, *6*, 278-284. *Regulation of TET protein stability by calpains.*

[181] T. Nakamura, Y. J. Liu, H. Nakashima, H. Umehara, K. Inoue, S. Matoba, M. Tachibana, A. Ogura, Y. Shinkai, T. Nakano, *Nature* **2012**, *486*, 415-419. *PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos.*

[182] H. Kobayashi, T. Sakurai, M. Imai, N. Takahashi, A. Fukuda, O. Yayoi, S. Sato, K. Nakabayashi, K. Hata, Y. Sotomaru, Y. Suzuki, T. Kono, *PLoS Genet.* **2012**, *8*, e1002440. *Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks.*

[183] M. M. Dawlaty, A. Breiling, T. Le, M. I. Barrasa, G. Raddatz, Q. Gao, B. E. Powell, A. W. Cheng, K. F. Faull, F. Lyko, R. Jaenisch, *Dev. Cell* **2014**, *29*, 102-111. *Loss of Tet enzymes compromises proper differentiation of embryonic stem cells.*

[184] M. M. Dawlaty, A. Breiling, T. Le, G. Raddatz, M. I. Barrasa, A. W. Cheng, Q. Gao, B. E. Powell, Z. Li, M. Xu, K. F. Faull, F. Lyko, R. Jaenisch, *Dev. Cell* **2013**, *24*, 310-323. *Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development.*

[185] M. M. Dawlaty, K. Ganz, B. E. Powell, Y. C. Hu, S. Markoulaki, A. W. Cheng, Q. Gao, J. Kim, S. W. Choi, D. C. Page, R. Jaenisch, *Cell Stem Cell* **2011**, *9*, 166-175. *Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development.*

[186] Z. Li, X. Cai, C. L. Cai, J. Wang, W. Zhang, B. E. Petersen, F. C. Yang, M. Xu, *Blood* **2011**, *118*, 4509-4518. *Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies.*

[187] W. A. Pastor, L. Aravind, A. Rao, *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 341-356. *TETonic shift: biological roles of TET proteins in DNA demethylation and transcription.*

[188] G. A. Kaas, C. Zhong, D. E. Eason, D. L. Ross, R. V. Vachhani, G. L. Ming, J. R. King, H. Song, J. D. Sweatt, *Neuron* **2013**, *79*, 1086-1093. *TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation.*

[189] H. Wu, Y. Zhang, *Genes Dev.* **2011**, *25*, 2436-2452. *Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation.*

[190] O. Yildirim, R. Li, J. H. Hung, P. B. Chen, X. Dong, L. S. Ee, Z. Weng, O. J. Rando, T. G. Fazzio, *Cell* **2011**, *147*, 1498-1510. *Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells.*

[191] P. Vella, A. Scelfo, S. Jammula, F. Chiacchiera, K. Williams, A. Cuomo, A. Roberto, J. Christensen, T. Bonaldi, K. Helin, D. Pasini, *Mol. Cell* **2013**, *49*, 645-656. *Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells.*

[192] K. Williams, J. Christensen, K. Helin, *EMBO Rep.* **2012**, *13*, 28-35. *DNA methylation: TET proteins-guardians of CpG islands?*

[193] A. Balasubramani, A. Rao, *Mol. Cell* **2013**, *49*, 618-619. *O-GlcNAcylation and 5-Methylcytosine Oxidation: An Unexpected Association between OGT and TETs.*

[194] B. Delatte, F. Fuks, *Briefings Funct. Genomics* **2013**, *12*, 191-204. *TET proteins: on the frenetic hunt for new cytosine modifications.*

[195] Y. Fardini, V. Dehennaut, T. Lefebvre, T. Issad, *Front. Endocrinol.* **2013**, *4*, 99. *O-GlcNAcylation: A New Cancer Hallmark?*

[196] R. Ito, S. Katsura, H. Shimada, H. Tsuchiya, M. Hada, T. Okumura, A. Sugawara, A. Yokoyama, *Genes Cells* **2014**, *19*, 52-65. *TET3-OGT interaction increases the stability and the presence of OGT in chromatin.*

[197] F. T. Shi, H. Kim, W. Lu, Q. He, D. Liu, M. A. Goodell, M. Wan, Z. Songyang, *J. Biol. Chem.* **2013**, *288*, 20776-20784. *Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells.*

[198] E. Solary, O. A. Bernard, A. Tefferi, F. Fuks, W. Vainchenker, *Leukemia* **2013**, *28*, 485-496. *The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases.*

[199] Q. Zhang, X. Liu, W. Gao, P. Li, J. Hou, J. Li, J. Wong, *J. Biol. Chem.* **2014**, *289*, 5986-5996. *Differential Regulation of Ten-Eleven Translocation Family of Dioxygenases by O-Linked beta-N-Acetylglucosamine Transferase OGT.*

[200] 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.*

[201] M. Mellén, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, *Cell* **2012**, *151*, 1417-1430. *MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System.*

[202] B. M. Colquitt, W. E. Allen, G. Barnea, S. Lomvardas, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 14682-14687. *Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity.*

[203] M. A. Hahn, R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S. G. Jin, Y. Jiang, G. P. Pfeifer, Q. Lu, *Cell Rep.* **2013**, *3*, 291-300. *Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis.*

[204] Y. Huang, L. Chavez, X. Chang, X. Wang, W. A. Pastor, J. Kang, J. A. Zepeda-Martinez, U. J. Pape, S. E. Jacobsen, B. Peters, A. Rao, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1361-1366. *Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells.*

[205] L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown, C. J. Schofield, *Chem. Commun.* **2014**, *50*, 1794-1796. *Structural insights into how 5-hydroxymethylation influences transcription factor binding.*

[206] J. Robertson, A. B. Robertson, A. Klungland, *Biochem. Biophys. Res. Commun.* **2011**, *411*, 40-43. *The presence of 5-hydroxymethylcytosine at the gene promoter and not in the gene body negatively regulates gene expression.*

[207] V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird, L. C. Sowers, *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. Iurlaro, G. Ficz, D. Oxley, E. A. Raiber, M. Bachman, M. Booth, S. Andrews, S. Balasubramanian, W. Reik, *Genome Biol.* **2013**, *14*, R119. *A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation.*

[209] H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang, X. Cheng, *Nucleic Acids Res.* **2012**, *40*, 4841-4849. *Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation.*

[210] K. E. Szulwach, X. Li, Y. Li, C. X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. Cheng, C. He, P. Jin, *Nat. Neurosci.* **2011**, *14*, 1607-1616. *5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging.*

[211] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res.* **2010**, *38*, e181. *Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA.*

[212] R. R. Zhang, Q. Y. Cui, K. Murai, Y. C. Lim, Z. D. Smith, S. Jin, P. Ye, L. Rosa, Y. K. Lee, H. P. Wu, W. Liu, Z. M. Xu, L. Yang, Y. Q. Ding, F. Tang, A. Meissner, C. Ding, Y. Shi, G. L. Xu, *Cell Stem Cell* **2013**, *13*, 237-245. *Tet1 Regulates Adult Hippocampal Neurogenesis and Cognition.*

[213] A. Rudenko, M. M. Dawlaty, J. Seo, A. W. Cheng, J. Meng, T. Le, K. F. Faull, R. Jaenisch, L. H. Tsai, *Neuron* **2013**, *79*, 1109-1122. *Tet1 is critical for neuronal activity-regulated gene expression and memory extinction.*

[214] S. Seisenberger, S. Andrews, F. Krueger, J. Arand, J. Walter, F. Santos, C. Popp, B. Thienpont, W. Dean, W. Reik, *Mol. Cell* **2012**, *48*, 849-862. *The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells.*

[215] J. Otani, H. Kimura, J. Sharif, T. A. Endo, Y. Mishima, T. Kawakami, H. Koseki, M. Shirakawa, I. Suetake, S. Tajima, *PLoS One* **2013**, *8*, e82961. *Cell Cycle-Dependent Turnover of 5-Hydroxymethyl Cytosine in Mouse Embryonic Stem Cells.*

[216] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 1049-1055. *5-Hydroxymethylcytosine is a predominantly stable DNA modification.*

[217] A. Eden, F. Gaudet, A. Waghmare, R. Jaenisch, *Science* **2003**, *300*, 455. *Chromosomal instability and tumors promoted by DNA hypomethylation.*

[218] S. B. Baylin, P. A. Jones, *Nat. Rev. Cancer* **2011**, *11*, 726-734. *A decade of exploring the cancer epigenome - biological and translational implications.*

[219] M. Esteller, *Nat. Rev. Genet.* **2007**, *8*, 286-298. *Cancer epigenomics: DNA methylomes and histone-modification maps.*

[220] M. Esteller, *N. Engl. J. Med.* **2008**, *358*, 1148-1159. *Epigenetics in cancer.*

[221] F. Gaudet, J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt, R. Jaenisch, *Science* **2003**, *300*, 489-492. *Induction of tumors in mice by genomic hypomethylation.*

[222] S. E. Goelz, B. Vogelstein, S. R. Hamilton, A. P. Feinberg, *Science* **1985**, *228*, 187-190. *Hypomethylation of DNA from benign and malignant human colon neoplasms.*

[223] P. A. Jones, S. B. Baylin, *Nat. Rev. Genet.* **2002**, *3*, 415-428. *The fundamental role of epigenetic events in cancer.*

[224] C. Lengauer, *Science* **2003**, *300*, 442-443. *Cancer. An unstable liaison.*

[225] C. B. Yoo, P. A. Jones, *Nat. Rev. Drug Discovery* **2006**, *5*, 37-50. *Epigenetic therapy of cancer: past, present and future.*

[226] T. K. Kelly, D. D. De Carvalho, P. A. Jones, *Nat. Biotechnol.* **2010**, *28*, 1069-1078. *Epigenetic modifications as therapeutic targets.*

[227] H. Yang, Y. Liu, F. Bai, J. Y. Zhang, S. H. Ma, J. Liu, Z. D. Xu, H. G. Zhu, Z. Q. Ling, D. Ye, K. L. Guan, Y. Xiong, *Oncogene* **2013**, *32*, 663-669. *Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation.*

[228] R. Uchiyama, H. Uhara, A. Uchiyama, E. Ogawa, Y. Takazawa, A. Ashida, H. Koga, K. Hayashi, Y. Kiniwa, R. Okuyama, *J. Dermatol. Sci.* **2014**, *73*, 161-163. *5-Hydroxymethylcytosine as a useful marker to differentiate between malignant melanomas and benign melanocytic nevi.*

[229] N. Konstandin, S. Bultmann, A. Szwagierczak, A. Dufour, B. Ksienzyk, F. Schneider, T. Herold, M. Mulaw, P. M. Kakadia, S. Schneider, K. Spiekermann, H. Leonhardt, S. K. Bohlander, *Leukemia* **2011**, *25*, 1649-1652. *Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia.*

[230] C. H. Hsu, K. L. Peng, M. L. Kang, Y. R. Chen, Y. C. Yang, C. H. Tsai, C. S. Chu, Y. M. Jeng, Y. T. Chen, F. M. Lin, H. D. Huang, Y. Y. Lu, Y. C. Teng, S. T. Lin, R. K. Lin, F. M. Tang, S. B. Lee, H. M. Hsu, J. C. Yu, P. W. Hsiao, L. J. Juan, *Cell Rep.* **2012**, *2*, 568-579. *TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases.*

[231] C. G. Lian, Y. Xu, C. Ceol, F. Wu, A. Larson, K. Dresser, W. Xu, L. Tan, Y. Hu, Q. Zhan, C. W. Lee, D. Hu, B. Q. Lian, S. Kleffel, Y. Yang, J. Neiswender, A. J. Khorasani, R. Fang, C. Lezcano, L. M. Duncan, R. A. Scolyer, J. F. Thompson, H. Kakavand, Y. Houvras, L. I. Zon, M. C. Mihm, Jr., U. B. Kaiser, T. Schatton, B. A. Woda, G. F. Murphy, Y. G. Shi, *Cell* **2012**, *150*, 1135-1146. *Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma.*

[232] W. Xu, H. Yang, Y. Liu, Y. Yang, P. Wang, S. H. Kim, S. Ito, C. Yang, P. Wang, M. T. Xiao, L. X. Liu, W. Q. Jiang, J. Liu, J. Y. Zhang, B. Wang, S. Frye, Y. Zhang, Y. H. Xu, Q. Y. Lei, K. L. Guan, S. M. Zhao, Y. Xiong, *Cancer Cell* **2011**, *19*, 17-30. *Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases.*

[233] W. Li, M. Liu, *J. Nucleic Acids* **2011**, *2011*, 870726. *Distribution of 5-hydroxymethylcytosine in different human tissues.*

[234] M. Ko, Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, A. Rao, *Nature* **2010**, *468*, 839-843. *Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2.*

[235] S. G. Jin, Y. Jiang, R. Qiu, T. A. Rauch, Y. Wang, G. Schackert, D. Krex, Q. Lu, G. P. Pfeifer, *Cancer Res.* **2011**, *71*, 7360-7365. *5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations.*

[236] M. C. Haffner, A. Chaux, A. K. Meeker, D. M. Esopi, J. Gerber, L. G. Pellakuru, A. Toubaji, P. Argani, C. Iacobuzio-Donahue, W. G. Nelson, G. J. Netto, A. M. De Marzo, S. Yegnasubramanian, *Oncotarget* **2011**, *2*, 627-637. *Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers.*

[237] M. E. Figueroa, O. Abdel-Wahab, C. Lu, P. S. Ward, J. Patel, A. Shih, Y. Li, N. Bhagwat, A. Vasanthakumar, H. F. Fernandez, M. S. Tallman, Z. Sun, K. Wolniak, J. K. Peeters, W. Liu, S. E. Choe, V. R. Fantin, E. Paietta, B. Lowenberg, J. D. Licht, L. A. Godley, R. Delwel, P. J. Valk, C. B. Thompson, R. L. Levine, A. Melnick, *Cancer Cell* **2010**, *18*, 553-567. *Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation.*

[238] S. J. Song, L. Poliseno, M. S. Song, U. Ala, K. Webster, C. Ng, G. Beringer, N. J. Brikbak, X. Yuan, L. C. Cantley, A. L. Richardson, P. P. Pandolfi, *Cell* **2013**, *154*, 311-324. *MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling.*

[239] K. Moran-Crusio, L. Reavie, A. Shih, O. Abdel-Wahab, D. Ndiaye-Lobry, C. Lobry, M. E. Figueroa, A. Vasanthakumar, J. Patel, X. Zhao, F. Perna, S. Pandey, J. Madzo, C. Song, Q. Dai, C. He, S. Ibrahim, M. Beran, J. Zavadil, S. D. Nimer, A. Melnick, L. A. Godley, I. Aifantis, R. L. Levine, *Cancer Cell* **2011**, *20*, 11-24. *Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation.*

[240] S. M. Langemeijer, R. P. Kuiper, M. Berends, R. Knops, M. G. Aslanyan, M. Massop, E. Stevens-Linders, P. van Hoogen, A. G. van Kessel, R. A. Raymakers, E. J. Kamping, G. E. Verhoef, E. Verburgh, A. Hagemeijer, P. Vandenberghe, T. de Witte, B. A. van der Reijden, J. H. Jansen, *Nat. Genet.* **2009**, *41*, 838-842. *Acquired mutations in TET2 are common in myelodysplastic syndromes.*

[241] F. Delhommeau, S. Dupont, V. Della Valle, C. James, S. Trannoy, A. Masse, O. Kosmider, J. P. Le Couedic, F. Robert, A. Alberdi, Y. Lecluse, I. Plo, F. J. Dreyfus, C. Marzac, N. Casadevall, C. Lacombe, S. P. Romana, P. Dessen, J. Soulier, F. Viguie, M. Fontenay, W. Vainchenker, O. A. Bernard, *N. Engl. J. Med.* **2009**, *360*, 2289-2301. *Mutation in TET2 in myeloid cancers.*

[242] C. Quivoron, L. Couronne, V. Della Valle, C. K. Lopez, I. Plo, O. Wagner-Ballon, M. Do Cruzeiro, F. Delhommeau, B. Arnulf, M. H. Stern, L. Godley, P. Opolon, H. Tilly, E. Solary, Y. Duffourd, P. Dessen, H. Merle-Beral, F. Nguyen-Khac, M. Fontenay, W. Vainchenker, C. Bastard, T. Mercher, O. A. Bernard, *Cancer Cell* **2011**, *20*, 25-38. *TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis.*

[243] L. Dang, D. W. White, S. Gross, B. D. Bennett, M. A. Bittinger, E. M. Driggers, V. R. Fantin, H. G. Jang, S. Jin, M. C. Keenan, K. M. Marks, R. M. Prins, P. S. Ward, K. E. Yen, L. M. Liau, J. D. Rabinowitz, L. C. Cantley, C. B. Thompson, M. G. Vander Heiden, S. M. Su, *Nature* **2009**, *462*, 739-744. *Cancer-associated IDH1 mutations produce 2-hydroxyglutarate.*

[244] P. S. Ward, J. Patel, D. R. Wise, O. Abdel-Wahab, B. D. Bennett, H. A. Coller, J. R. Cross, V. R. Fantin, C. V. Hedvat, A. E. Perl, J. D. Rabinowitz, M. Carroll, S. M. Su, K. A. Sharp, R. L. Levine, C. B. Thompson, *Cancer Cell* **2010**, *17*, 225-234. *The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate.*

[245] M. Xiao, H. Yang, W. Xu, S. Ma, H. Lin, H. Zhu, L. Liu, Y. Liu, C. Yang, Y. Xu, S. Zhao, D. Ye, Y. Xiong, K. L. Guan, *Genes Dev.* **2012**, *26*, 1326-1338. *Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors.*

[246] L. Lopez-Serra, M. Esteller, *Br. J. Cancer* **2008**, *98*, 1881-1885. *Proteins that bind methylated DNA and human cancer: reading the wrong words.*

[247] C. J. Burrows, *ACS Symp. Ser. Am. Chem. Soc.* **2009**, *2009*, 147-156. *Surviving an Oxygen Atmosphere: DNA Damage and Repair.*

[248] J. E. Klaunig, L. M. Kamendulis, *Annu. Rev. Pharmacol. Toxicol.* **2004**, *44*, 239-267. *THE ROLE OF OXIDATIVE STRESS IN CARCINOGENESIS.*

[249] J. Cadet, S. Loft, R. Olinski, M. D. Evans, K. Bialkowski, J. Richard Wagner, P. C. Dedon, P. Moller, M. M. Greenberg, M. S. Cooke, *Free Radic. Res.* **2012**, *46*, 367-381. *Biologically relevant oxidants and terminology, classification and nomenclature of oxidatively generated damage to nucleobases and 2-deoxyribose in nucleic acids.*

[250] S. Zuo, R. J. Boorstein, G. W. Teebor, *Nucleic Acids Res.* **1995**, *23*, 3239-3243. *Oxidative damage to 5-methylcytosine in DNA.*

[251] G. D. Castro, C. J. Stamato, J. A. Castro, *Free Radic. Biol. Med.* **1994**, *17*, 419-428. *5-methylcytosine attack by free radicals arising from bromotrichloromethane in a model system: structures of reaction products.*

[252] D. B. Bucher, B. M. Pilles, T. Pfaffeneder, T. Carell, W. Zinth, *ChemPhysChem* **2014**, *15*, 420–423. *Fingerprinting DNA Oxidation Processes: IR Characterization of the 5-Methyl-2'-Deoxycytidine Radical Cation.*

[253] E. Privat, L. C. Sowers, *Chem. Res. Toxicol.* **1996**, *9*, 745-750. *Photochemical deamination and demethylation of 5-methylcytosine.*

[254] H. Yamada, Y. Kitauchi, K. Tanabe, T. Ito, S. I. Nishimoto, *Chemistry* **2011**, *17*, 2225-2235. *Anthraquinone-Sensitized Photooxidation of 5-Methylcytosine in DNA Leading to Piperidine-Induced Efficient Strand Cleavage.*

[255] G. S. Madugundu, J. Cadet, J. R. Wagner, *Nucleic Acids Res.* **2014**, *42*, 7450-7460. *Hydroxyl-radical-induced oxidation of 5-methylcytosine in isolated and cellular DNA.*

[256] J. Cadet, J. R. Wagner, *Mutat. Res., Genet. Toxicol. Environ. Mutagen.* **2014**, *764-765*, 18-35. *TET enzymatic oxidation of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine.*

[257] J. R. Wagner, J. Cadet, *Acc. Chem. Res.* **2010**, *43*, 564-571. *Oxidation Reactions of Cytosine DNA Components by Hydroxyl Radical and One-Electron Oxidants in Aerated Aqueous Solutions.*

[258] H. Hong, H. Cao, Y. Wang, *Chem. Res. Toxicol.* **2006**, *19*, 614-621. *Identification and quantification of a guanine-thymine intrastrand cross-link lesion induced by Cu(II)/H2O2/ascorbate.*

[259] G. Boysen, L. B. Collins, S. Liao, A. M. Luke, B. F. Pachkowski, J. L. Watters, J. A. Swenberg, *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2010**, *878*, 375-380. *Analysis of 8-oxo-7,8-dihydro-2'-deoxyguanosine by ultra high pressure liquid chromatography-heat assisted electrospray ionization-tandem mass spectrometry.*

[260] F. Samson-Thibault, G. S. Madugundu, S. Gao, J. Cadet, J. R. Wagner, *Chem. Res. Toxicol.* **2012**, *25*, 1902-1911. *Profiling cytosine oxidation in DNA by LC-MS/MS.*

[261] E. Zarakowska, D. Gackowski, M. Foksinski, R. Olinski, *Mutat. Res., Genet. Toxicol. Environ. Mutagen.* **2014**, *764-765*, 58-63. *Are 8-oxoguanine (8-oxoGua) and 5-hydroxymethyluracil (5-hmUra) oxidatively damaged DNA bases or transcription (epigenetic) marks?*

[262] J. L. Ravanat, T. Douki, P. Duez, E. Gremaud, K. Herbert, T. Hofer, L. Lasserre, C. Saint-Pierre, A. Favier, J. Cadet, *Carcinogenesis* **2002**, *23*, 1911-1918. *Cellular background level of 8-oxo-7,8-dihydro-2'-deoxyguanosine: an isotope based method to evaluate artefactual oxidation of DNA during its extraction and subsequent work-up.*

[263] K. Taghizadeh, J. L. McFaline, B. Pang, M. Sullivan, M. Dong, E. Plummer, P. C. Dedon, *Nat. Protoc.* **2008**, *3*, 1287-1298. *Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry.*

[264] A. Klungland, R. Paulsen, V. Rolseth, Y. Yamada, Y. Ueno, P. Wiik, A. Matsuda, E. Seeberg, S. Bjelland, *Toxicol. Lett.* **2001**, *119*, 71-78. *5-Formyluracil and its nucleoside derivatives confer toxicity and mutagenicity to mammalian cells by interfering with normal RNA and DNA metabolism.*

[265] D. D. Levy, G. W. Teebor, *Nucleic Acids Res.* **1991**, *19*, 3337-3343. *Site directed substitution of 5-hydroxymethyluracil for thymine in replicating phi X-174am3 DNA via synthesis of 5-hydroxymethyl-2'-deoxyuridine-5'-triphosphate.*

[266] R. J. Boorstein, L. N. Chiu, G. W. Teebor, *Mol. Cell. Biol.* **1992**, *12*, 5536-5540. *A mammalian cell line deficient in activity of the DNA repair enzyme 5-hydroxymethyluracil-DNA glycosylase is resistant to the toxic effects of the thymidine analog 5-hydroxymethyl-2'-deoxyuridine.*

[267] W. Chaung, R. J. Boorstein, *Mutat. Res.* **1997**, *373*, 125-137. *Molecular spectrum of mutations induced by 5-hydroxymethyl-2'-deoxyuridine in (CHO)-PL61 cells.*

[268] K. Kemmerich, F. A. Dingler, C. Rada, M. S. Neuberger, *Nucleic Acids Res.* **2012**, *40*, 6016-6025. *Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung-/-Msh2-/- mice.*

[269] D. K. Rogstad, A. Darwanto, J. L. Herring, K. N. Rogstad, A. Burdzy, S. R. Hadley, J. W. Neidigh, L. C. Sowers, *Chem. Res. Toxicol.* **2007**, *20*, 1787-1796. *Measurement of the incorporation and repair of exogenous 5-hydroxymethyl-2'-deoxyuridine in human cells in culture using gas chromatography-negative chemical ionization-mass spectrometry.*

[270] S. Waschke, J. Reefschlager, D. Barwolff, P. Langen, *Nature* **1975**, *255*, 629-630. *5-hydroxymethyl-2'-deoxyuridine, a normal DNA constituent in certain Bacillus subtilis phages is cytostatic for mammalian cells.*

[271] A. Grove, A. Galeone, L. Mayol, E. P. Geiduschek, *J. Mol. Biol.* **1996**, *260*, 196-206. *On the connection between inherent DNA flexure and preferred binding of hydroxymethyluracil-containing DNA by the type II DNA-binding protein TF1.*

[272] A. Grove, A. Galeone, E. Yu, L. Mayol, E. P. Geiduschek, *J. Mol. Biol.* **1998**, *282*, 731-739. *Affinity, stability and polarity of binding of the TATA binding protein governed by flexure at the TATA Box.*

[273] G. A. Kassavetis, A. Kumar, G. A. Letts, E. P. Geiduschek, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 9196-9201. *A post-recruitment function for the RNA polymerase III transcription-initiation factor IIIB.*

[274] K. W. Plaxco, W. A. Goddard, 3rd, *Biochemistry* **1994**, *33*, 3050-3054. *Contributions of the thymine methyl group to the specific recognition of poly- and mononucleotides: an analysis of the relative free energies of solvation of thymine and uracil.*

[275] D. K. Rogstad, P. Liu, A. Burdzy, S. S. Lin, L. C. Sowers, *Biochemistry* **2002**, *41*, 8093-8102. *Endogenous DNA lesions can inhibit the binding of the AP-1 (c-Jun) transcription factor.*

[276] A. Grove, M. L. Figueiredo, A. Galeone, L. Mayol, E. P. Geiduschek, *J. Biol. Chem.* **1997**, *272*, 13084-13087. *Twin hydroxymethyluracil-A base pair steps define the binding site for the DNA-binding protein TF1.*

[277] S. Amente, A. Bertoni, A. Morano, L. Lania, E. V. Avvedimento, B. Majello, *Oncogene* **2010**, *29*, 3691-3702. *LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription.*

[278] S. Amente, L. Lania, E. V. Avvedimento, B. Majello, *Cell Cycle* **2010**, *9*, 3002-3004. *DNA oxidation drives Myc mediated transcription.*

[279] M. N. Gillespie, V. Pastukh, M. V. Ruchko, *Ann. N. Y. Acad. Sci.* **2009**, *1177*, 140-150. *Oxidative DNA modifications in hypoxic signaling.*

[280] B. Perillo, M. N. Ombra, A. Bertoni, C. Cuozzo, S. Sacchetti, A. Sasso, L. Chiariotti, A. Malorni, C. Abbondanza, E. V. Avvedimento, *Science* **2008**, *319*, 202-206. *DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression.*

[281] F. Forneris, C. Binda, E. Battaglioli, A. Mattevi, *Trends Biochem. Sci.* **2008**, *33*, 181-189. *LSD1: oxidative chemistry for multifaceted functions in chromatin regulation.*

[282] S. P. Moore, K. J. Toomire, P. R. Strauss, *DNA Repair* **2013**, *12*, 1152-1158. *DNA modifications repaired by base excision repair are epigenetic.*

[283] J. K. Zhu, *Annu. Rev. Genet.* **2009**, *43*, 143-166. *Active DNA demethylation mediated by DNA glycosylases.*

[284] C. A. Miller, J. D. Sweatt, *Neuron* **2007**, *53*, 857-869. *Covalent modification of DNA regulates memory formation.*

[285] K. Martinowich, D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. Fan, Y. E. Sun, *Science* **2003**, *302*, 890-893. *DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation.*

[286] D. K. Ma, M. C. Marchetto, J. U. Guo, G. L. Ming, F. H. Gage, H. Song, *Nat. Neurosci.* **2010**, *13*, 1338-1344. *Epigenetic choreographers of neurogenesis in the adult mammalian brain.*

[287] D. K. Ma, M. H. Jang, J. U. Guo, Y. Kitabatake, M. L. Chang, N. Pow-Anpongkul, R. A. Flavell, B. Lu, G. L. Ming, H. Song, *Science* **2009**, *323*, 1074-1077. *Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis.*

[288] J. J. Day, J. D. Sweatt, *Nat. Neurosci.* **2010**, *13*, 1319-1323. *DNA methylation and memory formation.*

[289] M. Gehring, K. L. Bubb, S. Henikoff, *Science* **2009**, *324*, 1447-1451. *Extensive demethylation of repetitive elements during seed development underlies gene imprinting.*

[290] M. Gehring, J. H. Huh, T. F. Hsieh, J. Penterman, Y. Choi, J. J. Harada, R. B. Goldberg, R. L. Fischer, *Cell* **2006**, *124*, 495-506. *DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation.*

[291] J. Penterman, D. Zilberman, J. H. Huh, T. Ballinger, S. Henikoff, R. L. Fischer, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 6752-6757. *DNA demethylation in the Arabidopsis genome.*

[292] A. P. Ortega-Galisteo, T. Morales-Ruiz, R. R. Ariza, T. Roldan-Arjona, *Plant Mol. Biol.* **2008**, *67*, 671-681. *Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks.*

[293] M. Fremont, M. Siegmann, S. Gaulis, R. Matthies, D. Hess, J. P. Jost, *Nucleic Acids Res.* **1997**, *25*, 2375-2380. *Demethylation of DNA by purified chick embryo 5-methylcytosine-DNA glycosylase requires both protein and RNA.*

[294] J. P. Jost, S. Schwarz, D. Hess, H. Angliker, F. V. Fuller-Pace, H. Stahl, S. Thiry, M. Siegmann, *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.*

[295] C. B. Millar, J. Guy, O. J. Sansom, J. Selfridge, E. MacDougall, B. Hendrich, P. D. Keightley, S. M. Bishop, A. R. Clarke, A. Bird, *Science* **2002**, *297*, 403-405. *Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice.*

[296] B. Zhu, Y. Zheng, H. Angliker, S. Schwarz, S. Thiry, M. Siegmann, J. P. Jost, *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.*

[297] B. Zhu, Y. Zheng, D. Hess, H. Angliker, S. Schwarz, M. Siegmann, S. Thiry, J. P. Jost, *Proc. Natl. Acad. Sci. U. S. A.* **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.*

[298] 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.*

[299] R. Metivier, 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. *Cyclical DNA methylation of a transcriptionally active promoter.*

[300] 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.*

[301] N. Navaratnam, J. R. Morrison, S. Bhattacharya, D. Patel, T. Funahashi, F. Giannoni, B. B. Teng, N. O. Davidson, J. Scott, *J. Biol. Chem.* **1993**, *268*, 20709-20712. *The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase.*

[302] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101-1105. *Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency.*

[303] K. Rai, I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, B. R. Cairns, *Cell* **2008**, *135*, 1201-1212. *DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45.*

[304] B. Teng, C. F. Burant, N. O. Davidson, *Science* **1993**, *260*, 1816-1819. *Molecular cloning of an apolipoprotein B messenger RNA editing protein.*

[305] K. Hirano, S. G. Young, R. V. Farese, Jr., J. Ng, E. Sande, C. Warburton, L. M. Powell-Braxton, N. O. Davidson, *J. Biol. Chem.* **1996**, *271*, 9887-9890. *Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48.*

[306] J. R. Morrison, C. Paszty, M. E. Stevens, S. D. Hughes, T. Forte, J. Scott, E. M. Rubin, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 7154-7159. *Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism.*

[307]   M. Muramatsu, K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, T. Honjo, *Cell* **2000**, *102*, 553-563. *Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme*.

[308]   P. Revy, T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A. G. Ugazio, N. Brousse, M. Muramatsu, L. D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer, A. Durandy, *Cell* **2000**, *102*, 565-575. *Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)*.

[309]   Z. Liutkeviciute, G. Lukinavicius, V. Masevicius, D. Daujotyte, S. Klimasauskas, *Nat. Chem. Biol.* **2009**, *5*, 400-402. *Cytosine-5-methyltransferases add aldehydes to DNA*.

[310]   Z. Liutkeviciute, E. Kriukiene, I. Grigaityte, V. Masevicius, S. Klimasauskas, *Angew. Chem. Int. Ed.* **2011**, *50*, 2090-2093. *Methyltransferase-Directed Derivatization of 5-Hydroxymethylcytosine in DNA*.

[311]   S. V. Cannon, A. Cummings, G. W. Teebor, *Biochem. Biophys. Res. Commun.* **1988**, *151*, 1173-1179. *5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue*.

[312]   S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67-79. *Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair*.

[313]   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*.

[314]   J. F. Mouret, M. Polverelli, F. Sarrazini, J. Cadet, *Chem Biol Interact* **1991**, *77*, 187-201. *Ionic and radical oxidations of DNA by hydrogen peroxide*.

[315]   M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Scholer, J. Walter, *EMBO J.* **2010**, *29*, 1877-1888. *Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes*.

[316]   E. A. Raiber, D. Beraldi, G. Ficz, H. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69. *Genome-wide distribution of 5-formylcytosine in ES cells is associated with transcription and depends on thymine DNA glycosylase*.

[317]   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*.

[318]   L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G. L. Xu, C. Luo, H. Jiang, C. He, *Nat. Chem. Biol.* **2012**, *8*, 328-330. *Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA*.

[319]   J. Walter, *Cell Stem Cell* **2011**, *8*, 121-122. *An Epigenetic Tet a Tet with Pluripotency*.

[320]   A. L. Jacobs, P. Schär, *Chromosoma* **2012**, *121*, 1-20. *DNA glycosylases: in DNA repair and beyond*.

[321]   K. M. Dickson, C. B. Gustafson, J. I. Young, S. Züchner, G. Wang, *Biochem. Biophys. Res. Commun.* **2013**, *439*, 522-527. *Ascorbate-induced generation of 5-hydroxymethylcytosine is unaffected by varying levels of iron and 2-oxoglutarate*.

[322]   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*.

[323]   R. Yin, S.-Q. Mao, B. Zhao, Z. Chong, Y. Yang, C. Zhao, D. Zhang, H. Huang, J. Gao, Z. Li, Y. Jiao, C. Li, S. Liu, D. Wu, W. Gu, Y.-G. Yang, G.-L. Xu, H. Wang, *J. Am. Chem. Soc.* **2013**, *135*, 10396-10403. *Ascorbic Acid Enhances Tet-Mediated 5-Methylcytosine Oxidation and Promotes DNA Demethylation in Mammals*.

[324]   C. S. Chen, Y. Fujimoto, G. Girdaukas, C. J. Sih, *J. Am. Chem. Soc.* **1982**, *104*, 7294-7299. *Quantitative-Analyses of Biochemical Kinetic Resolutions of Enantiomers*.

[325]   A. J. J. Straathof, J. A. Jongejan, *Enzyme Microb. Technol.* **1997**, *21*, 559-571. *The enantiomeric ratio: origin, determination and prediction*.

[326]   C. S. Chen, C. J. Sih, *Angew. Chem.* **1989**, *101*, 711-724. *Enantioselektive Biokatalyse in organischen Solventien am Beispiel Lipase-katalysierter Reaktionen*.

[327]   E. A. Barnsley, *Biochem. J.* **1993**, *291*, 323. *Specificity constants for two-substrate reactions*.

[328]   J. L. Rakels, A. J. Straathof, J. J. Heijnen, *Enzyme Microb. Technol.* **1993**, *15*, 1051-1056. *A simple method to determine the enantiomeric ratio in enantioselective biocatalysis*.

[329]   J. L. L. Rakels, P. Caillat, A. J. J. Straathof, J. J. Heijnen, *Biotechnol. Prog.* **1994**, *10*, 403-409. *Modification of the Enzyme Enantioselectivity by Product Inhibition*.

[330]   S. Kizaki, H. Sugiyama, *Org. Biomol. Chem.* **2014**, *12*, 104-107. *CGmCGCG is a versatile substrate with which to evaluate Tet protein activity*.

[331]   Z. Liutkevičiūtė, E. Kriukienė, J. Ličytė, M. Rudytė, G. Urbanavičiūtė, S. Klimašauskas, *J. Am. Chem. Soc.* **2014**, *136*, 5884-5887. *Direct Decarboxylation of 5-Carboxylcytosine by DNA C5-Methyltransferases*.

[332]   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*.

[333]   C. Brandmayr, Doktorarbeit, Ludwig-Maximilians-Universität (München), **2014**. *Investigation of the role and regulation of modified DNA and RNA nucleosides*.

[334]   A. Liu, H. Zhang, *Biochemistry* **2006**, *45*, 10407-10411. *Transition metal-catalyzed nonoxidative decarboxylation reactions*.

[335]   D. Martynowski, Y. Eyobo, T. Li, K. Yang, A. Liu, H. Zhang, *Biochemistry* **2006**, *45*, 10412-10421. *Crystal structure of alpha-amino-beta-carboxymuconate-epsilon-semialdehyde decarboxylase: insight into the active site and catalytic mechanism of a novel decarboxylation reaction*.

[336]   T. W. Nilsen, D. L. Wood, C. Baglioni, *J. Biol. Chem.* **1982**, *257*, 1602-1605. *Presence of 2',5'-oligo(A) and of enzymes that synthesize, bind, and degrade 2',5'-oligo(A) in HeLa cell nuclei*.

[337]   W. Yang, *Nucleus* **2013**, *4*, 166-175. *Distinct, but not completely separate spatial transport routes in the nuclear pore complex*.

[338]   J. E. Hagstrom, J. J. Ludtke, M. C. Bassik, M. G. Sebestyen, S. A. Adam, J. A. Wolff, *J. Cell Sci.* **1997**, *110 (Pt 18)*, 2323-2331. *Nuclear import of DNA in digitonin-permeabilized cells*.

[339]   Z. Zhang, W.-S. Wu, *Stem Cells Dev.* **2013**, *22*, 2268-2277. *Sodium Butyrate Promotes Generation of Human Induced Pluripotent Stem Cells Through Induction of the miR302/367 Cluster*.

[340]   J. Grunberg, K. Knogler, R. Waibel, I. Novak-Hofer, *BioTechniques* **2003**, *34*, 968-972. *High-yield production of recombinant antibody fragments in HEK-293 cells using sodium butyrate*.

[341]   J. C. Shen, E. J. Fox, E. H. Ahn, L. A. Loeb, *Sci. Rep.* **2014**, *4*, DOI: 10.1038/srep04894. *A Rapid Assay for Measuring Nucleotide Excision Repair by Oligonucleotide Retrieval*.

[342]   J. D. Dignam, *Method. Enzymol.* **1990**, *182*, 194-203. *Preparation of extracts from higher eukaryotes*.

[343]   J. D. Dignam, R. M. Lebovitz, R. G. Roeder, *Nucleic Acids Res.* **1983**, *11*, 1475-1489. *Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei*.

[344]   E. Kriukiene, Z. Liutkeviciute, S. Klimasauskas, *Chem. Soc. Rev.* **2012**, *41*, 6916-6930. *5-Hydroxymethylcytosine - the elusive epigenetic mark in mammalian DNA*.

[345]   K. M. Schermerhorn, S. Delaney, *Acc. Chem. Res.* **2014**, *47*, 1238-1246. *A Chemical and Kinetic Perspective on Base Excision Repair of DNA*.

[346]   U. Müller, C. Bauer, M. Siegl, A. Rottach, H. Leonhardt, *Nucleic Acids Res.* **2014**, *42*, 8592-8604. *TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation*.

[347]   A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, A. C. Drohat, *J. Am. Chem. Soc.* **2013**, *135*, 15813-15822. *Divergent mechanisms for enzymatic excision of 5-formylcytosine and 5-carboxylcytosine from DNA*.

[348]   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*.

[349]   A. Maiti, A. C. Drohat, *DNA Repair* **2011**, *10*, 545-553. *Dependence of substrate binding and catalysis on pH, ionic strength, and temperature for thymine DNA glycosylase: Insights into recognition and processing of G.T mispairs*.

[350]   D. Cortazar, C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schuermann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird, P. Schar, *Nature* **2011**, *470*, 419-423. *Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability*.

[351]   A. B. Robertson, A. Klungland, T. Rognes, I. Leiros, *Cell. Mol. Life Sci.* **2009**, *66*, 981-993. *DNA repair in mammalian cells: Base excision repair: the long and short of it*.

[352]   K. M. Kroeger, Y. L. Jiang, Y. W. Kow, M. F. Goodman, M. M. Greenberg, *Biochemistry* **2004**, *43*, 6723-6733. *Mutagenic effects of 2-deoxyribonolactone in Escherichia coli. An abasic lesion that disobeys the A-rule*.

[353]   K. M. Kroeger, J. Kim, M. F. Goodman, M. M. Greenberg, *Biochemistry* **2004**, *43*, 13621-13627. *Effects of the C4'-oxidized abasic site on replication in Escherichia coli. An unusually large deletion is induced by a small lesion*.

[354]   K. M. Kroeger, J. Kim, M. F. Goodman, M. M. Greenberg, *Biochemistry* **2006**, *45*, 5048-5056. *Replication of an oxidized abasic site in Escherichia coli by a dNTP-stabilized misalignment mechanism that reads upstream and downstream nucleotides*.

[355]   L. A. Loeb, B. D. Preston, *Annu. Rev. Genet.* **1986**, *20*, 201-230. *Mutagenesis by apurinic/apyrimidinic sites*.

[356]   K. M. Kroeger, M. F. Goodman, M. M. Greenberg, *Nucleic Acids Res.* **2004**, *32*, 5480-5485. *A comprehensive comparison of DNA replication past 2-deoxyribose and its tetrahydrofuran analog in Escherichia coli*.

[357] F. Santos, J. Peat, H. Burgess, C. Rada, W. Reik, W. Dean, *Epigenet. Chromatin* **2013**, *6*, 39. *Active demethylation in mouse zygotes involves cytosine deamination and base excision repair*.

[358] C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, R. M. Kohli, *Nat. Chem. Biol.* **2012**, *8*, 751-758. *AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation*.

[359] G. Rangam, K. M. Schmitz, A. J. Cobb, S. K. Petersen-Mahrt, *PLoS One* **2012**, *7*, e43279. *AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud*.

[360] N. Bhutani, J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel, H. M. Blau, *Nature* **2010**, *463*, 1042-1047. *Reprogramming towards pluripotency requires AID-dependent DNA demethylation*.

[361] S. Kangaspeska, B. Stride, R. Metivier, M. Polycarpou-Schwarz, D. Ibberson, R. P. Carmouche, V. Benes, F. Gannon, G. Reid, *Nature* **2008**, *452*, 112-115. *Transient cyclical methylation of promoter DNA*.

[362] M. M. Greenberg, *Acc. Chem. Res.* **2014**, *47*, 646-655. *Abasic and Oxidized Abasic Site Reactivity in DNA: Enzyme Inhibition, Cross-Linking, and Nucleosome Catalyzed Reactions*.

[363] Y. Liu, R. Prasad, W. A. Beard, P. S. Kedar, E. W. Hou, D. D. Shock, S. H. Wilson, *J. Biol. Chem.* **2007**, *282*, 13532-13541. *Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta*.

[364] M. Talpaert-Borle, M. Liuzzi, *Biochim. Biophys. Acta* **1983**, *740*, 410-416. *Reaction of apurinic/apyrimidinic sites with [14C]methoxyamine. A method for the quantitative assay of AP sites in DNA*.

[365] X. Zhou, R. G. Liberman, P. L. Skipper, Y. Margolin, S. R. Tannenbaum, P. C. Dedon, *Anal. Biochem.* **2005**, *343*, 84-92. *Quantification of DNA strand breaks and abasic sites by oxime derivatization and accelerator mass spectrometry: Application to γ-radiation and peroxynitrite*.

[366] B. X. Chen, K. Kubo, H. Ide, B. F. Erlanger, S. S. Wallace, Y. W. Kow, *Mutat. Res.* **1992**, *273*, 253-261. *Properties of a monoclonal antibody for the detection of abasic sites, a common DNA lesion*.

[367] K. Kubo, H. Ide, S. S. Wallace, Y. W. Kow, *Biochemistry* **1992**, *31*, 3703-3708. *A novel, sensitive, and specific assay for abasic sites, the most commonly produced DNA lesion*.

[368] H. Ide, K. Akamatsu, Y. Kimura, K. Michiue, K. Makino, A. Asaeda, Y. Takamori, K. Kubo, *Biochemistry* **1993**, *32*, 8276-8283. *Synthesis and damage specificity of a novel probe for the detection of abasic sites in DNA*.

[369] J. Nakamura, V. E. Walker, P. B. Upton, S. Y. Chiang, Y. W. Kow, J. A. Swenberg, *Cancer Res.* **1998**, *58*, 222-225. *Highly sensitive apurinic/apyrimidinic site assay can detect spontaneous and chemically induced depurination under physiological conditions*.

[370] J. Nakamura, J. A. Swenberg, *Cancer Res.* **1999**, *59*, 2522-2526. *Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues*.

[371] H. Atamna, I. Cheung, B. N. Ames, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 686-691. *A method for detecting abasic sites in living cells: age-dependent changes in base excision repair*.

[372] N. Kojima, T. Takebayashi, A. Mikami, E. Ohtsuka, Y. Komatsu, *J. Am. Chem. Soc.* **2009**, *131*, 13208-13209. *Construction of highly reactive probes for abasic site detection by introduction of an aromatic and a guanidine residue into an aminooxy group*.

[373] K. P. Roberts, J. A. Sobrino, J. Payton, L. B. Mason, R. J. Turesky, *Chem. Res. Toxicol.* **2006**, *19*, 300-309. *Determination of Apurinic/Apyrimidinic Lesions in DNA with High-Performance Liquid Chromatography and Tandem Mass Spectrometry*.

[374] T. Lindahl, *Nature* **1993**, *362*, 709-715. *Instability and decay of the primary structure of DNA*.

[375] P. Fattorini, G. Marrubini, S. Sorcaburu-Cigliero, P. Pitacco, P. Grignani, C. Previdere, *Electrophoresis* **2011**, *32*, 3042-3052. *CE analysis and molecular characterisation of depurinated DNA samples*.

[376] T. Lindahl, O. Karlstrom, *Biochemistry* **1973**, *12*, 5151-5154. *Heat-induced depyrimidination of deoxyribonucleic acid in neutral solution*.

[377] T. Lindahl, B. Nyberg, *Biochemistry* **1972**, *11*, 3610-3618. *Rate of depurination of native deoxyribonucleic acid*.

[378] S. Greer, S. Zamenhof, *J. Mol. Biol.* **1962**, *4*, 123-141. *Studies on depurination of DNA by heat*.

[379] L. A. Loeb, *Cancer Res.* **1989**, *49*, 5489-5496. *Endogenous carcinogenesis: molecular oncology into the twenty-first century--presidential address*.

[380] D. M. Wilson, 3rd, D. Barsky, *Mutat. Res.* **2001**, *485*, 283-307. *The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA*.

[381] P. J. McHugh, J. Knowland, *Nucleic Acids Res.* **1995**, *23*, 1664-1670. *Novel reagents for chemical cleavage at abasic sites and UV photoproducts in DNA*.

[382] H. Sugiyama, T. Fujiwara, A. Ura, T. Tashiro, K. Yamamoto, S. Kawanishi, I. Saito, *Chem. Res. Toxicol.* **1994**, *7*, 673-683. *Chemistry of thermal degradation of abasic sites in DNA. Mechanistic investigation on thermal DNA strand cleavage of alkylated DNA*.

[383]   C. Zhou, J. T. Sczepanski, M. M. Greenberg, *J. Am. Chem. Soc.* **2012**, *134*, 16734-16741. *Mechanistic studies on histone catalyzed cleavage of apyrimidinic/apurinic sites in nucleosome core particles.*

[384]   T. Lindahl, A. Andersson, *Biochemistry* **1972**, *11*, 3618-3623. *Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid.*

[385]   S. E. Bennett, J. Kitner, *Nucleosides, Nucleotides Nucleic Acids* **2006**, *25*, 823-842. *Characterization of the Aldehyde Reactive Probe Reaction with AP-Sites in DNA: Influence of AP-Lyase on Adduct Stability.*

[386]   J. Kalia, R. T. Raines, *Angew. Chem. Int. Ed.* **2008**, *47*, 7523-7526. *Hydrolytic Stability of Hydrazones and Oximes.*

[387]   A. Dirksen, T. M. Hackeng, P. E. Dawson, *Angew. Chem. Int. Ed.* **2006**, *45*, 7581-7584. *Nucleophilic Catalysis of Oxime Ligation.*

[388]   Y. C. Jeong, R. Sangaiah, J. Nakamura, B. F. Pachkowski, A. Ranasinghe, A. Gold, L. M. Ball, J. A. Swenberg, *Chem. Res. Toxicol.* **2004**, *18*, 51-60. *Analysis of M1G-dR in DNA by Aldehyde Reactive Probe Labeling and Liquid Chromatography Tandem Mass Spectrometry.*

[389]   C. Malvy, M. Lefrancois, J. R. Bertrand, J. Markovits, *Biochimie* **2000**, *82*, 717-721. *Modified alkaline elution allows the measurement of intact apurinic sites in mammalian genomic DNA.*

[390]   W. Chan, B. Chen, L. Wang, K. Taghizadeh, M. S. Demott, P. C. Dedon, *J. Am. Chem. Soc.* **2010**, *132*, 6145-6153. *Quantification of the 2-Deoxyribonolactone and Nucleoside 5'-Aldehyde Products of 2-Deoxyribose Oxidation in DNA and Cells by Isotope-Dilution Gas Chromatography Mass Spectrometry: Differential Effects of γ-Radiation and Fe2+-EDTA.*

[391]   M. Münzel, L. Lercher, M. Müller, T. Carell, *Nucleic Acids Res.* **2010**, *38*, e192. *Chemical discrimination between dC and 5MedC via their hydroxylamine adducts.*

[392]   D. Wei, W. D. Wilson, S. Neidle, *J. Am. Chem. Soc.* **2012**, *135*, 1369-1377. *Small-molecule Binding to the DNA Minor Groove Is Mediated by a Conserved Water Cluster.*

[393]   T. Santa, *Biomed. Chromatogr.* **2011**, *25*, 1-10. *Derivatization reagents in liquid chromatography/electrospray ionization tandem mass spectrometry.*

[394]   H. Hong, Y. Wang, *Anal. Chem.* **2006**, *79*, 322-326. *Derivatization with Girard Reagent T Combined with LC−MS/MS for the Sensitive Detection of 5-Formyl-2'-deoxyuridine in Cellular DNA.*

[395]   N. Narayana, M. A. Weiss, *J. Mol. Biol.* **2009**, *385*, 469-490. *Crystallographic Analysis of a Sex-Specific Enhancer Element: Sequence-Dependent DNA Structure, Hydration, and Dynamics.*

[396]   O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455-461. *AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.*

[397]   D. Tsikas, J. Sandmann, B. Beckmann, *Analyst* **2011**, *136*, 407-410. *Analysis of NO and its metabolites by mass spectrometry. Comment on 'Detection of nitric oxide in tissue samples by ESI-MS' by Z. Shen, A. Webster, K. J. Welham, C. E. Dyer, J. Greenman and S. J. Haswell.*

[398]   P. M. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew. Chem. Int. Ed.* **2008**, *47*, 8350-8358. *Postsynthetic DNA modification through the copper-catalyzed azide-alkyne cycloaddition reaction.*

[399]   H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004-2021. *Click Chemistry: Diverse Chemical Function from a Few Good Reactions.*

[400]   V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *114*, 2708-2711. *A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes.*

[401]   C. W. Tornoe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057-3064. *Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides.*

[402]   B. T. Worrell, J. A. Malik, V. V. Fokin, *Science* **2013**, *340*, 457-460. *Direct evidence of a dinuclear copper intermediate in Cu(I)-catalyzed azide-alkyne cycloadditions.*

[403]   V. Brantl, Masterarbeit, Ludwig-Maximilians-Universität (München), **2013**. *Etablierung einer neuen LC-MS basierten Quantifizierungsmethode für abasische Stellen in genomischer DNA.*

[404]   R. Rahimoff, Masterarbeit, Ludwig-Maximilians-Universität (München), **2015**. *In Bearbeitung.*

[405]   S. L. Allinson, Dianova, II, G. L. Dianov, *EMBO J.* **2001**, *20*, 6919-6926. *DNA polymerase beta is the major dRP lyase involved in repair of oxidative base lesions in DNA by mammalian cell extracts.*

[406]   A. Bapat, M. L. Fishel, M. R. Kelley, *Antioxid. Redox Signaling* **2009**, *11*, 651-668. *Going ape as an approach to cancer therapeutics.*

[407]   Y. Huang, W. A. Pastor, J. A. Zepeda-Martinez, A. Rao, *Nat. Protoc.* **2012**, *7*, 1897-1908. *The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine.*

[408]   H. Hong, Y. Wang, *J. Am. Chem. Soc.* **2005**, *127*, 13969-13977. *Formation of intrastrand cross-link products between cytosine and adenine from UV irradiation of d((Br)CA) and duplex DNA containing a 5-bromocytosine.*

[409]   S. Liu, J. Wang, Y. Su, C. Guerrero, Y. Zeng, D. Mitra, P. J. Brooks, D. E. Fisher, H. Song, Y. Wang, *Nucleic Acids Res.* **2013**. *Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA*.

[410]   A. B. Robertson, J. A. Dahl, A. Klungland, *Nat. Chem. Biol.* **2014**, *10*, 487-488. *DNA metabolism: Bases of DNA repair and regulation*.

[411]   M. W. Kellinger, C. X. Song, J. Chong, X. Y. Lu, C. He, D. Wang, *Nat. Struct. Mol. Biol.* **2012**, *19*, 831-833. *5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription*.

[412]   C.-X. Song, Keith E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678-691. *Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming*.

[413]   M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat Chem* **2014**, *6*, 435-440. *Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution*.

[414]   X. Lu, C. X. Song, K. Szulwach, Z. Wang, P. Weidenbacher, P. Jin, C. He, *J. Am. Chem. Soc.* **2013**, *135*, 9315-9317. *Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA*.

[415]   D. Ji, K. Lin, J. Song, Y. Wang, *Mol. BioSyst.* **2014**, *10*, 1749-1752. *Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation*.

[416]   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*.

[417]   K. Kai, Y. Wataya, H. Hayatsu, *J. Am. Chem. Soc.* **1971**, *93*, 2089-2090. *Bisulfite-catalyzed isotope labeling of cytidine 5'-phosphate at its 5 position*.

[418]   Y. Wataya, H. Hayatsu, *Biochemistry* **1972**, *11*, 3583-3588. *Effect of amines on the bisulfite-catalyzed hydrogen isotope exchange at the 5 position of uridine*.

[419]   M. Sono, Y. Wataya, H. Hayatsu, *J. Am. Chem. Soc.* **1973**, *95*, 4745-4749. *Role of bisulfite in the deamination and the hydrogen isotope exchange of cytidylic acid*.

[420]   M. Maeda, Y. Kawazoe, *Tetrahedron Lett.* **1975**, *16*, 1643-1646. *Chemical alteration of nucleic acids and their components. XI. Hydrogen-deuterium exchange of nucleosides and nucleotides catalyzed by platinum*.

[421]   A. Burdzy, K. T. Noyes, V. Valinluck, L. C. Sowers, *Nucleic Acids Res.* **2002**, *30*, 4068-4074. *Synthesis of stable-isotope enriched 5-methylpyrimidines and their use as probes of base reactivity in DNA*.

8 Anhang: Zusatzmaterialien der veröffentlichten Arbeiten

8.1 Zusatzmaterialien zu Abschnitt 3.1

Toni Pfaffeneder, Benjamin Hackner, Matthias Truß, Martin Münzel, Markus Müller, Christian A. Deiml, Christian Hagemeier und Thomas Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012. *The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA.*

Supporting Information

General methods

All non-aqueous reactions were performed using flame- or ovendried glassware under an atmosphere of dry nitrogen. Commercial reagents from *Sigma-Aldrich* or *Acros* were used as received unless otherwise noted. Non-aqueous reagents were transferred under nitrogen with a syringe or cannula. Solutions were concentrated *in vacuo* on a *Heidolph* rotary evaporator with a *Vario PC2001* diphragm pump by *Vacuubrand*. Chromatographic purification of products was accomplished using flash column chromatography on *Merck* Geduran Si 60 (40-63 µM) silica gel (normal phase) or *Fluka* silica gel 100 C_{18}-Reversed phase (15-35 µm). Thin layer chromatography (TLC) was performed on *Merck* 60 (silica gel F_{254}) plates. Visualization of the developed chromatogram was performed using fluorescence quenching, anisaldehyde or ninhydrine staining. ^{1}H and ^{13}C NMR spectra were recorded in deuterated solvents on *Bruker ARX 300*, *Varian VXR400S*, *Varian Inova 400* and *Bruker AMX 600* spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. Routine ESI spectra and high-resolution ESI spectra were obtained on the mass spectrometer *Thermo Finnigan* LTQ FT-ICR. Acetonitrile for HPLC-ESI-MS analysis was purchased from *VWR*, HPLC gradient grade. HCOOH was purchased from *Fluka*, p.a. for mass spectrometry. Molecular sieves were rigorously dried in high vacuum at 350°C over night. IR measurements were performed on a *Perkin Elmer Spectrum BX FT-IR* spectrometer (*Perkin Elmer*) with a diamond-ATR (Attenuated Total Reflection) setup. Melting points were determined with a *Büchi Melting Point B540*. The final concentration of the purified oligonucleotides and enzyme concentrations were determined with a *NanoDrop ND-1000* spectrophotometer (*Peqlab*). Extinction coefficients of the oligonucleotides at 260 nm were calculated by addition of the extinction coefficients of the individual nucleobases. These are dA 15.0 L/mmol·cm, dC 7.1 L/mmol·cm, dG 12.0 L/mmol·cm and dT 8.4 L/mmol·cm. For the dC derivatives, the dC value was used. Nuclease S1 (*Aspergillus oryzae*) was obtained from *Roche*, snake venom phosphodiesterase I (*Crotalus adamanteus*) from *USB corporation* and antarctic phosphatase from *New England Biolabs*.

High Performance Liquid Chromatography

HPLC was performed on *Waters* or *Merck-Hitachi* units. These were in detail: analytical HPLC *Waters* Alliance (2695 Separation Module, 2996 Photodiode Array Detector), *Merck* analytical (L-7400 UV detector, L-7100 pump), preparative HPLC: *Waters* Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump), *Merck* preparative (L-7150 pump, L-7420 UV detector, *Rheodyne* P/N 77 25i injection valve, ERC-3415 solvent degasser). For analytical HPLC Nucleosil 120-3 C18 from *Macherey Nagel* were used, for preparative

HPLC Nucleosil 100-7 C18 also from *Macherey Nagel*. For LC-MS and for analytical HPLC of enzymatic digestion products a Uptisphere120-3HDO column from *Interchim* was used.

Buffer Systems

buffer system I: analytical and preparative reversed phase HPLC

Buffer A: 0.1 M NHEt$_3$OAc

Buffer B: 0.1 M NHEt$_3$OAc in 80% MeCN

buffer system II: analytical HPLC of enzymatic digestion products; LC-MS

Buffer C: 2 mM NH$_4$CHOO

Buffer D: 2 mM NH$_4$CHOO in 80% MeCN

MALDI-MS

MALDI spectra were recorded on a *Bruker* autoflex II unit with an *MTP* AnchorChip var/384 target. Prior to the measurements the samples were desalted using *MF-Millipore* membrane filters (0.025 μM).

MALDI matrices:

HPA: 3-hydroxypicolinic acid (50 mg), 15-Crown-5 (10 μl), ammonium hydrogencitrate

(10 mg) in 500 μL ddH$_2$O and 500 μL MeCN

HPA Crown: 3-hydroxypicolinic acid (50 mg), 15-Crown-5 (10 μl), ammonium hydrogencitrate (10 mg) in in 500 μL ddH$_2$O and 500 μL MeCN

Preparation of Stock Solutions for Mass Spectrometry

The stock solutions were prepared by weighing the HPLC purified and lyophilized nucleosides on a balance in a volumetric flask. To take weighing errors into account calibration curves were determined in triplicate. The weighed nucleosides were filled up to 100 mL in a volumetric flask with *dd*H$_2$O to receive concentration 1 of each nucleoside, followed by further dilution of 10 mL to 100 mL to obtain concentration 2, resulting in 1/10 of concentration 1. With only rare exceptions the lower concentration 2 or further diluted samples were used for quantification experiments. All nucleosides were first dissolved in 1 mL of DMSO and filled up with H$_2$O. Importantly, it was taken care that all four solutions of one compound (3 solutions of the unlabeled nucleoside and 1 solution of the isotope-labeled nucleosides) were dissolved in the same amount of DMSO to keep the solutions comparable. All stock solutions were stored at −20 °C and warmed up to rt prior to addition to

the digestion mixture (spiking). No decomposition or byproducts were observed after several defrosting and spiking experiments.

ES cell culture and differentiation:

ES cells established from C57BL/6 mice were routinely cultured in Knockout™ D -MEM (Invitrogen) supplemented with 10% Knockout™ Serum Replacement (Invitrogen), 5% fetal calf serum (FCS, Biochrome) recombinant LIF produced in COS cells, ß-mercaptoethanol (100μM final), penicillin/streptomycin (Gibco), Glutamax (Invitrogen), and non-essential amino acids (NEA; Biochrome). Differentiation was induced in suspension in 10 cm bacterial dishes using IMDM medium (Gibco), supplemented with 20% FCS (Biochrome), ß-mercaptoethanol (100μM final), penicillin/streptomycin (Gibco) and Glutamax (Invitrogen). After 2 and 3 days of differentiation, EB bodies were harvested by sedimentation.

We thank *En Li* for providing the DNMT3a/b double knockout cells.

DNA preparation:

ES cells and embryoid bodies were resuspended in 3 mL TE (10 mM Tris pH 8.0; 1 mM EDTA). 3 mL of lysis buffer (20 mM Tris pH 8.0; 4 mM EDTA; 20 mM NaCl; 1% SDS) were added, supplemented with 100 μL proteinase K (20 mg/mL). After 4 h incubation at 56 °C, RNAse A was added (100 μL; 10 mg/mL) and incubation was continued for 1 h. DNA was extracted with 0.5 volumes phenol followed by extraction with 0.5 volumes chloroform. Precipitation was performed with NaOAc (pH 7.0; 0.3 M final) and 1.5 volumes isopropanol.

Enzymatic Digestion of DNA

For the analysis of synthesized strands after HPLC purification 1 nmol synthetic DNA was digested as follows, for the analysis of genomic DNA samples up to 30 μg were digested. In advance, the genomic DNA samples were heated (in 100 μL *dd*H₂O) to 100 °C for 5 min to denature the DNA and rapidly cooled on ice.

For the enzymatic digestion DNA mixtures (in 100 μL H₂O) were mixed with buffer E (10 μL, 300 mM ammonium acetate, 100 mM CaCl₂, 1 mM ZnSO₄, pH 5.7) and nuclease S1 (80 units, *aspergillus oryzae*) and incubated for 3 h at 37 °C. Addition of buffer F (12 μL, 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 h at 37 °C completed the digestion. Labeled nucleosides [D^2]5-hmC and [D^3]5-mC were added to the digest of the genomic DNA. The concentrations of the standard solutions were chosen to be in the expected range of the sample nucleoside concentration. After centrifugation (12100 g, 15 min) of the sample, the supernatant was removed and the volume reduced to a minimum of 80 μL and a maximum of 100 μL.

HPLC-ESI-MS

The samples (80-100 µL injection volume) were analyzed by HPLC-ESI-MS on a *Thermo Finnigan* LTQ Orbitrap XL and chromatographed by a *Dionex* Ultimate 3000 HPLC system with a flow of 0.15 mL/min over an Uptisphere120-3HDO column from *Interchim* (buffer system II). The column temperature was maintained at 30 °C. The gradient was 0 → 12 min; 0 % → 3 % buffer D; 12 → 60 min; 3 % → 60 % buffer D; 60 → 62 min; 60 % → 100 % buffer D; 62 → 70 min; 100 % buffer D; 70 → 85 min; 100 → 0 % buffer D; 85 → 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 use of a positive polarity mode over a full-scan range of *m/z* 200–1000 with a resolution of 30.000. Parameters of the mass spectrometer were tuned with a freshly mixed solution of adenosine (5 µM) 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 60 V.

HPLC-ESI-MS/MS and HPLC-ESI-MS³

The samples (80-100 µL injection volume) were analyzed by HPLC-ESI-MS on a *Thermo Finnigan* LTQ Orbitrap XL and chromatographed by a *Dionex* Ultimate 3000 HPLC system with a flow of 0.15 mL/min over an Uptisphere120-3HDO column from *Interchim* (buffer system II). The column temperature was maintained at 30 °C . The gradient was 0 → 12 min; 0 % → 3 % buffer D; 12 → 60 min; 3 % → 60 % buffer D; 60 → 62 min; 60 % → 100 % buffer D; 62 → 70 min; 100 % buffer D; 70 → 85 min; 100 → 0 % buffer D; 85 → 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 use of a positive polarity mode over a full-scan range of *m/z* 200-1000 with a resolution of 30.000. The fragmentation conditions were optimized so that the precursor ion was still visable (~10%) using a 1 mM fC solution directly injected into the mass-spectrometer [1]. The collision energy was 15-28% and the isolation width was set at 2 and 1 *m/z* for MS/MS respectively MS³. The activation Q was set to 0.25 to minimize the loss of fragment ions. The mass range for the trapped ions was set to *m/z* 100-300 for MS/MS and *m/z* 50-200 for MS³. 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 60 V.

Mass calibration curves

Mass calibration curves of the labeled and corresponding unlabeled synthesized nucleosides were obtained at five different concentration ratios. For each concentration an average value of three independent measurements was determined (**Figure S1**). Each labeled nucleoside solution was mixed with three solutions with different concentrations of the corresponding unlabeled nucleosides. The areas of labeled and unlabeled nucleosides of the LC-MS

measurements were determined using the Qualbrowser program by extraction of the accurate mass with a mass filter (**Table S1**) from the total ion current (TIC). The linear fits of the determined area ratios over the amount ratios gave R^2-values of minimum 0.9987 for mC and 0.9997 for hmC. The linear equations were used for calculation of the exact nucleoside contents in genomic DNA samples. Synthetic labeled nucleosides were added to the digest solutions and the areas of labeled and unlabeled nucleosides were determined as described above. The amount of each nucleoside was calculated from the obtained area ratios and the linear fit equations of the calibration curves.
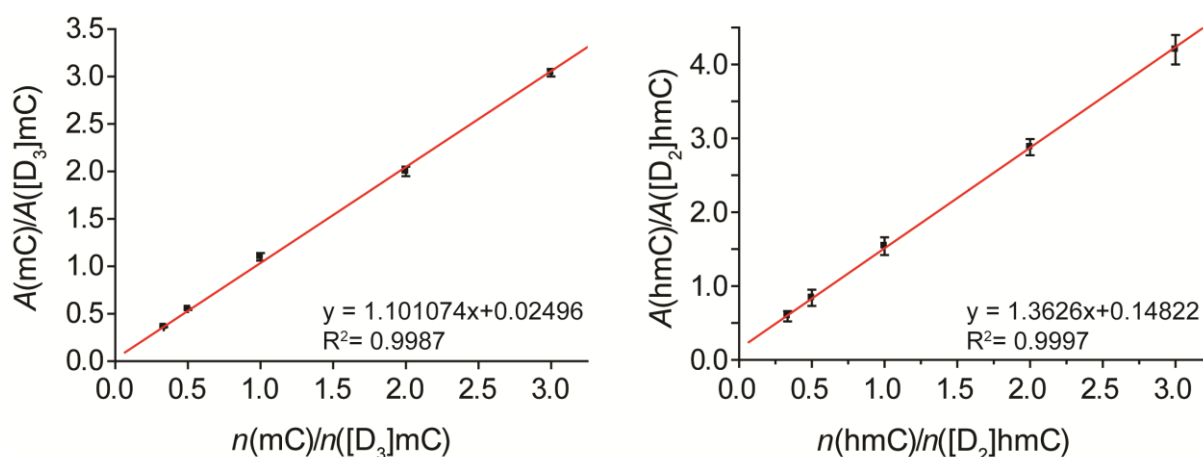
Figure S1. Mass calibration curves of the nucleosides mC and hmC.

Table S1. Applied mass ranges of analyzed nucleosides.

Nucleoside	Nucleoside Mass range *m/z*	Labeled nucleoside Mass range *m/z*
hmC	258.1024–258.1144	260.1167–260.1277
mC	242.1075–242.1195	245.1261–245.1381
fC	256.0868–256.0988	

HPLC calibration curves

The calibration curve for dG was measured to allow exact determination of the mC and hmC percentages. Therefore, the UV-peak areas of nine different amounts of dG were determined. These measurements were performed with two independent samples of dG. The resulting calibration curve with an R^2-value of 0.9997 is shown in **Figure S2**. Determination of the precise amounts of mC and hmC by the described isotope labeled mass spectrometric method and determination of the precise amount of dG by HPLC integration enabled calculation of the exact percentage of mC and hmC to dG (**Table S2**). We chose dG as a standard, because it forms base pairs with all four cytidine derivatives dC, mC, hmC and fC.

183
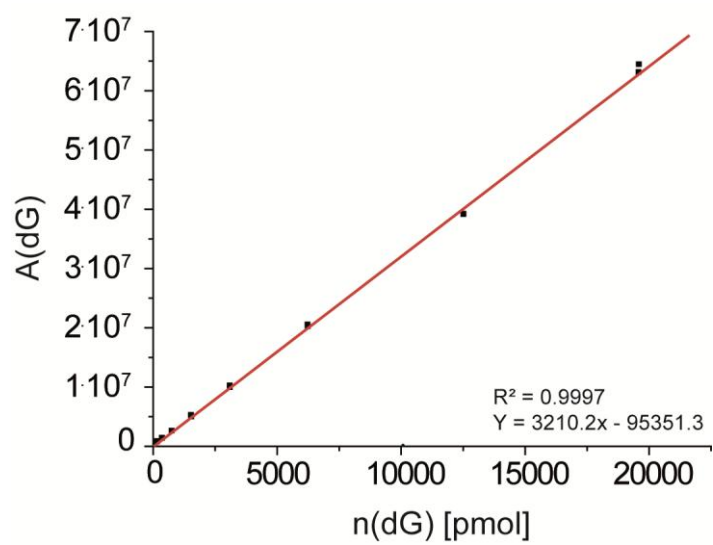
Figure S2. HPLC calibration curve of dG.

Measured values of mC and hmC

Table S2. Nucleoside percentages of mC to dG of mouse ES cells, embryoid bodies at 2 and 3 days of differentiation, myoblasts and DNMT3ab -/- cells (**Figure 3** in main text). Each sample is listed with the determined values, the standard deviation (SD) and the relative standard deviation (RSD). The average values over all independently prepared samples are listed in the last column.

Mouse		sample 1	sample 2	sample 3	average
Embryonic stem cells		4.97	5.46	5.19	**5.19**
	SD	0.37	0.27	0.26	0.36
	RSD / %	7.37	4.95	5.06	6.92
Embryoid body, 2 days differentiation		6.10			**6.10**
	SD	0.52			0.52
	RSD / %	8.46			8.46
Embryoid body, 3 days differentiation		6.22	6.28	4.62	**5.71**
	SD	0.49	0.78	0.35	0.95
	RSD / %	7.81	12.35	7.63	16.71
Myoblasts		3.53			**3.53**
	SD	0.43			0.43
	RSD / %	12.21			12.21
Embryonic stem cells, DNMT3ab-/-		0.54	0.45		**0.52**
	SD	0.13			0.13
	RSD / %	23.67			24.67

Table S3. Nucleoside percentages of hmC to dG of mouse ES cells, embryoid bodies at 2 and 3 days of differentiation, myoblasts and DNMT3ab -/- cells (**Figure 3** in main text). Each sample is listed with the determined values, the standard deviation (SD) and the relative standard deviation (RSD). The average values over all independently prepared samples are listed in the last column.

Mouse		sample 1	sample 2	sample 3	average
Embryonic stem cells		0.28	0.45	0.45	**0.39**
	SD	0.03	0.02	0.02	0.09
	RSD / %	9.51	4.31	3.90	22.98
Embryoid body, 2 days differentiation		0.20			**0.20**
	SD	0.02			0.02
	RSD / %	11.20			11.20
Embryoid body, 3 days differentiation		0.14	0.16	0.16	**0.15**
	SD	0.01	0.03	0.01	0.02
	RSD / %	6.70	20.14	4.52	13.76
Myoblasts		0			**0**
	SD				
	RSD / %				
Embryonic stem cells, DNMT3ab-/-		0.02	0.02		**0.02**
	SD	0.02			0.02
	RSD / %	115.63			86.60

HPLC-ESI-MS analysis of synthetic fC
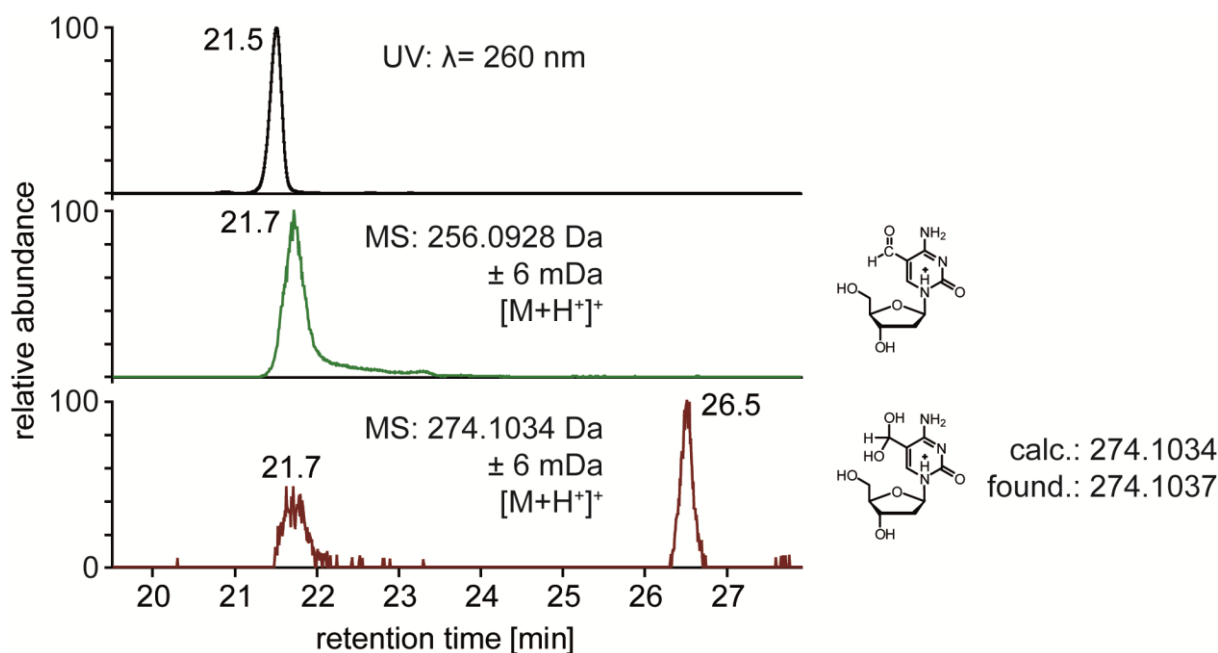
Figure S3. LC-MS analysis of synthetic fC. The delay between the retention time of UV-detection and MS-detection is 0.2 min. The hydrate of fC (26.5 min) can be separated by HPLC and also detected during the ionization process of fC (21.7 min).

Measured values of fC

Table S4. Correlation of the integral of fC to dG of mouse ES cells, embryoid bodies at 2 and 3 days of differentiation, myoblasts and DNMT3ab -/- cells (**Figure 3** in main text). Each sample is listed with the determined values, the standard deviation (SD) and the relative standard deviation (RSD). The average values over all independently prepared samples are listed in the last column.

Mouse		sample 1	sample 2	sample 3	average
Embryonic stem cells	Integral fC/dG	0.0032	0.0063	0.0012	**0.0035**
	SD	0.00031	0.00008	0.00016	0.00213
	RSD / %	9.86	1.30	13.21	60.88
Embryoid body, 2 days differentiation	Integral fC/dG	0			**0**
	SD				
	RSD / %				
Embryoid body, 3 days differentiation	Integral fC/dG	0	0	0	**0**
	SD				
	RSD / %				
Myoblasts	Integral fC/dG	0			**0**
	SD				
	RSD / %				
Embryonic stem cells, DNMT3ab-/-	Integral fC/dG	0	0		**0.02**
	SD				
	RSD / %				

Coinjection studies of fC:

In order to quantify the amount of the found fC in mouse ES cells, coinjection studies with synthetic fC were carried out. After enzymatic digestion of the mES cell DNA as described above, 10.13 pmol synthetic fC were added to the sample followed by centrifugation of the sample (12100 g, 15 min). The supernatant was removed, the volume reduced to a minimum of 80 µL and a maximum of 100 µL and subsequently subjected to LC-MS analysis. The sample was compared to a digest simulation containing 10.13 pmol fC, 15 nmol of each canonic nucleoside in 10 µL Buffer E and 12 µL Buffer F diluted to a final volume of 100 µL with ddH$_2$O.

The resulting mass signals for fC were integrated. The calculated sum of the integrals of the digest simulation and the genomic sample fits perfectly to the measured integral of the coinjection (**Fig. 3B** in main text).

Biotin-labeling of fC in DNA

fC was chemically labeled with an aldehyde reactive probe (ARP) through an oxime ligation reaction, which was originally established in order to quantify abasic sites in DNA [2].

Firstly, the reaction was optimized using synthetic oligodesoxynucleotides (ODN) containing either fC (ODN1), dC (ODN2) or hmC (ODN3) (*Table S5*).

Table S5. Synthetic short oligodesoxynucleotides (ODN) used in this work. MS-analysis was carried out by MALDI.

Strand No.	Sequence	Mass (calc.)	Mass (found)
ODN1	5'-GTA ATG XGC TAG G-3' (X = fC)	4043.0	4041.2
ODN2	5'-GTA ATG XGC TAG G-3' (X = dC)	4015.0	4013.2
ODN3	5'-GTA ATG XGC TAG G-3' (X = hmC)	4045.0	4043.8

In a typical reaction, 8 µM synthetic fC-containing DNA (ODN1) in 110 mM *p*-methoxy aniline, 44 mM NaOAc (pH 5.5) and 0.4 mM N-(aminooxyacetyl)-N'-(D-biotinoyl) hydrazide (Biomol) was incubated for 24 h at 25 °C. The usage of p-methoxy aniline enhanced the reaction rate as described by Dirksen *et al* [3]. The reaction was stopped by standard ethanol precipitation by adding 1/10 volumes 3 M NaOAc (pH 5.0) followed by the addition of 2–2.2 volumes absolute ethanol. After precipitation at −20 °C for at least 1 h, the samples were centrifuged (12100 g, 15 min) and the supernatant discarded. The pellets were washed with 75% ethanol (500 µL) twice followed by centrifugation (12100 g, 15 min) after each washing step and then taken up in ddH$_2$O. **Figure S4** shows the HPLC- and MALDI-analysis of ODN1–3 after ARP-treatment. Coinjection of the starting material proved the conversion of the fC-containing ODN1 was quantitative. The MALDI-TOF mass spectrogram revealed the mass of the product to be *m/z* = 4718.7, which corresponds to the calculated value of *m/z* =

4714.3 for the biotinoylated ODN1. In comparison, no reaction was observed for the control strands ODN2 and ODN3 (**Figure S4**). The exact mass of the biotinoylated fC-nucleoside was determined after enzymatic digestion of biotinoylated ODN1 and subsequent LC-MS analysis (**Figure S5**). The biotin-fC nucleoside has a retention time of 30 min with an observed mass of $m/z = 569.2132$ (calc.: $m/z = 569.2137$).

The incubation of genomic DNA samples with ARP was carried out in analogy. The final DNA concentration was adjusted to 0.3 µg/µL. After purification and enzymatic digestion a mass peak with the same retention time and exact mass could be observed (**Figure 4**, main text). The samples were also analyzed regarding to possible side-reactions with abasic sites, which could not be detected using the established LC-MS method.
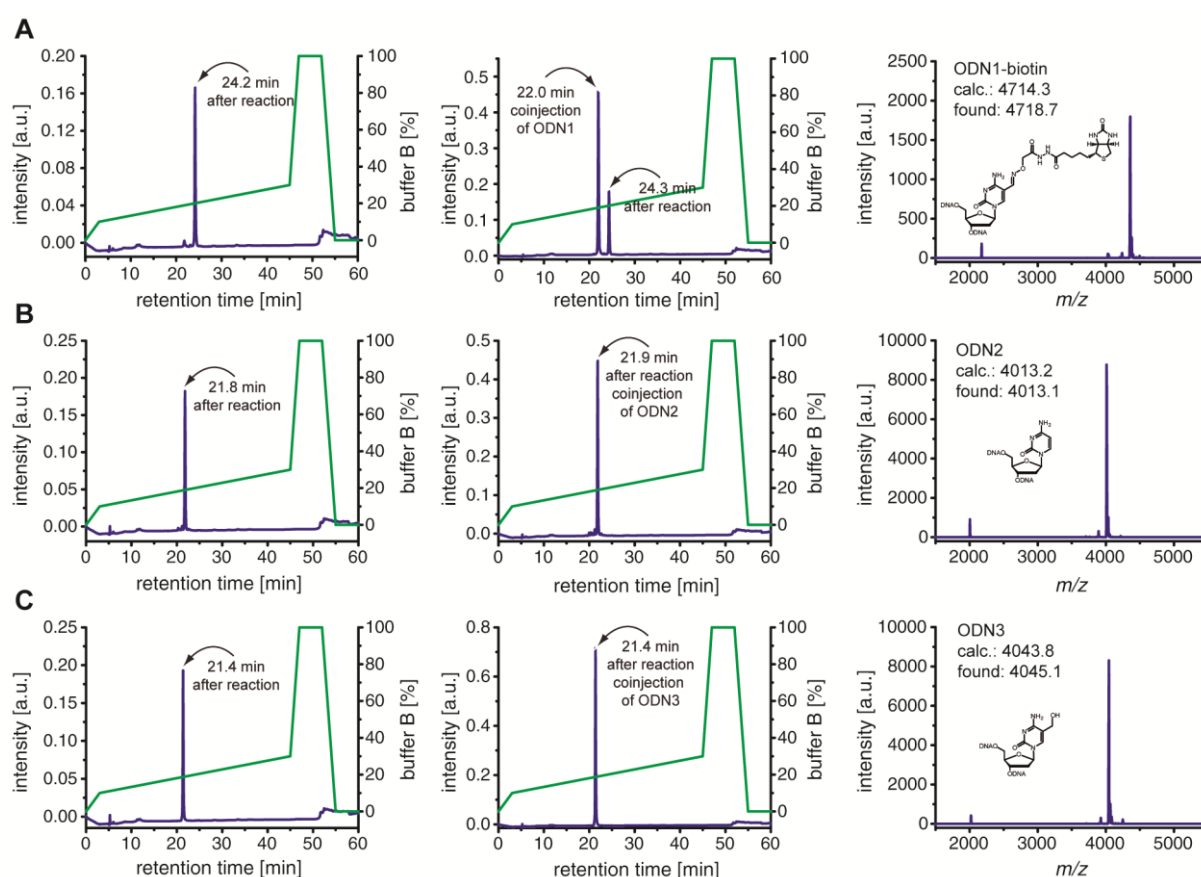
Figure S4. HPLC- and MALDI-analysis of ODN1 (A), ODN2 (B) and ODN3 (C) after incubation with ARP. Only the fC containing oligonucleotide ODN1 reacts with the probe.
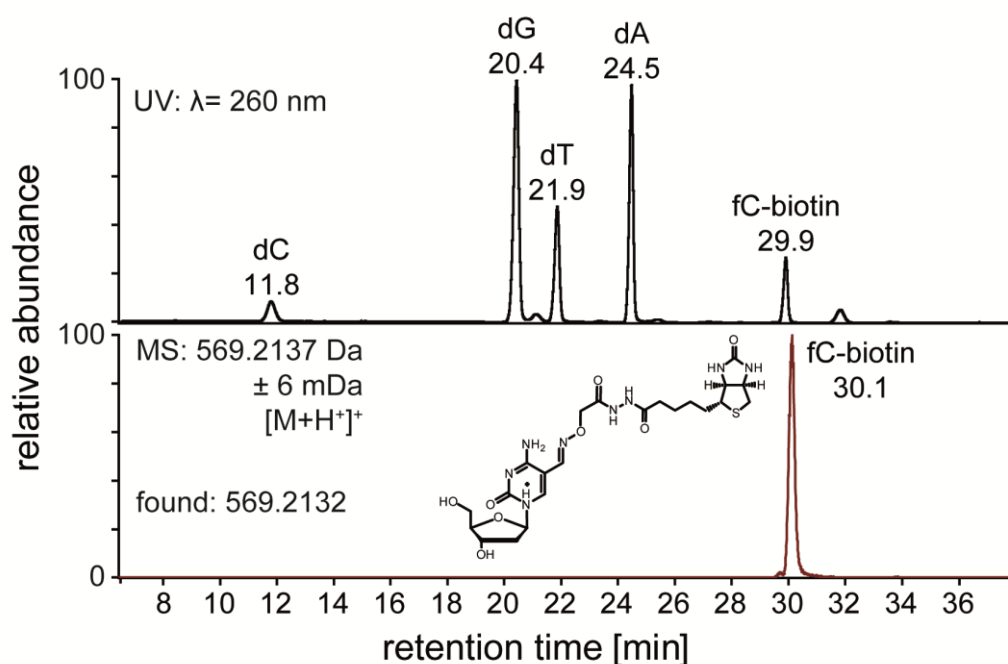
Figure S5. LC-MS analysis of the digested nucleoside mixture of ODN1 after ARP-treatment. Top: HPL-chromatogram of the nucleoside mixture obtained after enzymatic digest of ODN1; bottom: mass filter for fC-biotin. The delay between the retention time of UV-detection and MS-detection is 0.2 min.

Measured values of biotin-fC

Table S6. Correlation of the integral of biotin-fC to dG of mouse ES cells, embryoid bodies at 2 and 3 days of differentiation, myoblasts and DNMT3ab -/- cells (**Figure 4** in main text). Each sample is listed with the determined values, the standard deviation (SD) and the relative standard deviation (RSD). The average values over all independently prepared samples are listed in the last column.

Mouse		sample 1	sample 2	sample 3	average
Embryonic stem cells	Integral fC/dG	0.0205	0.0280	0.0265	**0.02881**
	SD	0.0085	0.0079		0.01100
	RSD / %	41.27	28.41		39.20
Embryoid body, 2 days differentiation	Integral fC/dG	**0.00281**			**0.00281**
	SD	0.00226			0.00226
	RSD / %	80.36			80.36
Embryoid body, 3 days differentiation	Integral fC/dG	0.00042	0.00076	0	**0.00040**
	SD	0.00013			0.00032
	RSD / %	30.88			80.31
Myoblasts	Integral fC/dG	0			**0**
	SD				
	RSD / %				
Embryonic stem cells, DNMT3ab-/-	Integral fC/dG	0	0		**0**
	SD				
	RSD / %				

Synthesis and Incorporation of fC into DNA

Generally, direct incorporation of aldehydes into DNA is considered very challenging.[4] To the best of our knowledge there is only one example, in which an unprotected formyl group could be successfully incorporated by solid phase synthesis.[5] Furthermore only very few examples for the incorporation of aldehydes by PCR are known, making the preparation of DNA strands that contain aldehydes very complicated.[4] In most studies, the formyl group was either protected, e.g. as an acetal, or it was generated by postsynthetic cleavage of a diol.[6] *Karino et al.* previously generated DNA strands that contained fC from a a diol precursor. This procedure, however, is complicated and the synthesis of the diol-phosphoramidite required ten synthetic steps from dU.[6d]
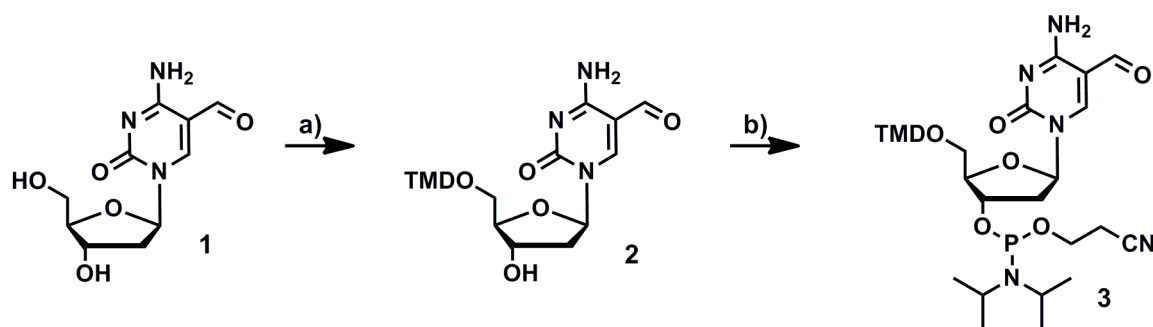
Scheme S1. Synthesis of the 5-formylcytosine-phosphoramidite **3**. a) DMT-Cl, pyr, 21 h, rt, 87 %, b) bis(di*iso*propylamino)(2-cyanoethoxy)phosphine, di*iso*propyltetrazolide, DCM, 15 h, rt, 92 %.

For the synthesis of the fC-phosphoramidite we initially planned to protect the 4-amino group as an acetate. However, the resulting compound was not stable in pyridine as solvent and therefore prohibited DMT protection. This can possibly be explained by the presence of the formyl group, which renders the already electron deficient heterocycle even more electron poor. It thereby becomes a good leaving group and the acetamide can consequently be cleaved by a plethora of weak nucleophiles. The electron deficiency imposed by the formyl group has the further effect that the aniline-type 4-amino group gets even less nucleophilic. We wanted to exploit this and thought that possibly no protection of the exocyclic amino group was needed for the incorporation into oligonucleotides. The synthesis of the fC-phosphoramidite **3** is depicted in Scheme 1. fC (**1**), which was synthesized as shown in , was first DMT protected (**2**) and subsequently converted to **3** with the *Bannwarth*-reagent.
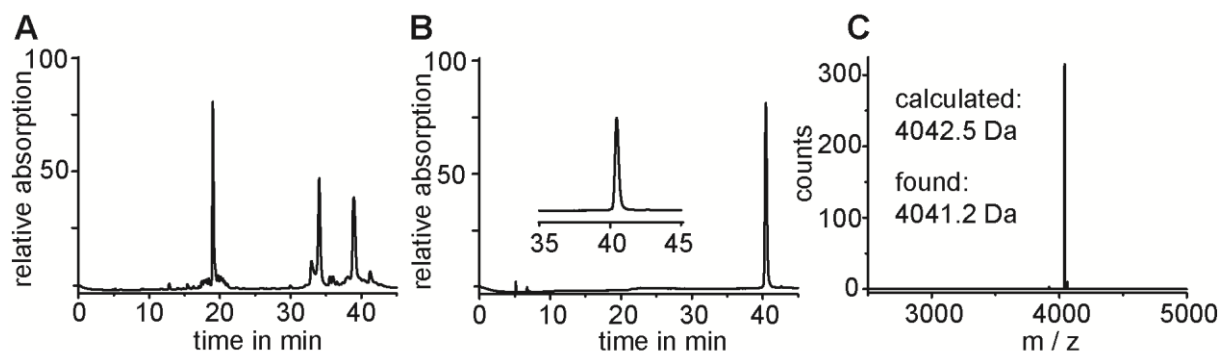
Figure S6. Incorporation of fC into DNA strands. Sequence: 5' GTA ATG C*GC TAG G 3' ODN1: C* = fC, A) Reversed phase HPLC chromatogram (0 - 50 % buffer B in 45 min) of ODN1 directly after cleavage from the resin. B) Reversed phase HPLC chromatogram (0 - 20 % buffer B in 45 min) of purified ODN1. C) MALDI spectrum of pure ODN1.

For the incorporation of fC into oligonucleotides standard solid phase DNA synthesis was performed with phosphoramidite **3**. Coupling times for **3** were doubled to ensure good yields. As shown in **Figure S6**, the incorporation of the DNA modifications was successful. However, incorporation of the aldehyde resulted in the formation of several byproducts (**Figure S6A**). With fC phosphoramidite **3** in principle two side reactions are possible. First of all the aldehyde could be oxidized in the oxidation step of each coupling cycle. Secondly, the presence of the unprotected 4-amino group in **3** could lead to a branching of the strands. All byproducts were isolated by HPLC, but no products of the expected side reactions could be detected. Desired ODN1, however, could be purified, as shown in **Figure S6B**. MALDI-MS proved the correct mass of the strand (**Figure S6C**). Thus we obtained the desired fC-containing strands in excellent purity and moderate yield. Deprotection was possible with NH_4OH or with NaOH.

To gain further evidence for the exclusive formation of fC we enzymatically digested ODN1. To this end we treated the oligonucleotides first with nuclease S1 for 3 h at 37 °C followed by incubation with antarctic phosphatase and snake venom phosphordiesterase for additional 3 h at 37 °C. The obtained nucleoside mixture was analyzed by HPLC-ESI-MS.

The chromatograms are depicted in **Figure S7**: For ODN1 only four peaks could be seen. To find the explanation for the missing signal for fC and to assign all peaks for ODN1 we subjected the nucleoside mixtures to a LC/MS study.

191
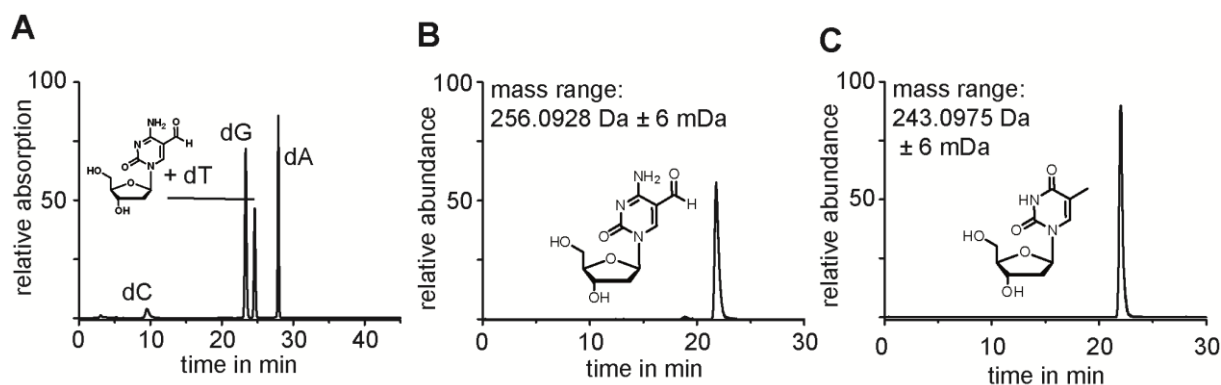
Figure S7. HPLC and LC/MS analysis of the digested nucleoside mixture. A) HPLC chromatogram of the nucleoside mixture obtained after enzymatic digest of ODN1 B) Mass filter for fC. C) Mass filter for dT.

The mass spectrum revealed that the expected nucleoside fC comigrates with dT on the HPLC column (**Figure S7A+C**). The mass data show the right high resolution mass expected for fC and therewith prove correct incorporation of the DNA modification into ODN1.

In conclusion we have developed a phosphoramidite building block for the incorporation of fC into DNA. This allows the efficient preparation of oligonucleotides for biochemical studies. These strands can now be used as probes that will help to decipher if oxidative demethylation pathways exist and will shine more light on the function of fC in cellular DNA.

Solid Phase DNA Synthesis

Oligonucleotide synthesis was performed on an Expedite 8909 Nucleic Acid Synthesis System (*PerSeptive Biosystems*) using standard DNA synthesis conditions (scale: 1 μM). CE-Phosphoramidites for dA (Bz-dA), dC (Bz-dC), dG (*i*Bu-dG), dT and CPG carriers were obtained from *Glen Research* or *Link Technologies*. The following concentrations were used: 0.1 M phosphoramidite in MeCN (10 ppm H_2O), activator: 0.25 M benzothiotetrazole in MeCN (10 ppm H_2O), Capping: Ac_2O (20 %), 2,6-lutidine (30 %) in MeCN (50 %, 30 ppm H_2O), Detritylation: 3 % dichloroacetic acid in toluene, Oxidation: 25 mM iodine in 2,6-lutidine/MeCN/H_2O (6/65/30). The activator BTT was synthesized according to a standard protocol.[7] The terminal DMT protecting group was kept on the oligonucleotides after synthesis and removed after cleavage from the resin (see deprotection and purification). Except for the special amidite standard coupling conditions were used. For the special amidite coupling times were doubled to ensure good yields.

Deprotection and purification of oligonucleotides

Deprotection and cleavage of the oligonucleotides from the CPG carrier was carried out with 0.4 M NaOH solution (1 mL) in MeOH/H_2O 4:1 for 18 h at room temperature (shaken at 1250 rpm) (Deprotection is also possible by NH_4OH treatment). After this, 600 μL 1M

192

triethylammoniumacetate was added to neutralize the NaOH. The mixture was centrifuged (13000 prm, 15 min) and the supernatant collected. subsequently, the solution was concentrated to ~ 30 % of the original volume in a speedvac (Christ RVC-2-33 IR). DNA purification was conducted on analytical and preparative HPLC (buffer system I) using *Nucleosil* columns (250 * 4 mm, C18ec, particle size 3 μm or 250 * 10 mm, C18ec, 5 μm) from *Machery-Nagel*. The fractions were checked for purity by analytical HPLC and MALDI-MS. The purified oligonucletides were concentrated using a *Christ* alpha 2-4 LD plus lyophyllizer.

Cleavage of the DMT group

The oligonucleotides still containing the trityl group were deprotected by addition of 100 μL of an 80% acetic acid solution. After incubation at r.t. for 20 min 100 μL of water together with 60 μL of a 3 M solution of sodium acetate were added. Finally, the ODN was precipitated by the addition of 1600 μL ethanol at -20°C (30 min). The strands were stored in *dd*H_2O.

Synthesis

5-Formyldeoxycytidine was synthesized as described previously.[8]

5'-(dimethoxytrityl)-5-formyl-2'deoxycytidine (2)
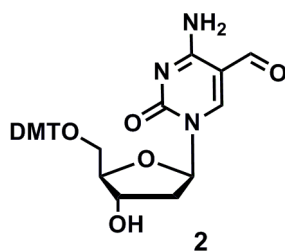
A round bottom flask was charged with rigorously dried 3 Å molecular sieves. 159 mg (0.62 mmol, 1.0 eq.) **1**, (final concentration: 66 mM) and 218 mg (0.62 mmol, 1.0 eq) DMT-Cl were dissolved in pyridine. The reaction mixture was stirred for 14 h and the molecular sieves subsequently filtered off. The filtrate was evaporated to dryness and the crude product purified by column chromatography (DCM/MeOH 29:1, 19:1).

Yield: 301 mg (87 %) of **2** as a beige solid.

¹H-NMR (599 MHz, CDCl₃) δ = 8.64 (s, 1H, C-C**H**=O), 8.33 (d, ^4J = 0.9, 1H, N-C**H**=C-CHO), 8.09 (s, 2H, N**H**₂), 7.35 – 7.32 (m, 2H, DMT), 7.29 – 7.25 (m, 3H, DMT), 7.23 (m, 4H, DMT), 6.81 (d, ^3J = 8.2, 4H, DMT), 6.24 (t, ^3J = 5.8, 1H, O-C**H**-N), 4.61 (m, 1H, O-C**H**-

CH$_2$-CH-N), 4.19 (dd, ^3J = 3.0, ^3J = 6.7, 1H, O-CH$_2$-C**H**-CH-O), 3.77 (s, 6H, O-C**H**$_3$), 3.51 (dd, ^3J = 2.8, ^2J = 10.8, 1H, 1x O-C**H**$_2$-CH-O), 3.40 (dd, ^3J = 3.0, ^2J = 10.8, 1H, 1x O-C**H**$_2$-CH-O), 2.77 – 2.70 (m, 1H, 1x O-CH-C**H**$_2$-CH-N), 2.39 – 2.33 (m, 1H, 1x O-CH-C**H**$_2$-CH-N).

^{13}C-NMR (151 MHz, CDCl$_3$) δ = 187.7, 162.6, 158.8, 158.8, 153.5, 153.1, 143.9, 135.2, 135.0, 130.1 (2C), 130.1 (2C), 128.2 (2C), 128.0 (2C), 127.3, 113.3 (4C), 104.9, 87.4, 86.9, 86.5, 70.5, 62.6, 55.2 (2C), 42.4.

HRMS (ESI–): calculated for C$_{31}$H$_{30}$N$_3$O$_7$$^-$ [M–H]$^-$: 556.2089, found: 556.2079

Melting Range: 129–131 °C

IR (ATR): ṽ = 3360 (w), 3219 (br, w), 2934 (w), 2599 (w), 1642 (s), 1506 (s), 1233 (m), 1172 (m), 1099 (m), 1035 (s), 789 (s), 725 (m), 698 (m).

3'-(di*iso*propylamino-cyanoethoxyphosphino)-5'-(dimethoxytrityl)-5-formyl-2'deoxycytidine (3)
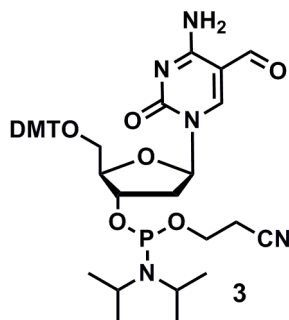
In a Schlenck tube 243 mg (0.44 mmol, 1.0 eq) **2** (final concentration: 50 mM), 39 mg (0.22 mmol, 0.5 eq) di*iso*propyltetrazolide, and 169 μL (0.52 mmol, 1.2 eq) bis(di*iso*-propylamino) (2-cyanoethoxy)phosphine, were dissolved in 10 mL rigorously degassed DCM and the solution degassed three more times (freeze, pump, thaw). The reaction mixture was stirred at room temperature for 15 h and subsequently concentrated in an argon atmosphere. The crude product was purified by column chromatography ((DCM/MeOH 49:1, 0.1 % NEt$_3$). The test tubes that were used to collect the fractions were flushed with argon prior to use.

Yield: 305 mg (92 %, colorless foam) of **3** as a mixture of two diastereomers on P.

The compound was directly used for solid phase DNA synthesis. Its identity was proven by successful incorporation into DNA.

^{31}P-NMR (81 MHz, Acetone-D$_6$) δ = 149.5, 149.3.

[1] H. Cao, Y. Wang, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1335-1341.

[2] a) K. Kubo, H. Ide, S. S. Wallace, Y. W. Kow, *Biochemistry* **1992**, *31*, 3703-3708; b) J. Nakamura, V. E. Walker, P. B. Upton, S.-Y. Chiang, Y. W. Kow, J. A. Swenberg, *Cancer Res.* **1998**, *58*, 222-225.

[3] A. Dirksen, T. M. Hackeng, P. E. Dawson, *Angew. Chem., Int. Ed.* **2006**, *45*, 7581-7584.

[4] V. Raindlova, R. Pohl, M. Sanda, M. Hocek, *Angew. Chem., Int. Ed.* **2010**, *49*, 1064-1066.

[5] A. Okamoto, K. Tainaka, I. Saito, *Tetrahedron Lett.* **2002**, *43*, 4581-4583.

[6] a) G. H. Clever, K. Polborn, T. Carell, *Angewandte Chemie International Edition* **2005**, *44*, 7204-7208; b) C. T. Wirges, J. Timper, M. Fischler, A. S. Sologubenko, J. Mayer, U. Simon, T. Carell, *Angewandte Chemie International Edition* **2009**, *48*, 219-223; c) H. Kamiya, H. Tsuchiya, N. Karino, Y. Ueno, A. Matsuda, H. Harashima, *J. Biochem. (Tokyo)* **2002**, *132*, 551-555; d) N. Karino, Y. Ueno, A. Matsuda, *Nucleic Acids Res.* **2001**, *29*, 2456-2463.

[7] R. Welz, S. Müller, *Tetrahedron Lett.* **2002**, *43*, 795-797.

[8] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367.

8.2 Zusatzmaterialien zu Abschnitt 3.2

Martin Münzel, Ulrike Lischke, Dimitrios Stathis, Toni Pfaffeneder, Felix A. Gnerlich, Christian A. Deiml, Sandra C. Koch, Konstantin Karaghiosoff und Thomas Carell, *Chem. Eur. J.* **2011**, *17*, 13782–13788. *Improved Synthesis and Mutagenicity of Oligonucleotides Containing5-Hydroxy-methylcytosine, 5-Formylcytosine and 5-Carboxylcytosine.*

Supporting Information

Buffer Systems for HPLC

buffer system I: analytical and preparative reversed phase HPLC

Buffer A: 0.1 M NHEt$_3$OAc

Buffer B: 0.1 M NHEt$_3$OAc in 80% MeCN

buffer system II: analytical HPLC of enzymatic digestion products; LC-MS

Buffer C: 2 mM NH$_4$CHOO

Buffer D: 2 mM NH$_4$CHOO in 80% MeCN

always used with standard gradient: 0 → 12 min; 0 % → 3 % buffer D; 12 → 60 min; 3 % → 60 % buffer D; 60 → 62 min; 60 % → 100 % buffer D; 62 → 70 min; 100 % buffer D; 70 → 85 min; 100 → 0 % buffer D; 85 → 95 min; 0 % buffer D.

MALDI-MS

MALDI spectra were recorded on a *Bruker* autoflex II unit with an *MTP* AnchorChip var/384 target. Prior to the measurements the samples were desalted using *MF-Millipore* membrane filters (0.025 µM). Desalting times depended on the molecular weight of the analyte. (4000 Da: 30 min, 12000 Da: over night)

MALDI matrix:

HPA Crown: 3-hydroxypicolinic acid (50 mg), 15-Crown-5 (10 µl), ammonium hydrogencitrate (10 mg) in 500 µL ddH$_2$O and 500 µL MeCN

Sequences of DNA Strands

ODN1: 5' d(GTA ATG XGC TAG G)3' X = dC, mC, hmC, fC, caC

ODN2: 5' d(GTA GCC AGG TCG CAC GCG TGC TAX GAT GCG AGA CTG C) 3' X = dC, mC, hmC, fC, caC

ODN3: 5' d(CCT AGC YCA TTA C)3' Y = dA, dG, dC, dT

ODN4: 5' d(Bio GCA GTC TCG CAT C)3'

ODN5: 5' d(GTA GCC AGG TCG CAC GCG TGC TA)3'

Stability assay of fC and caC at the oligonucleotide level

The oligonucleotides used for this assay where HPLC purified and tested for analytical purity prior to the stability analysis. A solution of ODN1 (18 pmol/μL, X = dC, fC, caC) was spiked with 0.01 % HCOOH and directly incubated at 37°C. Additionally, ODN1 was incubated in buffered solution at pH 7.0 and 5.0 at 37°C. Sequential injections (10 μL, 180 pmol DNA) every six hours and over a time frame of 18 h reveals the decomposition of the fC and caC strands in direct comparison with the control dC containing strand. The analytical HPLC runs where performed by application of buffer system I, with pH value strictly regulated to 7.00 (flow rate: 0.5 mL/min, gradient: 0→20 % buffer B in 45 min).

LC MS analysis of digested DNA strands
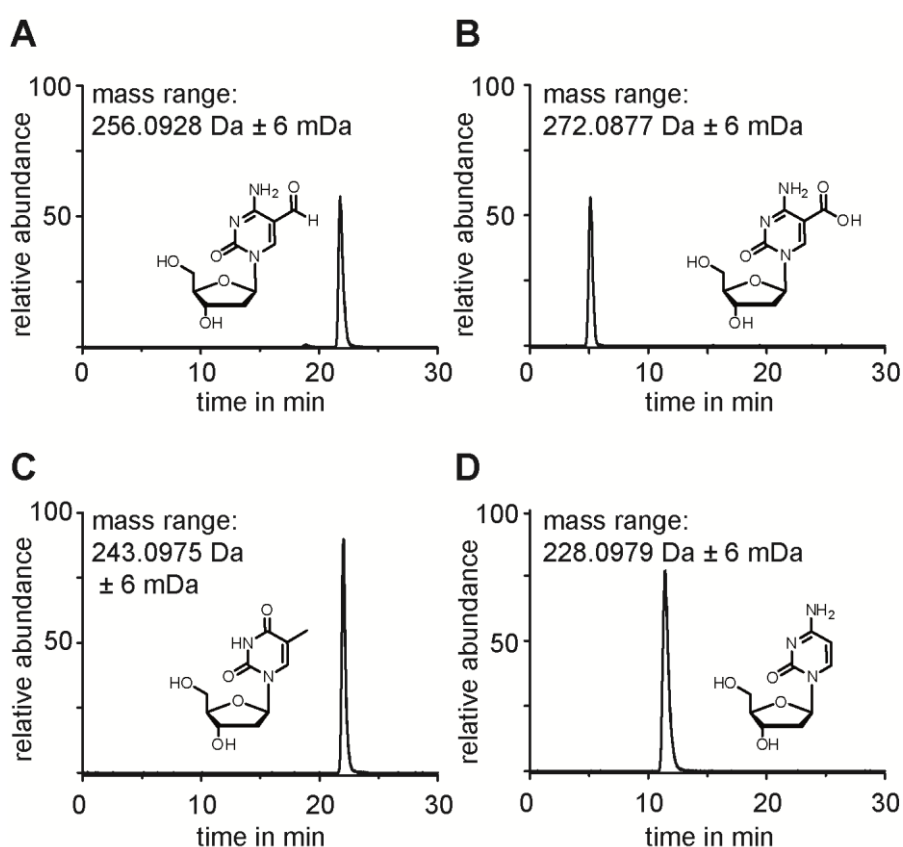
Figure S1: LC/MS analysis of the digested nucleoside mixtures. A) Mass filter for [5-CHO]dC. B) Mass filter for [5-COOH]dC. C) Mass filter for dT. d) Mass filter for dC.

Purification of human polymerase κ

Human polymerase kappa was cloned from human cDNA. The Stargate® system was used as a cloning strategy and the first codons were optimized for *E.coli* usage by using the primers below.

Primer:

5'P – AATG GG<mark>C</mark> CT<mark>G</mark> AA<mark>C</mark> GAT AAT AAA GCA <mark>GGA</mark> ATG GAA GGA T

5'P – TCCC TTG TTG GTG TTT CCT GTC CTC TTC

Furthermore the PCR-product was applied in a Stargate® reaction to form an entry clone (pPSG-IBA10) containing a short version of human polymerase kappa (polκs, 19-526). The pPSG-IBA33 was used as expression plasmid, which was transformed in Rosetta 2 (DE3) cells. Polκs was expressed using an autoinduction medium (ZYP-5052) adopted from Studier (Protein Expression and Purification 41 (2005) 207–234). The cells were grown at 37°C for 4 h and then for 16 h at 25°C and harvested by centrifugation. Subsequently the pellet was lysed in His-Buffer A (50 mM Tris pH = 7.4, 500 mM NaCl, 20 mM imidazol, 0.2 % Tween 20) using a french press and applied to a Ni-NTA superflow column. Polκs was eluted in buffer B (50 mM Tris pH=7.4, 500 mM NaCl, 250 mM imidazol). Furthermore the protein was transferred to Heparin buffer A (50 mM Tris 7.4, 100 mM NaCl, 5 mM DTT, 5 % glycerol) and injected into a heparin column. A high salt buffer (Hep B, 50 mM Tris pH = 7.4, 800 mM NaCl, 5mM DTT, 5 % glycerol) was used for elution. Afterwards the central peak fractions were pooled and concentrated using an Amicon 30K 15 ml. As a polishing step the protein was applied to Superdex G75 with gel filtration buffer (25 mM HEPES pH = 7.0, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP) as mobile phase. The protein was again concentrated and stored at -20°C by supplementing the buffer with 50 % glycerol.

Purification of yeast polymerase η

Pol η was purified as described before.[1]

Crystal Structure of fC

A characteristic feature of the structure of fC is a strong intramolecular hydrogen bond (Table 1) between the oxygen atom of the CHO group and the adjacent NH_2 group involving the hydrogen atom H22. Due to this hydrogen bonding the orientation of the CHO-group is fixed and there is no disorder of this group with respect to a rotation around the C_{ring}-CHO bond observed. The most fascinating feature is the crystal packing of fC. It is determined by an extensive network of intermolecular hydrogen bonds. To this network participates the NH_2 group, both OH groups and the oxygen atoms of both carbonyl groups of each molecule. Interestingly, while the primary OH-group acts only as donor, the secondary OH-group acts both as donor and as acceptor (Fig. S2). In the crystal every molecule undergoes O–H···O hydrogen bonds with four neighboring molecules acting two times as acceptor via the carbonyl oxygen atoms and two times as donor via both OH-groups. With two further molecules N–H···O hydrogen bonds are formed. This results in the formation of layers (Fig. S3), which are strongly puckered due to the fixed angle between the five-membered and the six-membered ring of fC. These layers are packed to form the crystal.
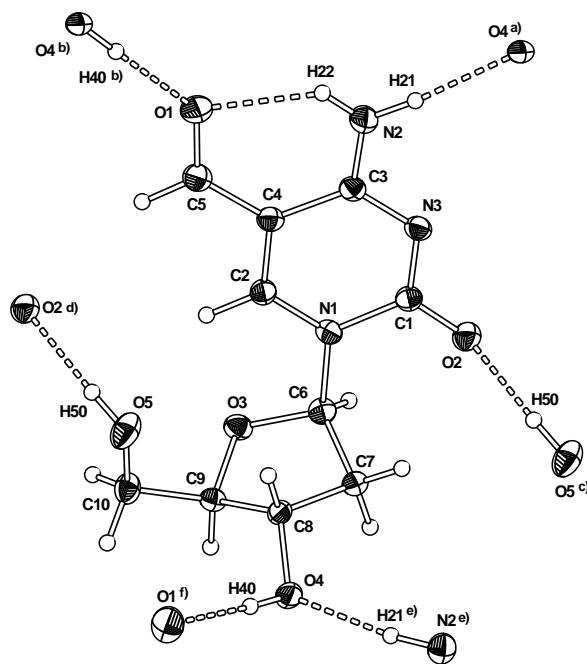
Figure S2: Crystal structure of fC showing the hydrogen bonds in which one molecule of fC is involved in the crystal. ORTEP representation, thermal ellipsoids are drawn at 50% probability level. Symmetry codes: a) 1-x, -0.5+y, 1-z; b) 2-x, -0.5+y, 2-z; c) -1+x, y, -1+z; d) 1+x, y, 1+z; e) 1-x, 0.5+y, 1-z; f) 2-x, 0.5+y, 2-z.
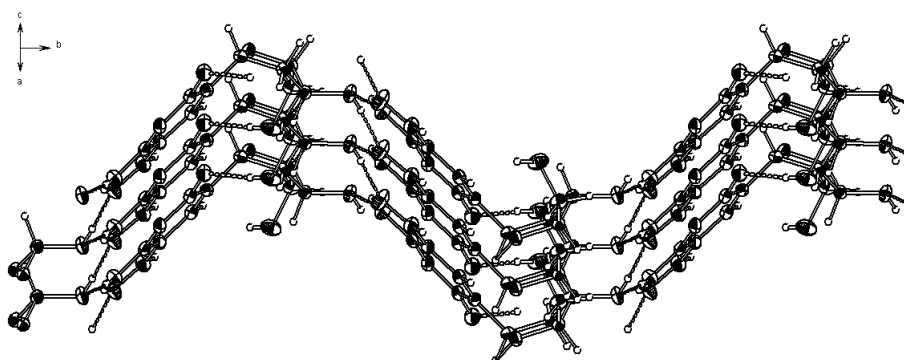
Figure S3: Crystal structure of fC, view of a puckered layer of hydrogen bonded molecules along the b-axis; thermal ellipsoids are drawn at 50% probability level.

Table S1: Hydrogen bonding parameters for fC in the crystal.

	D – H	H ··· A	D ··· A	D – H ··· A
N2 – H21 ··· O4[a)]	0.93(3)	1.92(3)	2.848(3)	173(3)
N2 – H22 ··· O1	0.83(3)	2.09(3)	2.745(3)	136(3)
O4 – H40 ··· O1[b)]	0.84(3)	1.86(3)	2.706(3)	176(3)
O5 – H50 ··· O2[c)]	0.86(3)	1.86(3)	2.719(3)	176(3)

Symmetry codes: [a)] 1-x, -0.5+y, 1-z; [b)] 2-x, -0.5+y, 2-z; [c)] -1+x, y, -1+z.

Table S2: Selected atom distances (in Å) in the molecular structure of fC in the crystal.

O1 – C5	1.242(3)	C2 – C4	1.371(4)
O2 – C1	1.244(3)	C3 – C4	1.446(4)
O3 – C6	1.420(3)	C4 – C5	1.435(4)
O3 – C9	1.450(3)	C6 – C7	1.525(4)
O4 – C8	1.424(3)	C7 – C8	1.517(4)
O6 – C10	1.421(3)	C8 – C9	1.531(4)
C9 – C10	1.513(4)	N1 – C2	1.339(3)
N1 – C6	1.509(3)	N1 – C1	1.429(3)
N2 – C3	1.331(4)	N3 – C3	1.335(3)
N3 – C1	1.344(3)		

Table S3: Selected bond angles (in °) in the molecular structure of fC in the crystal.

C6 – O3 – C9	110.3(2)	C6 – C9 – C10	115.2(2)
O3 – C9 – C8	104.4(2)	O3 – C9 – C10	110.3(2)
C1 – N1 – C2	120.5(2)	O5 – C10 – C9	112.0(2)
C1 – N1 – C6	117.3(2)	C2 – N1 – C6	122.2(2)
C1 – N3 – C3	120.0(2)	O2 – C1 – N3	123.4(2)
N1 – C1 – N3	119.6(2)	O2 – C1 – N1	117.0(2)
N1 – C2 – C4	121.3(2)	N2 – C3 – C4	121.3(2)
N3 – C3 – C4	122.1(2)	N2 – C3 – N3	118.5(2)
C2 – C4 – C3	116.6(2)	C2 – C4 – C5	118.4(2)
C3 – C4 – C5	125.0(2)	O1 – C5 – C4	124.0(3)
N1 – C6 – C7	112.0(2)	O3 – C6 – C7	106.5(2)
O3 – C6 – N1	108.3(2)	C6 – C7 – C8	102.1(2)
O4 – C8 – C9	113.5(2)	C7 – C8 – C9	101.7(2)
O4 – C8 – C7	110.4(2)		

Table S4: Details for X-ray data collection and structure refinement for compound fC

Empirical formula	$C_{10}H_{13}N_3O_5$
Formula mass	255.23
T[K]	173(2)
Crystal size [mm]	0.2×0.15×0.05
Crystal description	pale yellow block
Crystal system	monoclinic
Space group	$P2_1$
a [Å]	5.0787(4)
b [Å]	16.7550(12)
c [Å]	6.5815(5)
β [°]	103.791(9)
V [Å^3]	543.90(7)
Z	2
$\rho_{calcd.}$ [g cm^{-3}]	1.559
μ [mm^{-1}]	0.127
$F(000)$	268
Θ range [°]	4.30 – 33.51
Index ranges	$-4 \leq h \leq 7$
	$-24 \leq k \leq 21$
	$-10 \leq l \leq 9$
Reflns. collected	3819
Reflns. obsd.	2105
Reflns. unique	2933
	(R_{int} = 0.0322)
R_1, wR_2 (2σ data)	0.0502, 0.0799
R_1, wR_2 (all data)	0.0806, 0.0895
GOOF on F^2	0.956
Peak/hole [e Å^{-3}]	0.28/-0.24

Synthesis of compounds 1-11

5-formyl-3', 5'-(*O-tert*-butyl-dimethylsilyl)-2'deoxycytidine (8)
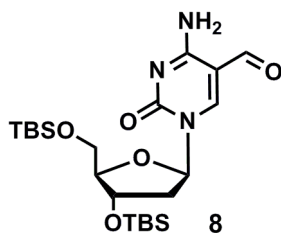
The reaction was carried out as described previously. The scale could be increased to 16 g. [2]

^{1}H-NMR (300 MHz, $CDCl_3$): δ (ppm) = 9.51 (s, 1H, C-C**H**=O), 8.57 (s, 1H, N-C**H**=C-CHO), 8.37 (s, 1H, 1x N**H**$_2$), 7.46 (s, 1H, 1x N**H**$_2$), 6.19 (t, ^3J = 6.1, 1H, O-C**H**-N), 4.40 – 4.32 (m, 1H, O-C**H**-CH$_2$-CH-N), 4.08 – 4.02 (m, 1H, O-CH-C**H**-CH$_2$-O), 3.95 (dd, ^3J = 2.7, ^2J = 11.7, 1H, 1x O-C**H**$_2$-CH-O), 3.78 (dd, ^3J = 2.6, ^2J = 11.6, 1H, 1x O-C**H**$_2$-CH-O), 2.59 (ddd, ^3J = 3.6, ^3J = 5.8, ^2J = 10.3, 1H, 1x O-CH-C**H**$_2$-CH-N), 2.20 – 2.08 (m, 1H, 1x O-CH-C**H**$_2$-CH-N), 0.89 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.88 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.10 (s, 3H, O-Si-C**H**$_3$), 0.08 (s, 6H, O-Si-C**H**$_3$), 0.07 (s, 3H, O-Si-C**H**$_3$).

^{13}C-NMR (75 MHz, $CDCl_3$): δ (ppm) = 187.1, 162.1, 153.1, 152.6, 104.9, 88.8, 87.9, 71.5, 62.6, 42.8, 25.9 (3C), 25.7 (3C), 18.4, 17.9, -4.5, -4.9, -5.2, -5.4.

HRMS (ESI+): calculated for $C_{22}H_{42}N_3O_5Si_2^+$ [M+H$^+$]$^+$: 484.2658, found: 484.2654

melting range: 150 - 152 °C

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3365 (w), 2952 (w), 2929 (w), 2857 (w), 1651 (s), 1245 (m), 1083 (s), 829 (s), 776 (s).

3',5'-(*tert*-butyl-dimethylsilyl)-4-*N*-benzoyl-5-formyl-2'deoxycytidine (2)
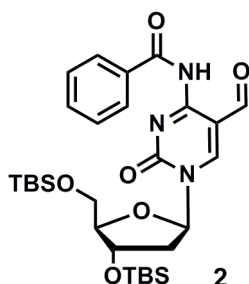
A round bottom flask was charged with 680 mg (1.41 mmol, 1.0 eq) **8** and 20 mL pyridine and was subsequently cooled to 0 °C. 875 µL (7.03 mmol, 5.0 eq) Bz-Cl were added slowly, the reaction mixture stirred at 0 °C for 30 min and for 2 h at room temperature. The reaction was stopped by addition of ice and diluted with 300 mL EtOAc. The solution was washed with sat. NaHCO$_3$ (1x 300 mL) and sat. NH$_4$Cl (1x 300 mL). The organic layer was separated, dried over MgSO$_4$ and subsequently evaporated to dryness. The crude product was purified by

column chromatography (*i*Hex/EtOAc 99:1→29:1→9:1) to give 435 mg (52 %) of **2** as a colorless oil along with 197 mg (20 %) of the double benzoylated derivative.

¹H-NMR (400 MHz, CDCl₃): δ (ppm) = 9.60 (br s, 1H, N**H**), 8.75 (s, 1H, C**H**O), 8.11 (ddd, ^4J = 2.8, ^3J = 3.9, ^3J = 8.5, 3H, Ph-**H**, N-C**H**=C-CHO), 7.63 – 7.44 (m, 3H, Ph-**H**), 6.22 (t, ^3J = 6.2, 1H, O-C**H**-N), 4.40 – 4.34 (m, 1H, O-C**H**-CH₂-CH-N), 4.10 (dd, ^3J = 2.6, ^3J = 5.7, 1H, O-CH-C**H**-CH₂-O), 3.97 (dd, ^3J = 2.6, ^2J = 11.6, 1H, 1x O-C**H**₂-CH-O), 3.80 (dd, ^3J = 2.5, ^2J = 11.6, 1H, 1x O-C**H**₂-CH-O), 2.71 – 2.61 (m, 1H, 1x O-CH-C**H**₂-CH-N), 2.13 (dt, ^3J = 6.1, ^2J = 13.5, 1H, 1x O-CH-C**H**₂-CH-N), 0.90 (2x s, 18H, O-Si-C(C**H**₃)₃), 0.11 (s, 3H, O-Si-C**H**₃), 0.10 (s, 3H, O-Si-C**H**₃), 0.09 (s, 3H, O-Si-C**H**₃), 0.07 (s, 3H, O-Si-C**H**₃).

¹³C-NMR (101 MHz, CDCl₃): δ (ppm) = 187.5, 174.7, 170.5, 159.4, 142.9, 133.6, 133.0, 130.1 (2C), 128.4 (2C), 91.6, 89.1, 88.4, 71.0, 62.6, 42.7, 25.9 (3C), 25.7 (3C), 18.4, 18.0, -4.5, -4.9, -5.3, -5.4.

HRMS (ESI+): calculated for $C_{29}H_{46}N_3O_6Si_2^+$ [M+H⁺]⁺: 588.2920, found: 588.2920.

IR (ATR): $\tilde{v}$ (cm⁻¹) = 2954 (w), 2929 (w), 2885 (w), 2856 (w), 1718 (m), 1692 (m), 1658 (s), 1567 (s), 1471 (s), 1250 (s), 1080 (s), 829 (s), 776 (s).

4-*N*-benzoyl-5-formyl-2'deoxycytidine (3)
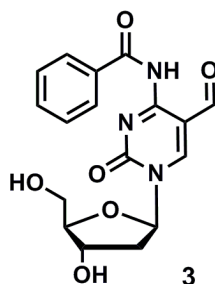
In a polypropylene tube 395 mg (0.67 mmol, 1.0 eq) of the TBS-protected nucleoside **2** were dissolved in dry EtOAc (final concentration 60 mM). Subsequently 272 µL (3.36 mmol, 5.0 eq) pyridine and 262 µL (10.1 mmol, 15 eq) HF (70 % in pyridine) were added and the reaction mixture stirred 14 h at rt. During this time a white solid precipitated. 1 mL/mmol TMSOMe were added and the reaction mixture stirred another 30 min. Subsequently the solid was collected by centrifugation (6000 rpm, 15 min). The crude product (purity >90 %) was not further purified and directly taken on to the next reaction. Yield: 198 mg (82 %).

¹H-NMR (400 MHz, pyridine-D₅/DMSO-D₆): δ (ppm) = 9.82 (br s, 1H, N**H**), 9.64 (s, 1H, C**H**O), 8.22 (d, ^3J = 7.1, 2H, Ph-**H**), 7.62 (s, 1H, N-C**H**=C-CHO), 7.55 (d, ^3J = 7.4, 1H, Ph-**H**), 7.48 (t, ^3J = 7.4, 2H, Ph-**H**), 6.60 (dd, ^3J = 4.9, ^3J = 6.3, 1H, O-C**H**-N), 4.90 (dd, ^3J = 5.8, ^3J = 10.5, 1H, O-C**H**-CH₂-CH-N), 4.46 (dt, ^3J = 2.9, ^3J = 4.7, 1H, O-CH-C**H**-CH₂-O), 4.22 (dd, ^3J = 3.0, ^2J = 12.0, 1H, 1x O-C**H**₂-CH-O), 4.12 (dd, ^3J = 2.9, ^2J = 12.0, 1H, 1x O-C**H**₂-CH-O), 2.94 (ddd, ^3J = 5.8, ^3J = 6.3, ^2J = 13.5, 1H, 1x O-CH-C**H**₂-CH-N), 2.69 (ddd, ^3J = 4.9, ^3J = 6.1, ^2J = 13.5, 1H, 1x O-CH-C**H**₂-CH-N).

¹³C-NMR (101 MHz, CDCl₃): δ (ppm) = 188.4, 159.1, 153.8, 152.2, 149.6, 134.1, 132.6, 128.6 (2C), 127.9 (2C), 106.5, 88.9, 87.9, 69.2, 60.5, 41.8.

HRMS (ESI+): calculated for $C_{17}H_{18}N_3O_6^+$ [M+H⁺]⁺: 360.1190, found: 360.1191.

Melting Range: >200 °C decomposition.

IR (ATR): $\tilde{\nu}$ (cm⁻¹) = 3477 (br, m), 3306 (br, m), 2988 (w), 2931 (w), 2906 (w), 2851 (w), 1721 (s), 1650 (s), 1570 (s), 1473 (s), 1251 (s), 1238 (s), 1092 (s), 795 (m).

5'-(dimethoxytrityl)-4-*N*-benzoyl-5-formyl-2'deoxycytidine (9)
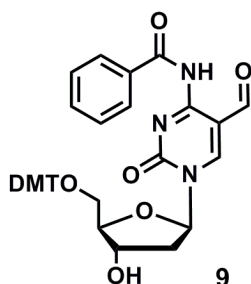
A round bottom flask was charged with rigorously dried 3 Å molecular sieves. 146 mg (0.41 mmol, 1.0 eq) of the nucleoside **3** (final concentration: 100 mM) and 151 mg (0.45 mmol, 1.1 eq) DMT-Cl were dissolved in pyridine. The reaction mixture was stirred for 14 h and the molecular sieves subsequently filtered off. The filtrate was evaporated to dryness and the crude product purified by column chromatography. Eluent for column chromatography: DCM/MeOH (49:1; 0.1 % NEt₃). Yield: 170 mg (63 %) of **9** as a pale yellow solid.

¹H-NMR (400 MHz, DMSO-D₆): δ (ppm) = 8.97 (s, 1H, C-C**H**=O), 8.81 (s, 1H, N-C**H**=C-CHO), 8.00 – 7.89 (m, 2H, Ph-**H**), 7.70 (t, ^3J = 7.3, 1H, Ph-**H**), 7.61 (t, ^3J = 7.3, 2H, Ph-**H**), 7.33 (m, 4H, DMT-**H**), 7.27 – 7.18 (m, 5H, DMT-**H**), 6.88 (dd, ^4J = 0.7, ^3J = 9.1, 4H, DMT-**H**), 6.09 (dd, ^3J = 5.1, ^3J = 6.4, 1H, O-C**H**-N), 5.42 (d, ^3J = 4.8, 1H, O**H**), 4.34 – 4.26 (m, 1H, O-C**H**-CH₂-CH-N), 4.07 (dd, ^3J = 4.3, ^3J = 7.9, 1H, O-CH₂-C**H**-CH-O), 3.71 (2x s, 6H, O-C**H₃**), 3.34 – 3.26 (m, 2H, O-C**H₂**-CH-O), 2.49 – 2.43 (m, 1H, 1x O-CH-C**H₂**-CH-N), 2.34 (ddd, ^3J = 4.5, ^3J = 6.3, ^2J = 11.3, 1H, 1x O-CH-C**H₂**-CH-N).

¹³C-NMR (101 MHz, DMSO-D₆): δ (ppm) = 189.0, 163.8, 159.3, 158.2 (2C), 154.3, 152.7, 144.5, 135.4, 135.2, 133.3, 133.2, 129.77 (2C), 129.76 (2C), 129.1 (4C), 128.0 (2C), 127.7 (2C), 126.9, 114.1 (2C), 113.3 (2C), 105.8, 87.7, 86.4, 86.0, 69.4, 63.0, 55.0 (2C), 40.7.

HRMS (ESI+): calculated for $C_{38}H_{36}N_3O_8^+$ [M+H⁺]⁺: 662.2497, found: 662.2500.

Melting Range: 115 – 123 °C

IR (ATR): $\tilde{\nu}$ (cm⁻¹) = 2934 (w), 1716 (m), 1650 (s), 1601 (m), 1565 (s), 1508 (m), 1470 (s), 1247 (s), 1174 (m), 1089 (m), 1068 (m), 1031 (m), 827 (m), 790 (m), 701 (s).

3'-(di*iso*propylamino-cyanoethoxyphosphino)-5'-(dimethoxytrityl)-4-*N*-benzoyl-5-formyl-2'deoxycytidine (4)
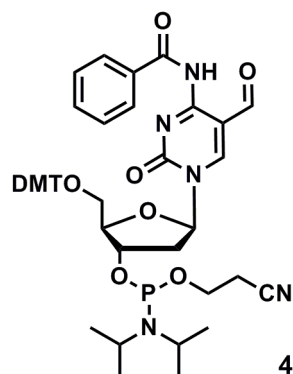
In a Schlenck tube 100 mg (0.15 mmol, 1.0 eq) of DMT-protected nucleoside **9**, 14 mg (0.13 mmol, 0.5 eq) di*iso*propyltetrazolide, and 55 µL (0.18 mmol, 1.2 eq) bis(di*iso*propylamino)(2-cyanoethoxy)phosphine, were dissolved in 10 mL rigorously degassed DCM and the solution degassed three more times (freeze, pump, thaw). The reaction mixture was stirred at rt for 15 h and subsequently concentrated in an argon atmosphere. The crude product was purified by column chromatography. The test tubes that were used to collect the fractions were flushed with argon prior to use. Eluent for column chromatography: DCM/MeOH (49:1, 0.1 % NEt₃). Yield: 113 mg (95 %, colorless foam) of **4** as a mixture of two diastereomers on P.

³¹P-NMR (81 MHz, Acetone): δ (ppm) = 149.64, 149.45.

HRMS (ESI+): calculated for $C_{47}H_{53}N_5O_9P^+$ [M+H]⁺: 862.3575, found: 862.3580.

3',5'-(*tert*-butyl-dimethylsilyl)-5-carboxymethyl-2'deoxycytidine (10)
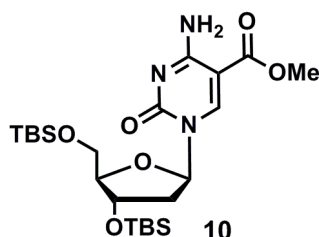
In a glass autoclave 1.00 g (1.72 mmol, 1.0 eq) **1** and 49 mg (0.19 mmol, 0.1 eq) Pd(MeCN)₂Cl₂ were dissolved in 20 mL methanol and subsequently 600 µL (3.44 mmol, 2.0 eq) DIPEA added. The autoclave was flushed once with CO and the reaction mixture then stirred for 18 h at 60 °C with a CO pressure of 3 bar. After completion of the reaction the gas was discharged and the solvent evaporated *in vacuo*. The crude product was purified by column chromatography (DCM/MeOH 99:1, dry loaded) to give 696 mg (84 %) of **10** as a yellowish solid.

^{1}H-NMR (300 MHz CDCl$_3$): δ (ppm) = 8.59 (s, 1H, N-C**H**=C-COOMe), 7.93 (s, 1H, 1x N**H**$_2$), 7.31 (s, 1H, 1x N**H**$_2$), 6.15 (dd, J=5.9, 7.3, 1H, O-C**H**-N), 4.36 – 4.29 (m, 1H, O-C**H**-CH$_2$-CH-N), 4.03 (dd, ^3J = 2.9, ^3J = 5.6, 1H, O-CH$_2$-C**H**-CH-O), 3.83 (dd, ^3J = 3.2, ^2J = 11.3, 1H, 1x O-C**H**$_2$-CH-O), 3.79 (s, 3H, C-CO-O-C**H**$_3$), 3.74 (dd, ^3J = 3.1, ^2J = 11.3, 1H, 1x O-C**H**$_2$-CH-O), 2.58 (ddd, ^3J = 2.4, ^3J = 5.8, ^2J = 13.4, 1H, 1x O-CH-C**H**$_2$-CH-N), 1.94 (ddd, ^3J = 6.1, ^3J = 7.3, ^2J = 13.4, 1H, 1x O-CH-C**H**$_2$-CH-N), 0.86 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.85 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.05 (s, 6H, Si-C**H**$_3$), 0.04 (s, 6H, Si-C**H**$_3$).

^{13}C-NMR (75 MHz, CDCl$_3$): δ (ppm) = 165.1, 163.7, 154.0, 147.4, 95.1, 88.7, 87.7, 72.7, 63.0, 51.8, 42.8, 25.8 (3C), 25.7 (3C), 18.2, 17.9, -4.7, -5.0, -5.6, -5.7.

HRMS (ESI+): calculated for C$_{23}$H$_{44}$N$_3$O$_6$Si$_2$$^+$ [M+H$^+$]$^+$: 514.2763, found: 514.2762

IR (ATR): $\tilde{v}$ (cm^{-1}) = 2930 (m), 2857 (m), 1720 (s), 1674 (s), 1501 (m), 1101 (s), 1075 (s), 822 (s), 776 (s).

melting range: 129 - 131 °C

3',5'-(*tert*-butyl-dimethylsilyl)-4-*N*-acetyl-5-carboxymethyl-2'deoxycytidine (5)

In a round bottom flask 550 mg (1.07 mmol, 1.0 eq) **57** were dissolved in 50 mL THF. Subsequently 7 mL (53 mmol, 50 eq) Ac$_2$O, 87 μL (1.07 mmol, 1.0 eq) pyridine and 7 mg (0.05 mmol, 0.1 eq) DMAP were added. The mixture was stirred for 15 h at rt. After completion of the reaction 5 mL MeOH were added and the solution was concentrated *in vacuo* to give 585 mg (98 %) of **5** as a yellowish solid.

^{1}H-NMR (599 MHz, CDCl$_3$): δ (ppm) = 10.71 (s, 1H, N**H**), 8.80 (s, 1H, N-C**H**=C-COOMe), 6.13 (dd, ^3J = 6.1, ^3J = 7.1, 1H, O-C**H**-N), 4.35 (dt, ^3J = 2.2, ^3J = 5.5, 1H, O-C**H**-CH$_2$-CH-N), 4.11 (q, ^3J = 2.7, 1H, O-CH$_2$-C**H**-CH-O), 3.88 – 3.84 (m, 4H, CO-O-C**H**$_3$, 1x O-C**H**$_2$-CH-O), 3.77 (dd, ^3J = 2.9, ^2J = 11.4, 1H, 1x O-C**H**$_2$-CH-O), 2.69 (ddd, ^3J = 2.3, ^3J = 5.9, ^2J = 13.5, 1H, 1x O-CH-C**H**$_2$-CH-N), 2.64 (s, 3H, N-CO-C**H**$_3$), 1.99 (ddd, ^3J = 6.0, ^3J = 7.3, ^2J = 13.4, 1H, 1x O-CH-C**H**$_2$-CH-N), 0.89 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.84 (s, 9H, O-Si-C(C**H**$_3$)$_3$), 0.08 (s, 3H, Si-C**H**$_3$), 0.07 (s, 3H, Si-C**H**$_3$), 0.06 (s, 3H, Si-C**H**$_3$), 0.04 (s, 3H, Si-C**H**$_3$).

^{13}C-NMR (151 MHz, CDCl$_3$): δ (ppm) = 172.0, 164.9, 159.6, 153.1, 148.7, 95.7, 89.4, 88.7, 72.9, 63.1, 52.4, 42.9, 27.2, 25.8 (3C), 25.7 (3C), 18.2, 18.0, -4.6, -4.9, -5.6, -5.7.

HRMS (ESI+): calculated for C$_{25}$H$_{46}$N$_3$O$_7$Si$_2$$^+$ [M+H$^+$]$^+$: 556.2869, found: 556.2866
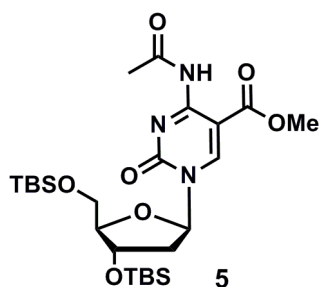
Melting Range: 132 - 134 °C

IR (ATR): $\tilde{v}$ (cm^{-1}) = 2927 (w), 2857 (w), 1644 (s), 1537 (m), 1284 (s), 1255 (s), 1119 (s), 1006 (m), 817 (s), 761 (s).

4-*N*-acetyl-5-carboxymethyl-2'deoxycytidine (6)
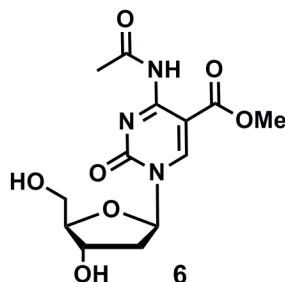
In a polypropylene tube 473 mg (0.85 mmol, 1.0 eq) of the TBS-protected nucleoside **5** were dissolved in dry EtOAc (final concentration 60 mM). Subsequently 332 μL (12.8 mmol, 5.0 eq) pyridine and 343 μL (4.26 mmol, 15 eq) HF (70 % in pyridine) were added and the reaction mixture stirred 14 h at rt. During this time a white solid precipitated. 1 mL/mmol TMSOMe were added and the reaction mixture stirred another 30 min. Subsequently the solid was collected by centrifugation (6000 rpm, 15 min). The crude product was further purified by column chromatography (DCM/MeOH 19:1) to give 212 mg (76 %) of **6** as a colorless solid.

^{1}H-NMR (400 MHz, CD$_3$OD): δ (ppm) = 9.12 (s, 1H, N-C**H**=C-COOMe), 6.18 (t, ^3J = 6.0, 1H, O-C**H**-N), 4.38 (dt, ^3J = 4.6, ^3J = 6.2, 1H, O-C**H**-CH$_2$-CH-N), 3.99 (dd, ^3J = 3.4, ^3J = 7.5, 1H, O-CH$_2$-C**H**-CH-O), 3.89 – 3.82 (m, 4H, CO-O-C**H**$_3$, 1x O-C**H**$_2$-CH-O), 3.76 (dd, ^3J = 3.4, ^2J = 12.0, 1H, 1x O-C**H**$_2$-CH-O), 2.47 (ddd, ^3J = 4.8, ^3J = 6.4, ^3J = 13.7, 1H, O-CH-C**H**$_2$-CH-N), 2.26 – 2.16 (m, 1H, O-CH-C**H**$_2$-CH-N), 2.02 (s, 3H, N-CO-C**H**$_3$).

^{13}C-NMR (101 MHz, CD$_3$OD): δ (ppm) = 173.6, 166.8, 165.4, 156.7, 150.2, 97.2, 89.4, 88.7, 71.4, 62.3, 52.6, 42.8, 20.6.

HRMS (ESI+): calculated for C$_{13}$H$_{18}$N$_3$O$_7$$^+$ [M+H$^+$]$^+$: 328.1139, found: 328.1138

IR (ATR): $\tilde{v}$ (cm^{-1}) = 3305 (br, m), 2928 (w), 1634 (s), 1584 (m), 1484 (m), 1437 (s), 1326 (m), 1296 (s), 1255 (s), 1101 (s), 792 (s).

melting range: > 200 °C decomposition

5'-(dimethoxytrityl)-4-*N*-acetyl-5-carboxymethyl-2'deoxycytidine (11)
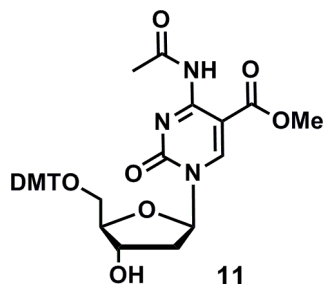
A round bottom flask was charged with rigorously dried 3 Å molecular sieves. 173 mg (0.53 mmol, 1.0 eq) of the nucleoside **6** (final concentration: 66 mM) and 185 mg (0.53 mmol, 1.0 eq) DMT-Cl were dissolved in pyridine. The reaction mixture was stirred for 14 h and the molecular sieves subsequently filtered off. The filtrate was evaporated to dryness and the crude product purified by column chromatography. Eluent for column chromatography: DCM/MeOH (49:1). Yield: 158 mg (48 %) of **11** as a colorless solid.

^{1}H-NMR (400 MHz, CD$_3$CN): δ (ppm) = 10.69 (s, 1H, N**H**), 8.89 (s, 1H, N-C**H**=C-COOMe), 7.46 – 7.41 (m, 2H, DMT), 7.35 – 7.21 (m, 7H, DMT), 6.88 – 6.82 (m, 4H, DMT), 6.05 (t, ^3J = 5.9, 1H, O-C**H**-N), 4.35 (m, 1H, O-C**H**-CH$_2$-CH-N), 4.05 (dd, ^3J = 3.9, ^3J = 7.5, 1H, O-CH$_2$-C**H**-CH-O), 3.76 (s, 6H, CH=C-O-C**H**$_3$), 3.44 – 3.38 (m, 1H, 1x O-C**H**$_2$-CH-O), 3.23 (dd, ^3J = 3.9, ^2J = 10.9, 1H, 1x O-C**H**$_2$-CH-O), 3.18 (s, 3H, CO-O-C**H**$_3$), 2.56 – 2.47 (m, 4H, 1x O-CH-C**H**$_2$-CH-N, N-CO-C**H**$_3$), 2.34 – 2.26 (m, 1H, 1x O-CH-C**H**$_2$-CH-N).

^{13}C-NMR (101 MHz, CD$_3$CN): δ (ppm) = 172.0, 166.5, 160.7, 159.82, 159.81, 154.2, 150.1, 145.9, 136.94, 136.87, 131.08 (2C), 131.04 (2C), 129.1 (2C), 128.9 (2C), 128.0, 114.1 (4C), 97.1, 88.9, 87.8, 87.4, 71.1, 63.7, 56.0 (2C), 52.8, 42.3, 27.1.

HRMS (ESI+): calculated for C$_{34}$H$_{36}$N$_3$O$_9^+$ [M+H$^+$]$^+$: 630.2446, found: 630.2443

IR (ATR): $\tilde{\nu}$ (cm^{-1}) = 3277 (m), 3223 (m), 2952 (m), 2935 (w), 2837, 1659 (s), 1547 (m), 1506 (s), 1439 (m), 1329 (m), 1296 (s), 1255 (s), 1174 (s), 1098 (s), 837 (m), 792 (s), 701 (s).

melting range: 97 - 98 °C

3'-(di*iso*propylamino-cyanoethoxyphosphino)-5'-(dimethoxytrityl)-5-carboxymethyl-2'deoxycytidine (7)
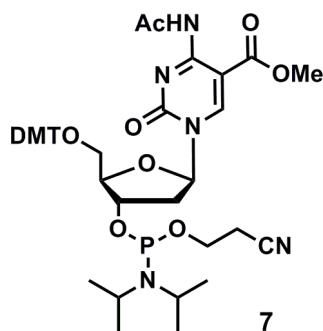
In a Schlenck tube 158 mg (0.25 mmol, 1.0 eq) of DMT-protected nucleoside **11** (final concentration: 50 mM), 22 mg (0.13 mmol, 0.5 eq) di*iso*propyltetrazolide, and 98 μL (0.30 mmol, 1.2 eq) bis(di*iso*propylamino)(2-cyanoethoxy)phosphine, were dissolved in 10 mL rigorously degassed DCM and the solution degassed three more times (freeze, pump, thaw). The reaction mixture was stirred at rt for 15 h and subsequently concentrated in an argon atmosphere. The crude product was purified by column chromatography. The test tubes that were used to collect the fractions were flushed with argon prior to use. Eluent for column chromatography: DCM/MeOH (49:1, 0.1 % NEt$_3$). Yield: 196 mg (94 %, colorless foam) of **7** as a mixture of two diastereomers on P.

The compound was directly used for solid phase DNA synthesis. Its identity was proven by successful incorporation into DNA.

^{31}P-NMR (81 MHz, Acetone): δ (ppm) = 149.46, 149.32.

HRMS (ESI+): calculated for $C_{43}H_{53}N_5O_{10}P^+$ [M+H$^+$]$^+$: 830.3525, found: 830.3530.

[1]   S. Schorr, S. Schneider, K. Lammens, K.-P. Hopfner, T. Carell, *Proceedings of the National Academy of Sciences* **2010**, *107*, 20720-20725.
[2]   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) M. Münzel, D. Globisch, C. Trindler, T. Carell, *Organic Letters* **2010**, *12*, 5671-5673.

8.3 Zusatzmaterialien zu Abschnitt 3.3

Stefan Schiesser, Benjamin Hackner, <u>Toni Pfaffeneder</u>, Markus Müller, Christian Hagemeier, Matthias Truss und Thomas Carell, *Angew. Chem.* **2012***, 124*, 6622–6626; *Angew. Chem., Int. Ed.* **2012**, *51*, 6516–6520. *Mechanism and Stem-Cell Activity of 5-Carboxycytosine Decarboxylation Determined by Isotope Tracing.*

Supporting Information

General methods:

All non-aqueous reactions were performed using flame- or oven-dried glassware under an atmosphere of dry nitrogen. Commercial reagents from *Sigma-Aldrich* or *Acros* were used as received unless otherwise noted. Non-aqueous reagents were transferred under nitrogen with a syringe or cannula. Solutions were concentrated *in vacuo* on a *Heidolph* rotary evaporator with a *Vario PC2001* diphragm pump by *Vacuubrand*. Chromatographic purification of products was accomplished using flash column chromatography on *Merck* Geduran Si 60 (40-63 µM) silica gel (normal phase) or *Fluka* silica gel 100 C_{18}-Reversed phase (15-35 µm). Thin layer chromatography (TLC) was performed on *Merck* 60 (silica gel F_{254}) plates. Visualization of the developed chromatogram was performed using fluorescence quenching or anisaldehyde staining. ^{1}H, ^{13}C and ^{15}N NMR spectra were recorded in deuterated solvents on *Bruker ARX 300*, *Varian VXR400S*, *Varian Inova 400* and *Bruker AMX 600* spectrometers and calibrated to the residual solvent peak. Routine ESI spectra and high-resolution ESI spectra were obtained on the mass spectrometer *Thermo Finnigan* LTQ FT-ICR. Acetonitrile for HPLC-ESI-MS analysis was purchased from *VWR*, HPLC gradient grade. HCOOH was purchased from *Fluka*, p.a. for mass spectrometry. Molecular sieves were rigorously dried in high vacuum at 350 °C over night. IR measurements were performed on a *Perkin Elmer Spectrum BX FT-IR* spectrometer (*Perkin Elmer*) with a diamond-ATR (Attenuated Total Reflection) setup. Melting points were determined with a *Büchi Melting Point B540*. The concentration of the synthesized DNA was determined with a *NanoDrop ND-1000* spectrophotometer (*Peqlab*). Extinction coefficients at 260 nm were calculated by addition of the extinction coefficients of the individual nucleobases. These are dA 15.0 L/mmol·cm, dC 7.1 L/mmol·cm, dG 12.0 L/mmol·cm and dT 8.4 L/mmol·cm. For the dC derivatives, the dC value was used. Nuclease S1 (*Aspergillus oryzae*) was obtained from *Roche*, snake venom phosphodiesterase I (*Crotalus adamanteus*) from *USB corporation* and antarctic phosphatase from *New England Biolabs*.

High performance liquid chromatography:

The following devices were used for the purification and analysis of the synthetic DNA strands. HPLC was performed on *Waters* or *Merck-Hitachi* units. These were in detail: analytical HPLC *Waters* Alliance (2695 Separation Module, 2996 Photodiode Array Detector), *Merck* analytical (L-7400 UV detector, L-7100 pump), preparative HPLC: *Waters* Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump), *Merck* preparative (L-7150 pump, L-7420 UV detector, *Rheodyne* P/N 77 25i injection valve, ERC-3415 solvent

degasser). For analytical HPLC Nucleosil 120-3 C18 from *Macherey Nagel* were used, for preparative HPLC Nucleosil 100-7 C18 also from *Macherey Nagel*. For preparative HPLC Nucleosil VP 250/10 C18 and VP 250/32 C18 from *Macherey Nagel* were used.

Buffer systems:

buffer system I: analytical and preparative reversed phase HPLC

Buffer A: 0.1 M NHEt$_3$OAc

Buffer B: 0.1 M NHEt$_3$OAc in 80% MeCN

Buffer C: H$_2$O

Buffer D: MeCN

buffer system II: LC-MS

Buffer A: 0.01% FA in H$_2$O

Buffer B: 0.01% FA in 100% MeCN

LC-MS instruments:

LC-ESI-MS with high resolution was carried out on a *Thermo Finnigan* LTQ Orbitrap XL coupled to a *Dionex* Ultimate 3000 HPLC system.

HPLC-ESI-analysis using an Orbitrap XL mass spectrometer:

Samples (80-100 µL injection volume) for the isotope-tracing experiments in nuclear extracts were subjected to a *Thermo Finnigan* LTQ Orbitrap XL and chromatographed by a *Dionex* Ultimate 3000 HPLC system with a flow of 0.15 mL/min over an Uptisphere120-3HDO (3 µm, 2.1 mm x 15 mm) column from *Interchim* (buffer system II). The flow rate was 150 µL min^{-1} and the column temperature was maintained at 30 °C. The gradient was 0 → 12 min; 0% → 1% buffer B; 12 → 20 min; 1% → 2% buffer B; 20 → 30 min; 2% → 10% buffer B; 30 → 35 min; 10% → 80% buffer B; 35 → 45 min; 80% buffer B; 45 → 50 min; 80% → 0% buffer B; 50 → 60 min; 0% buffer B. The elution was monitored at 260 nm (*Dionex* Ultimate 3000 Diode Array Detector). The effluent up to 5 min (total run time of 60 min) was diverted to waste by a Valco valve in order to protect the mass spectrometer. The chromatographic effluent was then directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of *m/z* 100-500 with a resolution of 60000. 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, 275 °C; capillary voltage, 35 V, tube lens 65 V.

Preparation of nuclear extracts:

Intact nuclei were prepared with a slighty modified Dignam method, after cell counting.[1-2] B*riefly the cell pellet was resuspended in 5 volumes buffer A (10 mM HEPES pH7.9, 1.5mM MgCl₂, 10mM KCl, 0.5 mM fresh DTT, complete protease Inhibitor (Roche)) incubated on ice for 10 min, pelleted (2000 rpm, 15min) and resuspended in 2 volumes buffer A. Homogenzing was achieved using a douncer (25 strokes) and lysis was checked with trypan blue staining. Nuclei were pelleted again (2000 rpm, 15 min) to remove residual cytoplasmatic lysate, resuspended in Buffer A and aliquoted.*

Demethylation assay:

For the demethylation assay 660 pmol of ODN1 (5´-CCT-TTC-CGA-AGG-GAC-GTT-GAC-caCGG-ATG-CYC-3´; Y = Biotinylated-dU) and ODN2 (5´GAG-CAT-CCG-GTC-AAC-GTC-CCT-TcaCG-GAA-AGG-3´) were hybridized and evaporated to dryness. Then freshly prepared mESC nuclear extract or non mESC nuclear extract were added (180 µl) to the DNA and incubated for 1 h at 37°C 500 rpm. After incubation the DNA was reisolated using Streptavidin coated magnetic beads (M-270, *Invitrogen*) according to the manufactures instructions using a 2 fold excess of biotinylated DNA. Additional washing steps included 2 min incubation with a low salt- (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris HCl pH 8.0, 150 mM NaCl) and high salt-buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris HCl pH 8.0, 500 mM NaCl) as well as ddH₂O, each 3 times. Then the DNA was removed from the beads using a method established by Holmberg and others.[3] Briefly the streptavidin-biotin interaction was broken using ddH₂O elevating from 22-99° C. Then the DNA was digested using the protocol described below without prior spiking of the samples.

Enzymatic Digestion of DNA for LC-MS analysis

The enzymatic digestion was performed as previously reported.[4]

Synthesis of [^{15}N₂]-caC-phosphoramidite (10)

^{15}N₂-Uracil (1)

The compound was synthesized with slight modifications according to literature procedure.[5] A flame dried round bottom flask was charged with 2.10 g (33.85 mmol, 1.0 eq.) ^{15}N-urea and 62 g polyphosphorous acid. After addition of 2.39 mL (34.19 mmol, 1.0 eq.) propiolic acid, the reaction mixture was stirred at 95 °C for 19 h. After cooling to room temperature 45 mL H₂O were added and the mixture was stored at -20 °C for 9 h. The precipitated solid was filtrated, washed with cold water and was dried several times to give **1** (1.20 g, 31 %) as a light brown solid. **^{1}H-NMR** (400 MHz, CDCl₃, ppm): δ = 10.98 (dd, ^3J=1.6, 1H, (C=O)-N-**H**-

(C=O)), 10.79 (dd, ^3J=1.7, 1H, (C=O)-N$\underline{\text{H}}$-CH), 7.40- 7.36 (m, 1H, NH-C$\underline{\text{H}}$=CH), 5.46-5.52 (m, 1H, CH=C$\underline{\text{H}}$-(C=O)). ^{13}**C-NMR** (151 MHz, CDCl$_3$, ppm): δ = 164.3 (CH-($\underline{\text{C}}$=O)-NH), 151.5 (NH-($\underline{\text{C}}$=O)-NH), 142.1 (NH-$\underline{\text{C}}$H=CH), 100.2 (CH=$\underline{\text{C}}$H-(C=O). ^{15}**N-NMR** (40 MHz, d^6-DMSO, ppm): δ = -222.0 (d, J=2.7 Hz), -248.4 (d, J=2.7 Hz). **HRMS (EI+):** calc. for C$_4$H$_4$^{15}N$_2$O$_2$ [M+H$^+$]: 114.0208, found: 114.0207. **Melting range:** 323-325 °C. **IR (ATR):** v (cm^{-1}) = 3080 (w), 2916 (w), 2156 (m), 2003 (m), 1973 (m), 1956 (m), 1716 (s), 1660 (s), 1411 (m), 1229 (m), 820 (s), 755 (s).

3´,5´-(Toluoyl)-2´-deoxy-(N^1,N^3-^{15}N)-uridine (2)

To a stirred solution of ^{15}N$_2$-uracil (**1**) (1.10 g, 9.64 mmol, 1.0 eq.) in 30 mL HMDS was added TMS-Cl (4.93 mL, 38.57 mmol, 4.0 eq.) and heated to 120 °C for 1 h. The clear solution was evaporated to dryness and coevaporated with dry CHCl$_3$ two times. The solid was diluted in 30 ml dry CHCl$_3$ and 4.12 g (10.61 mmol, 1.1 eq.) Hoffers chlorosugar were added. The chlorosugar was synthesized according to Clever and co-workers.[6] The reaction mixture was stirred for 2.5 h at room temperature and diluted with 300 mL DCM. The solution was washed with sat. NH$_4$Cl (2 × 300 mL). The organic layer was separated, dried over MgSO$_4$ and subsequently evaporated to dryness. The crude product was purified by coevaporation with EtOAc and filtration from EtOAc to give 2.28 g (51 %) of **2** as a colorless solid. ^{1}**H-NMR** (600 MHz, CDCl$_3$, ppm): δ = 8.71 (d, 1H, N-$\underline{\text{H}}$), 7.94 (d, ^3J=8.2, 2H, C-C$\underline{\text{H}}$=CH-C-CH$_3$), 7.89 (d, ^3J=8.2, 2H, C-C$\underline{\text{H}}$=CH-C-CH$_3$), 7.52 (d, ^3J=8.2, 1H, N-C$\underline{\text{H}}$=CH-C), 7.27 (d, ^3J=8.2, 2H, C-CH=C$\underline{\text{H}}$-C-CH$_3$), 7.25 (d, ^3J=8.2, 2H, C-CH=C$\underline{\text{H}}$-C-CH$_3$), 6.40 (t, 1H, N-C$\underline{\text{H}}$-O), 5.61-5.56 (m, 2H, O-C$\underline{\text{H}}$-CH$_2$-CH-N, N-CH=C$\underline{\text{H}}$-C), 4.70 (qd, ^3J=3.3, ^2J=12.3, 2H, O-C$\underline{\text{H}}_2$-CH-O), 4.53 (q, 1H, O-CH$_2$-C$\underline{\text{H}}$-O), 2.74 (dd, ^3J=6.5, ^2J=14.9, 1H, O-CH-C$\underline{\text{H}}_2$-CH-N), 2.42 (s, 6H, C-CH=CH-C-C$\underline{\text{H}}_3$) 2.33-2.25 (m, 1H, O-CH-C$\underline{\text{H}}_2$-CH-N). ^{13}**C-NMR** (151 MHz, CDCl$_3$, ppm): δ = 166.0 (2C, O-($\underline{\text{C}}$=O)-C), 162.6 (C-($\underline{\text{C}}$=O)-NH), 144.6 (N-($\underline{\text{C}}$=O)-N), 138.9 (N-$\underline{\text{C}}$=C), 138.8 (2C, H$_3$C-$\underline{\text{C}}$=C), 129.4 (4C, $\underline{\text{C}}$=C-(C=O)-O), 129.3 (4C, H$_3$C-C=$\underline{\text{C}}$-C), 126.3 (2C, O-(C=O)-$\underline{\text{C}}$-C), 102.9 (N-CH-$\underline{\text{C}}$-(C=O)), 85.3 (O-$\underline{\text{C}}$H-N), 83.0 (O-$\underline{\text{C}}$H-CH$_2$-O), 74.6 (O-$\underline{\text{C}}$H-CH$_2$-CH-N), 64.0 (O-$\underline{\text{C}}$H$_2$-CH-O), 38.4 (N-CH-$\underline{\text{C}}$H$_2$-CH-O), 21.7 (2C, $\underline{\text{C}}$H$_3$-C=C). ^{15}**N-NMR** (40 MHz, CDCl$_3$, ppm): δ = -224.6 (d, J=3.0 Hz), -236.6 (d, J=3.0 Hz). **HRMS (ESI+):** calc. for C$_{25}$H$_{25}$^{15}N$_2$O$_7$ [M+H]$^+$: 467.1597, found: 467.1598. **Melting range:** 214-215 °C. **IR (ATR):** v (cm^{-1}) = 2156 (w), 1710 (s), 1682 (m), 1450 (m), 1365 (m), 1278 (s), 1122 (m), 1041 (m), 803 (w), 765 (s).

3´,5´-(Toluoyl)-5-iodo-2´-deoxy-(N^1,N^3-^{15}N)-uridine (3)

The compound was synthesized with slight modifications according to Asakura.[7] A round bottom flask was charged with 2.28 g (4.89 mmol, 1.0 eq.) **2**, 5.36 g (9.78 mmol, 2.0 eq.) dried CAN, 0.79 g (5.87 mmol, 1.2 eq.) dried LiI and 60 mL dry MeCN. The reaction mixture was stirred at 80 °C for 2 h and stopped by addition of ice and 100 mL brine. The mixture was diluted with 200 mL EtOAc and 50 mL $CHCl_3$. The aqueous layer was extracted with DCM several times. The combined organic layers were washed with brine, dried over $MgSO_4$ and subsequently evaporated to dryness. The crude product was purified by column chromatography (DCM/MeOH 70:1) to give 2.31 g (80 %) of **3** as a colorless solid. **^{1}H-NMR** (600 MHz, $CDCl_3$, ppm): δ = 8.68 (d, 1H, N-**H**), 7.96-7.92 (m, 4H, C-C**H**=CH-C-CH$_3$), 7.27 (d, ^3J=9.4, 4H, C-CH=C**H**-C-CH$_3$), 6.36 (t, 1H, N-C**H**-O), 5.61 (d, ^3J=6.5, 1H, O-C**H**-CH$_2$-CH-N), 4.78-4.70 (m, 2H, O-C**H$_2$**-CH-O), 4.57 (q, 1H, O-CH$_2$-C**H**-O), 2.77 (dd, ^3J=6.1, ^2J=14.5, 1H, O-CH-C**H$_2$**-CH-N), 2.42 (s, 6H, C-CH=CH-C-C**H$_3$**) 2.31-2.23 (m, 1H, O-CH-C**H$_2$**-CH-N). **^{13}C-NMR** (151 MHz, $CDCl_3$, ppm): δ = 166.0 (2C, O-(**C**=O)-C), 159.5 (C-(**C**=O)-NH), 144.6 (N-(**C**=O)-N), 143.6 (N-**C**=C-I), 143.5 (2C, H$_3$C-**C**=C), 129.6 (4C, **C**=C-(C=O)-O), 129.4 (4C, H$_3$C-C=**C**-C), 126.2 (2C, O-(C=O)-**C**-C), 85.8 (O-**C**H-N), 83.4 (O-**C**H-CH$_2$-O), 74.9 (O-**C**H-CH$_2$-CH-N), 68.9 (N-CH-**C**-(C=O)), 64.1 (O-**C**H$_2$-CH-O), 38.7 (N-CH-**C**H$_2$-CH-O), 21.8 (2C, **C**H$_3$-C=C). **^{15}N-NMR** (41 MHz, $CDCl_3$, ppm): δ = -228.2, -229.8. **HRMS** (ESI+): calc. for $C_{25}H_{24}I^{15}N_2O_7$ $[M+H]^+$: 593.0563, found: 593.0570. **Melting range:** 196-197 °C. **IR (ATR):** ν (cm^{-1}) = 3644 (w), 2472 (w), 2142 (s), 2004 (s), 1973 (s), 1681 (m), 1446 (m), 1255 (w), 1777 (w), 700 (w).

3´,5´-(Toluoyl)-5-iodo-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (4)

A round bottom flask was charged with 2.26 g (32.78 mmol, 9.0 eq.) Triazol and 60 mL dry MeCN. The mixture was cooled to 0 °C and 0.68 mL (7.28 mmol, 2.0 eq.) $POCl_3$ were added. After stirring at 0 °C for 10 min, 4.37 mL (31.32 mmol, 8.6 eq.) NEt_3 were added dropwise and the reaction mixture was stirred at 0 °C for 20 min. Then 2.15 g (3.64 mmol, 1.0 eq.) **(3)** was added and the mixture was stirred at 30 °C over night. After the addition of 3.05 mL (21.85 mmol, 6.0 eq.) NEt_3 und 1.08 mL (43.70 mmol, 12.0 eq.) H_2O the mixture was stirred at rt for 10 min and extracted with 150 mL DCM. The combined organic layers were washed with saturated $NaHCO_3$ and brine, dried over $MgSO_4$ and subsequently evaporated to dryness. The crude intermediate was dissolved in 60 mL Dioxan and 17 mL concentrated NH_4OH were added. After stirring at rt for 10 min , the reaction was stopped by the addition of sat. NH_4Cl-solution, extracted with 200 mL DCM, dried over $MgSO_4$ and subsequently evaporated to dryness. The crude product was purified by column chromatography (DCM/MeOH 60:1) to give 1.78 g (82 %) of **4** as a colorless solid. **^{1}H-NMR** (600 MHz, $CDCl_3$,

ppm): δ = 7.98 (s, 1H, N-C$\underline{H}$=C-I), 7.94 (d, ^3J=8.2, 2H, C-C$\underline{H}$=CH-C-CH$_3$), 7.91 (d, ^3J=8.2, 2H, C-C$\underline{H}$=CH-C-CH$_3$), 7.26 (d, ^3J=8.2, 4H, C-CH=C$\underline{H}$-C-CH$_3$), 6.37 (dd, ^3J=5.8, 8.0, 1H, N-C$\underline{H}$-O), 5.59 (d, ^3J=6.4, 1H, O-C$\underline{H}$-CH$_2$-CH-N), 4.77 (dd, ^3J=2.9, ^2J=12.2, 1H, O-C$\underline{H}_2$-CH-O), 4.68 (dd, ^3J=3.5, ^2J=12.3, 1H, O-C$\underline{H}_2$-CH-O), 4.59 (d, , ^3J=1.9, 1H, O-CH$_2$-C$\underline{H}$-O), 2.98 (dd, ^3J=5.4, ^2J=14.4, 1H, O-CH-C$\underline{H}_2$-CH-N), 2.42 (d, ^3J=7.8, 6H, C-CH=CH-C-C$\underline{H}_3$), 2.20-2.15 (m, 1H, O-CH-C$\underline{H}_2$-CH-N). **^{13}C-NMR** (151 MHz, CDCl$_3$, ppm): δ = 166.3 (2C, O-($\underline{C}$=O)-C), 163.6 (C-($\underline{C}$-NH$_2$)=N), 146.3 (N-($\underline{C}$=O)-N), 144.7 (N-$\underline{C}$=C-I), 144.6 (2C, H$_3$C-$\underline{C}$=C), 129.9 (4C, $\underline{C}$=C-(C=O)-O), 129.6 (4C, H$_3$C-C=$\underline{C}$-C), 126.6 (2C, O-(C=O)-$\underline{C}$-C), 87.3 (O-$\underline{C}$H-N), 83.7 (O-$\underline{C}$H-CH$_2$-O), 75.4 (O-$\underline{C}$H-CH$_2$-CH-N), 64.4 (O-$\underline{C}$H$_2$-CH-O), 56.0 (N-CH-($\underline{C}$-I)-C), 39.7 (N-CH-$\underline{C}$H$_2$-CH-O), 21.9 (2C, $\underline{C}$H$_3$-C=C). **^{15}N-NMR** (41 MHz, CDCl$_3$, ppm): δ = -169.4, -215.6. **HRMS** (ESI+): calc. for C$_{25}$H$_{25}$I^{15}N$_2$O$_6$ [M+H]$^+$: 592.0723, found: 592.0726. **Melting range:** 144-148 °C. **IR (ATR):** v (cm^{-1}) = 3644 (w), 2473 (w), 2155 (m), 2004 (m), 1722 (m), 1609 (m), 1268 (s), 1097 (m), 950 (w), 763 (s).

3´,5´-(*tert*-Butyl-dimethylsilyl)-5-iodo-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (5)

To a stirred solution of **4** (1.78 g, 3.01 mmol, 1.0 eq.) in 200 mL dry MeOH was added K$_2$CO$_3$ (2.08 g, 15.04 mmol, 5.0 eq.). After stirring at rt for 19 h, the reaction mixture was evaporated to dryness. The crude product was solved in 30 mL DMF and cooled to 0 °C without any further purification. Imidazole (2.46 g, 36.10 mmol, 12.0 eq.) and TBS-Cl (5.44, 36.10 mmol, 12.0 eq.) were added. The reaction mixture was stirred at rt for 4 days and stopped by addition of ice and 100 mL sat. NaHCO$_3$-solution. The mixture was diluted and extracted with 200 mL EtOAc and the combined organic layers were washed with sat. NH$_4$Cl-solution, dried over MgSO$_4$ and subsequently evaporated to dryness. The crude product was purified by column chromatography (DCM/MeOH 60:1) to give 1.1 g (63 %) of **5** as a colorless solid **^{1}H-NMR** (200 MHz, CDCl$_3$, ppm): δ = 8.16 (s, 1H, N-C$\underline{H}$=(C-I)), 6.23 (t, ^3J=5.8, 1H, N-C$\underline{H}$-O), 4.37 (q, 1H, O-C$\underline{H}$-CH$_2$-CH-N), 4.01 (q, ^3J=2.5, 1H, O-CH-C$\underline{H}$-CH$_2$-O), 3.90 (dd, ^3J=2.5, ^2J=11.5, 1H, O-C$\underline{H}_2$-CH-O), 3.76 (dd, ^3J=2.4, ^2J=11.5, 1H, O-C$\underline{H}_2$-CH-O), 2.48- 2.44 (m, 1H, O-CH-C$\underline{H}_2$-CH-N), 2.01- 1.96 (m, 1H, O-CH-C$\underline{H}_2$-CH-N), 0.93 (s, 9H, Si-C-C$\underline{H}_3$), 0.89 (s, 9H, Si-C-C$\underline{H}_3$), 0.13 (s, 6H, Si-C$\underline{H}_3$), 0.08 (s, 6H, Si-C$\underline{H}_3$). **^{13}C-NMR** (151 MHz, CDCl$_3$, ppm): δ =161.2 (C-($\underline{C}$-NH$_2$)=N), 147.5 (N-$\underline{C}$H=(C-I)), 143.9 (N-($\underline{C}$=O)-N), 87.7 (O-$\underline{C}$H-CH$_2$-O), 88.1 (O-$\underline{C}$H-N), 72.2 (O-$\underline{C}$H-CH$_2$-CH-N), 62.8 (O-$\underline{C}$H$_2$-CH-O), 55.9 (N-CH-($\underline{C}$-I)-C), 42.6 (N-CH-$\underline{C}$H$_2$-CH-O), 25.9 (6C, ($\underline{C}$H$_3$)$_3$-C-Si),18.5 ((CH$_3$)$_3$-$\underline{C}$-Si), 18.0 ((CH$_3$)$_3$-$\underline{C}$-Si), -4.6 (O-Si-$\underline{C}$H$_3$), -4.9 (O-Si-$\underline{C}$H$_3$), -5.1 (O-Si-$\underline{C}$H$_3$), -5.2 (O-Si-$\underline{C}$H$_3$). **^{15}N-NMR** (41 MHz, CDCl$_3$, ppm): δ = -174.2, -212.6. **HRMS** (ESI+): calc. for C$_{21}$H$_{41}$I^{15}N$_2$O$_4$Si$_2$ [M+H]$^+$: 584.1615, found: 584.1618. **Melting range:** 195-196 °C. **IR (ATR):** v (cm^{-1}) = 3643 (w), 2850 (w), 2517 (w), 2172 (s), 2004 (s), 1650 (s), 1470 (m), 1271 (m), 1094 (m), 836 (m), 775 (m).

3´,5´-(*tert*-Butyl-dimethylsilyl)-5-carboxymethyl-2´-deoxy-(*N*¹,*N*³-¹⁵N)-cytidine (**6**)

5´-Iodo-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxy-(*N*¹,*N*³-¹⁵N)-cytidine (**5**) (0.20 g, 0.34 mmol, 1.0 eq.), bis(acetonitrile)palladium(II) chloride (20.0 mg, 0.08 mmol, 0.2 eq.) and *N*,*N*-diisopropylethylamine (0.24 mL, 1.38 mmol, 4.1 eq.) were dissolved in 8 mL three times degassed methanol in a high pressure glas autoclave. The autoclave was flushed with CO to remove residual air and after heating to 60 °C, the CO pressure was set to 3.5 bar. The reaction mixture was stirred at 60 °C for 17 h. After evaporation to dryness, the product was purified by column chromatography (EtOAc) to give product **6** as a colorless powder (0.12 g, 0.23 mmol, 67%). **¹H-NMR** (600 MHz, CDCl₃, ppm): δ = 8.64 (s, 1H, N-C$\underline{H}$=C), 7.94 (s br, 1H, N$\underline{H}_2$), 6.17-6.19 (m, 2H, N$\underline{H}_2$, N-C$\underline{H}$-O), 4.35 (dt, *J*=6.0, 2.5 Hz, 1H, C$\underline{H}$-O-Si), 4.06 (q, *J*=2.9 Hz, 1H, C$\underline{H}$-CH₂-O), 3.85 (dd, *J*=11.3, 3.1 Hz, 1H, C$\underline{H}_2$-O-Si), 3.82 (s, 3H, OC$\underline{H}_3$), 3.77 (dd, *J*=11.3, 3.1 Hz, 1H, C$\underline{H}_2$-O-Si), 2.62 (dddd, *J*=13.5, 5.9, 2.4, 1.1 Hz, 1H, N-CH-C$\underline{H}_2$), 1.94-1.99 (m, 1H, O-CH-C$\underline{H}_2$), 0.89 (s, 9H, Si-C-C$\underline{H}_3$), 0.87 (s, 9H, Si-C-C$\underline{H}_3$), 0.08 (s, 3H, Si-C$\underline{H}_3$), 0.07 (s, 3H, Si-C$\underline{H}_3$), 0.06 (s, 3H, Si-C$\underline{H}_3$), 0.06 (s, 3H, Si-C$\underline{H}_3$). **¹³C-NMR** (151 MHz, CDCl₃, ppm): δ =165.3 ($\underline{C}$O-O), 163.6 (d, *J*=6.1 Hz, $\underline{C}$-NH₂), 154.1 (dd, J=10.0, 7.7 Hz, $\underline{C}$=O-N), 147.6 (d, *J*=13.9 Hz, $\underline{C}$H=C), 95.0 (d, *J*=1.7 Hz, $\underline{C}$=CO), 88.8 ($\underline{C}$H-CH₂-O), 87.9 (d, *J*=10.2 Hz, N-$\underline{C}$H-CH₂), 72.7 ($\underline{C}$H-O-Si), 63.1 ($\underline{C}$H₂-O-Si), 51.9 (O$\underline{C}$H₃), 42.9 (O-$\underline{C}$H-CH₂), 25.8 (Si-C-$\underline{C}$H₃), 25.8 (Si-C-$\underline{C}$H₃), 18.3 (Si-$\underline{C}$), 18.0 (Si-$\underline{C}$), -4.6 (Si-$\underline{C}$H₃), -4.9 (Si-$\underline{C}$H₃), -5.6 (Si-$\underline{C}$H₃), -5.7 (Si-$\underline{C}$H₃). **¹⁵N-NMR** (40 MHz, CDCl₃, ppm): δ =-172.7, -211.4. **HRMS** (EI+): calc. for C₂₂H₄₀N¹⁵N₂O₆Si₂ [M-CH₃]⁺: 500.2391, found: 500.2391. **Melting range:** 203-204 °C. **IR** (cm⁻¹): v= 3409 (w), 2952 (w), 2928 (w), 1709 (m), 1667 (m), 1496 (m), 1321 (m), 1257 (s), 1087 (s), 1064 (s), 1013 (s), 831 (s), 789 (s), 775 (s).

3´,5´-(*tert*-Butyl-dimethylsilyl)-4-*N*-dibenzoyl-5-carboxymethyl-2´-deoxy-(*N*¹,*N*³-¹⁵N)-cytidine (7)

To a solution of 3´,5´-(*tert*-butyl-dimethylsilyl)-5-carboxymethyl-2´-deoxy-(*N*¹,*N*³-¹⁵N)-cytidine (**6**) (1.07 g, 2.07 mmol, 1.0 eq.) in pyridine (29.4 mL) benzoyl chloride (3.3 mL, 28.43 mmol, 13.7 eq.) was added at 0 °C. The mixture was stirred at 0 °C for 1.5 h and additional 19 h at room temperature. After quenching with ice (33.5 g) and addition of a saturated NaHCO₃-solution, the product was extracted with EtOAc (441 mL), the organic phase washed with a saturated NH₄Cl-solution (441 mL), dried over MgSO₄ and the solvent evaporated *in vacuo*. After column chromatography (iHex/EtOAc 4:1) product **7** was obtained as a colorless solid (1.21 g, 1.67 mmol, 81%). **¹H-NMR** (600 MHz, CDCl₃, ppm): δ = 8.88 (s, 1H, N-C$\underline{H}$=C), 7.85-7.87 (m, 4H, CO-C$_q$-CH-C$\underline{H}$), 7.48-7.51 (m, 2H, CO-C$_q$-CH-CH-C$\underline{H}$), 7.37-7.40 (m, 4H, C$_q$-C$\underline{H}$), 6.06-6.08 (m, 1H, N-C$\underline{H}$-O), 4.35-4.37 (m, 1H, C$\underline{H}$-O-Si), 4.12 (dd, *J*=5.3, 2.8 Hz, 1H,

CH-CH$_2$-O), 3.88 (dd, J=11.4, 3.1 Hz, 1H, CH$_2$-O-Si), 3.78 (dd, J=11.4, 2.9 Hz, 1H, CH$_2$-O-Si), 3.72 (s, 3H, OCH$_3$), 2.64-2.67 (m, 1H, O-CH-CH$_2$), 1.97-2.02 (m, 1H, O-CH-CH$_2$), 0.88 (s, 18H, Si-C-CH$_3$), 0.09 (s, 3H, Si-CH$_3$), 0.08 (s, 3H, Si-CH$_3$), 0.08 (s, 3H, Si-CH$_3$), 0.07 (s, 3H, Si-CH$_3$). ^{13}C-NMR (151 MHz, CDCl$_3$, ppm): δ =172.4 (N-CO), 164.0 (d, J=6.9 Hz, C-NH$_2$), 162.6 (CO-O), 153.1 (dd, J=10.6, 4.6 Hz, N-CO-N), 149.2 (d, J=13.8 Hz, CH=C), 134.1 (C$_q$-CH), 132.9 (C$_q$-CH-CH-CH), 129.3 (C$_q$-CH-CH), 128.8 (C$_q$-CH-CH), 103.9 (d, J=1.5 Hz, C=CO), 89.5 (CH-CH$_2$-O), 89.0 (d, J=9.0 Hz, N-CH-CH$_2$), 72.9 (CH-O-Si), 63.1 (CH$_2$-O-Si), 52.4 (OCH$_3$), 42.6 (O-CH-CH$_2$), 25.9 (Si-C-CH$_3$), 25.7 (Si-C-CH$_3$), 18.3 (Si-C), 18.0 (Si-C), -4.6 (Si-CH$_3$), -4.9 (Si-CH$_3$), -5.5 (Si-CH$_3$), -5.6 (Si-CH$_3$). ^{15}N-NMR (40 MHz, CDCl$_3$, ppm): δ =-112.3, -202.0. HRMS (ESI+): calc. for C$_{37}$H$_{51}$N^{15}N$_2$NaO$_8$Si$_2$ [M+Na]$^+$: 746.3048, found: 746.3046. Melting range: 84-85 °C. IR (cm^{-1}): v= 2953 (w), 2929 (w), 2857 (w), 1683 (s), 1615 (w), 1495 (m), 1277 (m), 1242 (s), 1107 (s), 1079 (s), 831 (s), 776 (s), 700 (s), 676 (s).

4-*N*-Dibenzoyl-5-carboxymethyl-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (8)

A solution of 3´,5´-(*tert*-butyl-dimethylsilyl)-4-*N*-dibenzoyl-5-carboxymethyl-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (**7**) (1.12 g, 1.54 mmol, 1.0 eq.), HF*pyridine (0.61 mL, 6.81 mmol, 4.4 eq.), pyridine (0.59 mL, 7.15 mmol, 4.6 eq.) in EtOAc (25.0 mL) was stirred at room temperature for 14 h. The reaction was quenched with methoxytrimethylsilane (7.3 mL, 52.85 mmol, 34.3 eq.) and stirred for additional 40 min. After evaporation to dryness and purification by column chromatography product **8** was obtained as colorless solid (0.59 g, 1.19 mmol, 77%). ^{1}H-NMR (600 MHz, CDCl$_3$, ppm): δ = 9.33 (s, 1H, N-CH=C), 7.82-7.84 (m, 4H, CO-C$_q$-CH-CH), 7.49-7.51 (m, 2H, CO-C$_q$-CH-CH-CH), 7.37-7.39 (m, 4H, C$_q$-CH), 6.05-6.07 (m, 1H, N-CH-O), 4.42 (dd, J=10.8, 5.6 Hz, 1H, CH-OH), 4.02-4.04 (m, 1H, CH-CH$_2$-O), 3.91 (dd, J=12.1, 2.9 Hz, 1H, CH$_2$-OH), 3.80 (dd, J=12.1, 2.6 Hz, 1H, CH$_2$-OH), 3.66 (s, 3H, OCH$_3$), 3.39 (s br, 2H, OH), 2.49-2.53 (m, 1H, O-CH-CH$_2$), 1.19-2.23 (m, 1H, O-CH-CH$_2$). ^{13}C-NMR (151 MHz, CDCl$_3$, ppm): δ =172.7 (N-CO), 164.0 (d, J=7.0 Hz, C-NH$_2$), 162.8 (CO-O), 153.6 (dd, J=10.9, 4.8 Hz, N-CO-N), 151.0 (d, J=13.6 Hz, CH=C), 133.7 (C$_q$-CH), 133.2 (C$_q$-CH-CH-CH), 129.4 (C$_q$-CH-CH), 128.9 (C$_q$-CH-CH), 104.4 (d, J=1.4 Hz, C=CO), 88.5 (d, J=8.2 Hz, N-CH-CH$_2$), 87.8 (CH-CH$_2$-O), 69.7 (CH-OH), 61.1 (CH$_2$-OH), 52.5 (OCH$_3$), 41.4 (O-CH-CH$_2$). ^{15}N-NMR (40 MHz, CDCl$_3$, ppm): δ =-114.2, -199.8. HRMS (ESI+): calc. for C$_{25}$H$_{24}$N^{15}N$_2$O$_8$ [M+H]$^+$: 496.1499, found: 496.1500. Melting range: 128-130 °C. IR (cm^{-1}): v= 3398 (w), 1666 (s), 1612 (m), 1498 (s), 1241 (s), 1195 (m), 1176 (m), 1092 (s), 750 (s), 703 (s), 666 (s).

4-*N*-Dibenzoyl-5-carboxymethyl-5´-(dimethoxytrityl)-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (9)

To a solution of 4-*N*-dibenzoyl-5-carboxymethyl-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (**8**) (0.34 g, 0.69 mmol, 1.0 eq.) in pyridine (9.3 mL) was added 3 Å molecular sieve and DMT-Cl (0.21 g, 0.62 mmol, 0.9 eq.). The reaction mixture was stirred at room temperature for 16.5 h. After quenching with methanol (11.7 mL) the solvent was evaporated *in vacuo*. Purification by column chromatography gave product **9** as a yellowish foam (0.31 g, 0.39 mmol, 63%). ^{1}**H-NMR** (200 MHz, CDCl$_3$, ppm): δ = 9.15 (s, 1H, N-C**H**=C), 7.84-7.88 (m, 4H, CO-C$_q$-CH-C**H**), 7.23-7.54 (m, 15H, CO-C$_q$-CH-CH-C**H**, CO-C$_q$-C**H**, C$_q$-C$_q$-C**H**-CH-C$_q$-O, C$_q$-C$_q$-C**H**-CH-C**H**), 6.85 (d, *J*=8.8 Hz, 4H, C$_q$-C$_q$-CH-C**H**-C$_q$-O), 6.10 (t, *J*=6.0 Hz, 1H, N-C**H**-O), 4.38 (dd, *J*=9.7, 4.9 Hz, 1H, C**H**-OH), 4.08-4.14 (m, 1H, C**H**-CH$_2$-O), 3.78 (s, 6H, C$_{Ar}$-OC**H$_3$**), 3.60 (dd, *J*=10.7, 3.0 Hz, 1H, C**H$_2$**-O), 3.28 (dd, *J*=10.8, 3.1 Hz, 1H, C**H$_2$**-O), 3.07 (s, 3H, CO-OC**H$_3$**), 2.55-2.68 (m, 1H, O-CH-C**H$_2$**), 2.41 (s br, 1H, O**H**), 2.16-2.29 (m, 1H, O-CH-C**H$_2$**). ^{13}**C-NMR** (101 MHz, CDCl$_3$, ppm): δ =172.4 (N-**C**O), 164.3 (dd, *J*=6.9, 1.3 Hz, **C**-NH$_2$), 162.2 (d, *J*=1.5 Hz, **C**O-O), 158.6 (**C$_q$**-OCH$_3$), 158.6 (**C$_q$**-OCH$_3$), 153.2 (dd, *J*=10.6, 4.6 Hz, N-**C**O-N), 149.1 (d, *J*=13.7 Hz, **C**H=C), 144.3 (**C$_q$**-C$_q$-CH-CH-CH), 135.6 (**C$_q$**-C$_q$-CH-CH-C$_q$), 135.4 (**C$_q$**-C$_q$-CH-CH-C$_q$), 134.1 (CO-**C$_q$**-CH), 132.9 (CO-C$_q$-CH-CH-**C**H), 130.0 (C$_q$-C$_q$-**C**H-CH-C$_q$), 129.9 (C$_q$-C$_q$-**C**H-CH-C$_q$), 129.4 (CO-C$_q$-CH-**C**H), 128.7 (**C$_{Ar}$**H), 128.0 (**C$_{Ar}$**H), 127.9 (C$_q$-C$_q$-CH-**C**H-CH), 127.0 (C$_q$-C$_q$-CH-CH-**C**H), 113.1 (**C**H-C$_q$-OCH$_3$), 113.1 (**C**H-C$_q$-OCH$_3$), 104.1 (dd, *J*=2.0, 0.7 Hz, **C**=CO), 88.0 (d, *J*=8.8 Hz, N-**C**H-CH$_2$), 86.9 (**C**H-CH$_2$-O), 86.6 (O-**C$_q$**-C$_q$), 70.4 (**C**H-OH), 62.3 (**C**H$_2$-OH), 55.2 (2C, CH-C$_q$-O**C**H$_3$), 51.8 (CO-O**C**H$_3$), 41.7 (O-CH-**C**H$_2$). ^{15}**N-NMR** (40 MHz, CDCl$_3$, ppm): δ =-112.0, -202.3. **HRMS** (ESI+): calc. for C$_{46}$H$_{40}$N^{15}N$_2$O$_{10}$ [M-H]$^+$: 796.2660, found: 796.2660. **Melting range:** 132-134 °C. **IR** (cm^{-1}): ν= 1681 (m), 1608 (m), 1496 (m), 1304 (m), 1276 (m), 1244 (m), 1175 (m), 1116 (m), 1094 (m), 1030 (m), 827 (m), 790 (m), 777 (m), 727 (m), 699 (s).

4-*N*-Dibenzoyl-5-carboxymethyl-3´-(di*iso*propylamino-cyanoethoxyphosphino)-5´-(dimethoxytrityl)-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (10)

To a solution of 4-*N*-dibenzoyl-5-carboxymethyl-5´-(dimethoxytrityl)-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (**9**) (50 mg, 62.67 µmol, 1.0 eq.) in three times degassed DCM (1.3 mL) were added di*iso*propylammonium tetrazolide (6 mg, 35.04 µmol, 0.6 eq.) and bis(di*iso*propylamino)-2-cyanoethoxyphosphine (25 µL, 78.71 µmol, 1.26 eq.). The resulting mixture was stirred at room temperature for 17 h. After evaporation to dryness the product was obtained after column chromatography (DCM/MeOH 200:1+1% pyridine) as a white foam of the two diastereomers (73 mg, 73.14 µmol, quant.). ^{1}**H-NMR** (200 MHz, *d*^{6}-acetone, ppm): δ = 9.21 (s, 1H, N-C**H**=C), 9.21 (s, 1H, N-C**H**=C), 7.88-7.94 (m, 8H, CO-C$_q$-CH-C**H**), 7.26-7.64 (m, 30H, CO-C$_q$-CH-CH-C**H**, CO-C$_q$-C**H**, C$_q$-C$_q$-C**H**-CH-C$_q$-O, C$_q$-C$_q$-C**H**-C**H**-C**H**), 6.90-6.96 (m,

8H, C$_q$-C$_q$-CH-C**H**-C$_q$-O), 6.09 (t, *J*=5.1 Hz, 2H, N-C**H**-O), 4.56-4.71 (m, *J*=9.7, 2H, C**H**-OH), 4.26-4.34 (m, 2H, C**H**-CH$_2$-O), 3.80 (s, 6H, C$_{Ar}$-OC**H$_3$**), 3.80 (s, 6H, C$_{Ar}$-OC**H$_3$**), 3.52-3.72 (m, 8H, P-O-C**H$_2$**, N-C**H** or C**H$_2$**-CN), 3.33-3.44 (m, 2H, C**H$_2$**-O), 3.07 (s, 3H, CO-OC**H$_3$**), 3.07 (s, 3H, CO-OC**H$_3$**), 2.42-2.77 (m, 8H, C**H$_2$**-CH, N-C**H** or C**H$_2$**-CN),1.03-1.19 (m, 24H,CH-C**H$_3$**). **^{31}P-NMR** (81 MHz, *d^6*-acetone, ppm): δ = 149.7, 149.4. **HRMS** (ESI+): calc. for C$_{55}$H$_{58}$N$_3$^{15}N$_2$NaO$_{11}$P [M+H]$^+$: 1020.3709, found: 1020.3710.

Synthesis of oligonucleotides

Oligonucleotide synthesis was performed on an ABI 394 DNA/RNA synthesizer (*Applied Biosystems*) using standard DNA synthesis conditions (DMT on). Phosphoramidites and CPG carriers were obtained from *Glen Research*. The 5-octadiin-5´-DMT-3´-CED-dU-phosphoramidite was synthesized according to literature procedure.[8] The crude oligonucleotide was cleaved from the resin and deprotected by incubation in a 0.4 M NaOH-solution (methanol/H$_2$O 4:1) at 27 °C for 2 d. After HPLC-purification the DMT group was cleaved by incubation with a 2% DCA-solution for 30 min at 27 °C, precipitated by addition of a 3 M NaOAc-solution followed by another HPLC purification. The biotin label was introduced by standard click conditions with a biotin-azid from *baseclick*.[9] The DNA-strands were characterized by MALDI-TOF (*Bruker Autoflex II*).

3´,5´-(*tert*-Butyl-dimethylsilyl)-5-carboxymethyl-2´-deoxy-3,6-dihydro-cytidine (**12**)

To 3´,5´-(*tert*-Butyl-dimethylsilyl)-5-carboxymethyl-2´-deoxy-cytidine **11** (50.0 mg, 97.32 µmol, 1.0 eq.) dissolved in methanol (5.0 mL) was added sodium borohydride (3.6 mg, 95.16 µmol, 0.98 eq.). After stirring for 20 min at room temperature a saturated NH$_4$Cl solution (33 mL) was added and extracted with EtOAc (33 mL). The organic phase was washed with a saturated NH$_4$Cl solution (2 × 30 mL), dried over MgSO$_4$ and evaporated to dryness to give **12** as a colorless solid (49 mg, 95.00 µmol, 98%). **^{1}H-NMR** (600 MHz, CDCl$_3$): δ (ppm) = 9.39 (s, br, 1H, N**H**), 6.31 (s, br, 2H, N**H$_2$**), 6.27 (dd, *J*=8.0, 6.2 Hz, 1H, O-C**H**-N), 4.39 (dt, *J*=6.3, 3.2 Hz, 1H, O-C**H**-CH$_2$-CH-N), 4.02 (d, *J*=12.5 Hz, 1H, NC**H$_2$**), 3.91 (d, *J*=12.6 Hz, 1H, NC**H$_2$**), 3.75-3.77 (m, 1H, O-CH-C**H**-CH$_2$-O), 3.68-3.72 (m, 2H, O-C**H$_2$**-CH-O), 3.66 (s, 3H, OC**H$_3$**), 2.13 (ddd, *J*=13.2, 7.9, 6.3 Hz, 1H, O-CH-C**H$_2$**-CH-N), 1.87 (ddd, *J*=13.0, 6.1, 3.3 Hz, 1H, O-CH-C**H$_2$**-CH-N), 0.90 (s, 9H, Si-C-C**H$_3$**), 0.89 (s, 9H, Si-C-C**H$_3$**), 0.08 (s, 3H, Si-C**H$_3$**), 0.07 (s, 3H, Si-C**H$_3$**), 0.07 (s, 3H, Si-C**H$_3$**), 0.06 (s, 3H, Si-C**H$_3$**). **^{13}C-NMR** (151 MHz, CDCl$_3$, ppm): δ =167.8 (**C**O-O), 154.6 (**C**=O-N), 151.4 (**C**-NH$_2$), 85.9 (**C**H-CH$_2$-O), 83.3 (N-**C**H-CH$_2$), 71.9 (**C**H-O-Si), 70.5 (**C**-CO-OCH$_3$), 62.8 (**C**H$_2$-O-Si), 50.4 (O**C**H$_3$), 37.6 (N-**C**H$_2$), 36.1 (**C**H$_2$-CH-O), 25.9 (Si-C-**C**H$_3$), 25.8 (Si-C-**C**H$_3$), 18.4 (Si-**C**H$_3$),

18.0 (Si-$\underline{C}H_3$), -4.7 (Si-$\underline{C}H_3$), -4.8 (Si-$\underline{C}H_3$), -5.5 (Si-$\underline{C}H_3$), -5.6 (Si-$\underline{C}H_3$). **HRMS** (ESI+): calc. for $C_{23}H_{46}N_3O_6Si_2^+$ [M+H$^+$]$^+$: 516.2920, found: 516.2915. **Melting range**: 62-64 °C. **IR** (cm^{-1}): v= 3422 (w), 3311 (w), 2952 (w), 2929 (w), 2857 (w), 1682 (m), 1629 (m), 1554 (m), 1462 (m), 1435 (m), 1335 (m), 1252 (m), 1098 (s), 1025 (m), 832 (s), 774 (s).

3´,5´-(*tert*-Butyl-dimethylsilyl)-2´-deoxy-5,6-dihydrocytidine (**13**)

To a solution of 3´,5´-(*tert*-Butyl-dimethylsilyl)-5-carboxymethyl-2´-deoxy-3,6-dihydrocytidine **12** (2.16 g, 4.19 mmol, 1.0 eq.) in a acetonitrile/water mixture (173 mL/52 mL) LiOH (1.22 g, 50.94 mmol, 12.2 eq.) was added and the resulting solution was stirred at room temperature for 2 h. After addition of a saturated NaHCO$_3$-solution (387 mL) and acidification with 2 M hydrochloric acid (221 mL) the aqueous phase was extracted with diethyl ether (3 × 450 mL), dried over MgSO$_4$ and evaporated to dryness. After column chromatography (DCM/MeOH 50:1→5:1) the product was obtained as a colorless solid (184 mg, 0.40 mmol, 10%). **^{1}H-NMR** (600 MHz, CDCl$_3$): δ (ppm) = 7.98 (s, br, 2H, N$\underline{H}_2$), 6.33-6.35 (m, 1H, O-C$\underline{H}$-N), 4.31-4.34 (m, 1H, O-C$\underline{H}$-CH$_2$-CH-N), 3.74 (q, J=3.3 Hz, 1H, O-CH-C$\underline{H}$-CH$_2$-O), 3.68 (dd, J=3.4, 1.1 Hz, 2H, C$\underline{H}_2$-O), 3.54 (dt, J=12.1, 6.0 Hz, 1H, N-C$\underline{H}_2$), 3.18 (ddd, J=12.1, 10.0, 4.8 Hz, 1H, N-C$\underline{H}_2$), 2.51-2.64 (m, 2H, N-CH$_2$-C$\underline{H}_2$), 1.92-1.95 (m, 2H, O-CH-C$\underline{H}_2$-CH-N), 0.89 (s, 9H, Si-C-C$\underline{H}_3$), 0.87 (s, 9H, Si-C-C$\underline{H}_3$), 0.06 (s, 9H, Si-C$\underline{H}_3$), 0.05 (s, 3H, Si-C$\underline{H}_3$), 0.05 (s, 3H, Si-C$\underline{H}_3$), 0.04 (s, 3H, Si-C$\underline{H}_3$). **^{13}C-NMR** (151 MHz, CDCl$_3$, ppm): δ =164.6 ($\underline{C}$-NH$_2$), 156.2 ($\underline{C}$=O-N), 85.9 ($\underline{C}$H-CH$_2$-O), 83.9 (N-$\underline{C}$H-CH$_2$), 71.9 ($\underline{C}$H-O-Si), 62.9 ($\underline{C}$H$_2$-O-Si), 37.4 ($\underline{C}$H$_2$-CH-O), 35.6 (N-$\underline{C}$H$_2$), 28.5 (N-CH$_2$-$\underline{C}$H$_2$), 25.9 (Si-C-$\underline{C}$H$_3$), 25.8 (Si-C-$\underline{C}$H$_3$), 18.3 (Si-$\underline{C}$H$_3$), 18.0 (Si-$\underline{C}$H$_3$), -4.7 (Si-$\underline{C}$H$_3$), -4.8 (Si-$\underline{C}$H$_3$), -5.4 (Si-$\underline{C}$H$_3$), -5.6 (Si-$\underline{C}$H$_3$). **HRMS** (ESI+): calc. for $C_{21}H_{44}N_3O_4Si_2^+$ [M+H$^+$]$^+$: 458.2865, found: 458.2862. **Melting range**: 179-180 °C. **IR** (cm^{-1}): v= 2954 (m), 2929 (m), 2857 (m), 1631 (m), 1552 (m), 1472 (m), 1416 (m), 1298 (m), 1252 (s), 1090 (s), 1033 (m), 1005 (m), 831 (s), 774 (s).

3´,5´-(*tert*-Butyl-dimethylsilyl)-2´-deoxycytidine

To a solution of 3´,5´-(*tert*-Butyl-dimethylsilyl)-2´-deoxy-5,6-dihydrocytidine (**13**) (50.0 mg, 0.11 mmol, 1.0 eq.) in 1,4-dioxane (3 mL) 2,3-dichloro-5,6-dicyano-*p*-benzoquinone (76.0 mg, 0.33 mmol, 3.0 eq.) was added and the resulting solution was stirred at room temperature for 25 h. After evaporation to dryness and purification by column chromatography (DCM/MeOH 15:1) the product was obtained as a colorless solid (26.0 mg, 0.06 mmol, 55%).

^{1}H-NMR (300 MHz, CDCl$_3$): δ (ppm) = 7.96 (d, J=7.4 Hz, 1H, N-C$\underline{H}$-CH), 6.25 (dd, J=6.3, 5.2 Hz, 1H, O-C$\underline{H}$-N), 5.66 (d, J=7.4 Hz, 1H, N-CH-C$\underline{H}$), 4.33-4.39 (m, 1H, Si-O-C$\underline{H}$), 3.86-

3.93 (m, 1H, Si-O-C$\underline{H}_2$), 3.73-3.78 (m, 1H, Si-O-C$\underline{H}_2$), 2.40 (ddd, J=13.3, 6.4, 5.6 Hz, 1H, N-CH-C$\underline{H}_2$), 2.06 (ddd, J=13.4, 6.4, 5.2 Hz, 1H, N-CH-C$\underline{H}_2$), 0.91 (s, 9H, Si-C-C$\underline{H}_3$), 0.87 (s, 9H, Si-C-C$\underline{H}_3$), 0.10 (s, 9H, Si-C$\underline{H}_3$), 0.09 (s, 3H, Si-C$\underline{H}_3$), 0.05 (s, 3H, Si-C$\underline{H}_3$), 0.05 (s, 3H, Si-C$\underline{H}_3$). ^{13}**C-NMR** (75 MHz, CDCl$_3$, ppm): δ =165.6 ($\underline{C}$-NH$_2$), 155.8 ($\underline{C}$=O-N), 141.3 (N-$\underline{C}$H-CH), 93.7 (N-CH-$\underline{C}$H), 87.3 ($\underline{C}$H-CH$_2$-O), 85.9 (N-$\underline{C}$H-CH$_2$), 70.3 ($\underline{C}$H-O-Si), 62.0 ($\underline{C}$H$_2$-O-Si), 42.2 ($\underline{C}$H$_2$-CH-O), 25.9 (Si-C-$\underline{C}$H$_3$), 25.7 (Si-C-$\underline{C}$H$_3$), 18.3 (Si-$\underline{C}$H$_3$), 17.9 (Si-$\underline{C}$H$_3$), -4.6 (Si-$\underline{C}$H$_3$), -4.9 (Si-$\underline{C}$H$_3$), -5.5 (Si-$\underline{C}$H$_3$), -5.5 (Si-$\underline{C}$H$_3$). **HRMS** (ESI+): calc. for C$_{21}$H$_{42}$N$_3$O$_4$Si$_2^+$ [M+H$^+$]$^+$: 456.2708, found: 456.2713. **Melting range**: 99-102 °C. **IR** (cm^{-1}): ν= 2954 (m), 2929 (m), 2857 (m), 1625 (m), 1486 (m), 1472 (m), 1252 (s), 1112 (s), 833 (s), 775 (s).

Decarboxylation of caC containing oligo nucleotides with amino acids

A solution of ODN2 (10µM DNA, 0.6 M L-cysteine and 0.1 M L-histidine or 0.3 M L-arginine) was incubated at 50 °C for 20 h (*eppendorf Mastercycler personal*). The DNA was precipitated by addition of 50 µL 3 M NaOAc and 1 mL EtOH at -20 °C. After centrifugation the pellet was washed with 1 mL 75% EtOH, digested and analyzed by LC-MS.

References

[1]   J. D. Dignam, *Method. Enzymol.* **1990**, *182*, 194-203.
[2]   J. D. Dignam, R. M. Lebovitz, R. G. Roeder, *Nucleic Acids Res.* **1983**, *11*, 1475-1489.
[3]   A. Holmberg, A. Blomstergren, O. Nord, M. Lukacs, J. Lundeberg, M. Uhlén, *Electrophoresis* **2005**, *26*, 501-510.
[4]   M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375-5377.
[5]   C. J. LaFrancois, J. Fujimoto, L. C. Sowers, *Chem. Res. Toxicol.* **1998**, *11*, 75-83.
[6]   G. H. Clever, C. Kaul, T. Carell, *Angew. Chem. Int. Ed.* **2007**, *46*, 6226-6236.
[7]   J. Asakura, M. J. Robins, *J. Org. Chem.* **1990**, *55*, 4928-4933.
[8]   J. Gierlich, G. A. Burley, P. M. Gramlich, D. M. Hammond, T. Carell, *Org. Lett.* **2006**, *8*, 3639-3642.
[9]   P. M. Gramlich, S. Warncke, J. Gierlich, T. Carell, *Angew. Chem. Int. Ed.* **2008**, *47*, 3442-3444.

8.4 Zusatzmaterialien zu Abschnitt 3.4

Stefan Schiesser,[#] Toni Pfaffeneder,[#] Keyarash Sadeghian, Benjamin Hackner, Barbara Steigenberger, Arne S. Schröder, Jessica Steinbacher, Gengo Kashiwazaki, Georg Höfner, Klaus T. Wanner, Christian Ochsenfeld und Thomas Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593−14599. *Deamination, Oxidation and C–C bond cleavage reactivity of 5-hydroxy-methylcytosine, 5-formylcytosine and 5-carboxycytosine.*

Supporting Information

Deamination, Oxidation and C-C bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxycytosine

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

[†]Center for Integrated Protein Science (CiPS^M) at the Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstrasse 5–13, 81377 Munich.

[‡]Chair for Theoretical Chemistry at the Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstrasse 5–13, 81377 Munich.

[#]Center for Drug Research at the Department of Pharmacy, Ludwig-Maximilians Universität München, Butenandtstr. 5–13, 81377 Munich.

E-mail: Thomas.Carell@cup.uni-muenchen.de

Table of Contents

General Methods

Chemicals were purchased from *Sigma-Aldrich, Fluka, ABCR* or *Acros organics* and used without further purification. Solutions were concentrated *in vacuo* on a *Heidolph* rotary evaporator. The solvents for organic synthesis were of reagent grade and purified by distillation. Chromatographic purification of products was accomplished using flash column chromatography on *Merck* Geduran Si 60 (40 – 63 μm) silica gel (normal phase). Thin layer chromatography (TLC) was performed on *Merck* 60 (silica gel F_{254}) plates. Visualization of the developed chromatogram was performed using fluorescence quenching or staining solutions. ^{1}H-, ^{13}C- and ^{15}N-NMR spectra were recorded in deuterated solvents on *Varian Oxford 200, Bruker ARX 300*, *Varian VXR400S*, *Varian Inova 400* and *Bruker AMX 600* spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, qi = quintet, m = multiplet, br. = broad. High-resolution ESI spectra were obtained on the mass spectrometers *Thermo Finnigan* LTQ FT-ICR. IR measurements were performed on *Perkin Elmer Spectrum BX FT-IR* spectrometer with a diamond-ATR (Attenuated Total Reflection) setup. Melting ranges of small molecules were measured on a Büchi B-540. Melting points of DNA hairpins (see Table S1) were measured using a *JASCO V-650 Spectrophotometer* with a *JASCO* peltier element and 2 μM DNA-solutions in 100 mM NaCl and 10 mM sodium phosphate buffer of the respective pH-value. The pH-values were adjusted using an *MP 220* pH-meter (*Mettler Toledo*) with 5 N NaOH and 2 N HCl-solutions, except for sodium phosphate buffers. The later were adjusted using mixtures of monobasic and dibasic sodium phosphate solutions of the desired molarity.

Acetonitrile for HPLC-purification of nucleoside standards were purchased from *VWR*. Acetonitrile of LC-MS grade was purchased from *Carl Roth GmbH + Co. KG*. Formic acid was purchased from *Fluka*, p.a. for mass spectrometry. Water was purified by a Milli-Q Plus system from *Merck Millipore*. Nuclease S1 (*Aspergillus oryzae*) was obtained from *Sigma Aldrich*, snake venom phosphodiesterase I (*Crotalus adamanteus*) from *USB corporation* and antarctic phosphatase from *New England Biolabs*.

Oligonucleotide synthesis was performed on an ABI 394 DNA/RNA synthesizer (*Applied Biosystems*) using standard DNA synthesis conditions (DMT off). Phosphoramidites and CPG carriers were obtained from *Glen Research*. The crude oligonucleotide was cleaved from the resin and deprotected by incubation in a 0.4 M NaOH-solution (methanol/H_2O 4:1) at 27 °C for 2 d and purified by HPLC (see below). The DNA-strands were characterized by MALDI-TOF (*Bruker Autoflex II*).

Table S1. Melting points of synthesized hairpins at pH 5.0 and pH 7.4.

	T_M hairpin 1 [°C]	T_M hairpin 2 [°C]	T_M hairpin 3 [°C]	T_M hairpin 4 [°C]	T_M hairpin 5 [°C]	T_M hairpin 6 [°C]
pH 5.0	76.4	74.8	75.2	70.0	71.0	71.8
pH 7.4	78.6	76.6	77.0	73.6	70.9	72.8

High performance liquid chromatography:

The following devices were used for the purification and analysis of the nucleoside standards and synthetic DNA strands. Analytical HPLC: *Waters* Alliance (2695 Separation Module, 2996 Photodiode Array Detector); preparative HPLC: *Waters* Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump). For analytical HPLC a CC 250/4 Nucleosil 120-3 C18 column from *Macherey Nagel* was used (flow: 0.5 mL/min). For preparative HPLC a VP 250/10 Nucleosil 100-7 C18 column (flow: 5 mL/min) and a VP 250/32 Nucleosil 100-7 C18 column (flow: 15 mL/min) from *Macherey Nagel* were used.

In general the unprotected nucleosides were purified with an unbuffered gradient of water and acetonitrile unless otherwise stated. In these cases a gradient of buffer A (0.1 M NHEt$_3$OAc in water) and buffer B (0.1 M NHEt$_3$OAc in 80% MeCN) were used.

Synthesis and preparation of nucleoside standards for mass spectrometry

Experimental procedures for the synthesis of mdC, [D$_3$]-mdC, hmdC, fdC, cadC, hmdU, were published by *Münzel et al.*[1] and *Globisch et al.*[2] A summary of the performed reactions of the synthesized MS standards [^{15}N$_2$]-dC, [D$_2$,^{15}N$_2$]-hmdC, [^{15}N$_2$]-fdC, [^{15}N$_2$]-cadC, [D$_3$]-dT, [^{15}N$_2$]-dU, [D$_2$]-hmdU, fU and [^{15}N$_2$]-fdU is given in Scheme S1. The key steps in the synthesis of [D$_2$,^{15}N$_2$]-hmdC, [^{15}N$_2$]-fdC, [^{15}N$_2$]-cadC were Pd(0) catalyzed functionalizations of the protected ^{15}N$_2$-labeled 5-iodo-2'-deoxycytidine (**4**) according to the methodology of *Münzel et al.*[2,3] The key step in the synthesis of the latter compound **4** was the amination of the respective uridine derivative.[4] Amination was achieved by using methodologies of *Hodge et al.*[5] and *Felczak et al.*[6] and was earlier described by us.[4]

8-hydroxy-2′-deoxy-(^{15}N$_5$)-guanosine ([^{15}N$_5$]-8-oxo-dG) (99 atom% ^{15}N) was purchased from *Cambridge Isotope Laboratories*, 8-hydroxy-2′-deoxyguanosine (8-oxo-dG) from *BIOLOG*, 2′-deoxycytidine (dC) and 2′-deoxyuridine (dU) from *ChemGenes*. The stock solutions of MS standards were prepared by weighing the HPLC purified and lyophilized nucleosides (2 to 10 mg) on a balance in a volumetric flask. To take errors into account the weighing procedure was performed in triplicate for each unlabeled nucleoside. The weighed nucleosides were dissolved in 5 mL DMSO and filled up to 100 mL in a volumetric flask with deionized H$_2$O (purified by a Milli-Q Plus system from Merck Millipore) to receive the stock-concentrations, which were further diluted in 1/10-dilution steps as needed. The extinction coefficients of the unlabeled nucleosides were determined at 20 °C on a UV-Vis spectrometer (*Jasco V-650*) and applied to obtain the concentration of the isotope labeled nucleosides. All stock solutions were stored at -20 °C and warmed up to r.t. prior to its addition to the digestion mixture (spiking).
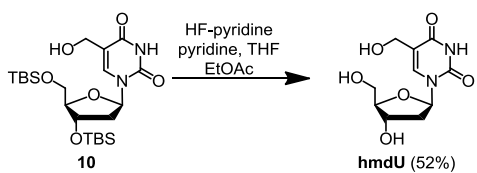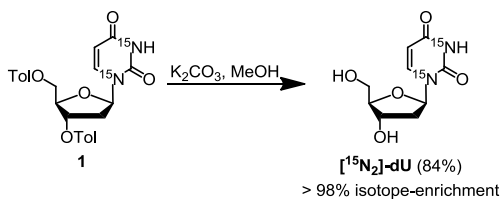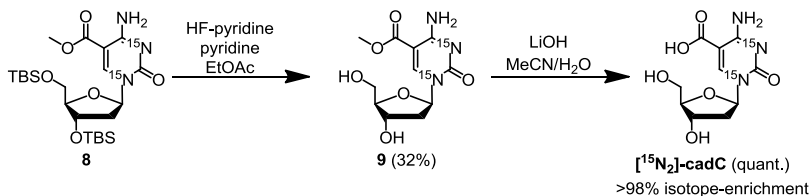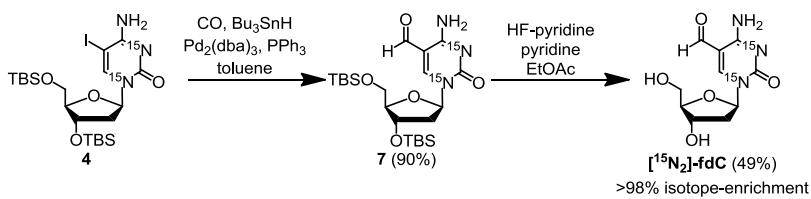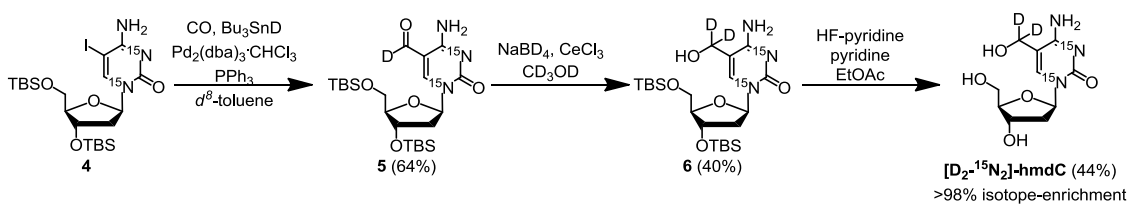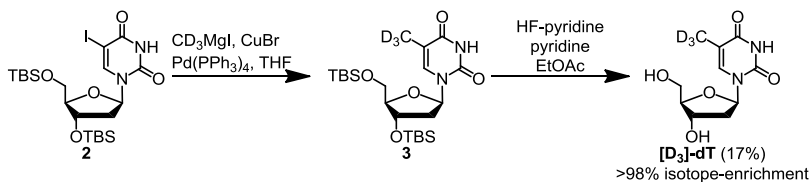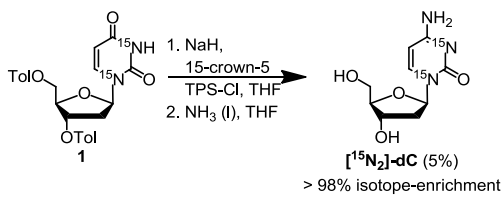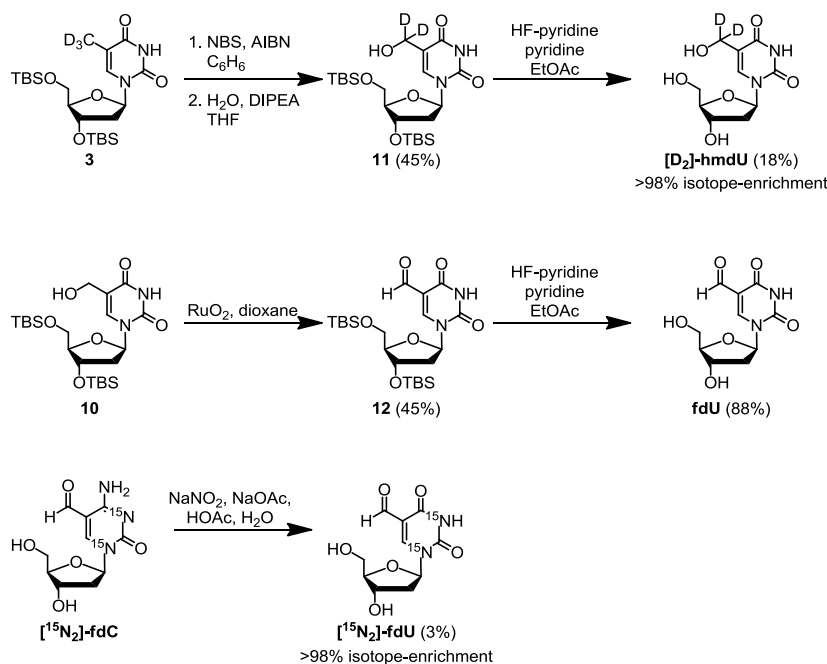
Scheme S1: Synthesis of standards for mass spectrometry.

2′-Deoxy-(N^1,N^3-^{15}N)-cytidine ([^{15}N$_2$]-dC)
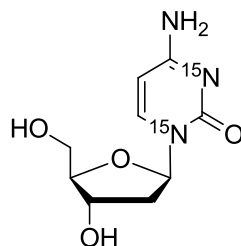
The synthesis was carried out using modifications of the methods of *Bischofberger et al.* and *Hodge et al.*[5,7] A stirred solution of 3′,5′-(toluoyl)-2′-deoxy-(N^1,N^3-^{15}N)-uridine **1**[4] (149.1 mg, 0.32 mmol, 1.0 eq.) and 15-crown-5 (0.05 mL, 0.25 mmol, 0.8 eq.) in THF (12 mL) was cooled to 0 °C, NaH in dispersion oil (60%, 38.5 mg, 0.98 mmol, 3.0 eq.) was added and the suspension stirred for 45 minutes. Then TPS-Cl (194.3 mg, 0.64 mmol, 2.0 eq.) was added and the mixture was further stirred at room temperature over night. The reaction was quenched with MeOH (2 mL) and a half sat. aqueous NH$_4$Cl solution (20 mL) was added. The reaction mixture was extracted with EtOAc (100 mL) and washed with sat. NH$_4$Cl (100 mL). The organic layer was separated, dried over MgSO$_4$ and subsequently evaporated to dryness. The crude product was then dissolved in THF (4 mL) and cooled to -78 °C. Then gaseous NH$_3$ was condensed (approx. 5 mL) into an overpressure vessel and the reaction was stirred at room temperature for 16 h. The reaction mixture was then evaporated to dryness and dissolved in MeOH (10 mL) and K$_2$CO$_3$ (221.0 mg, 1.6 mmol. 5.0 eq.) was added. The reaction was stirred for 1.5 h at room temperature and evaporated to dryness. The crude product was purified by RP-HPLC (0% to 10% MeCN in water in 45 minutes) yielding in 4.5 mg (5%) of [^{15}N$_2$]-dC as a white solid.

¹H-NMR (400 MHz, [D₆]-DMSO, ppm): δ = 7.76 (d, J=7.6 Hz, 1H, ¹⁵N-C**H**=CH), 7.11 (s, br, 1H, N**H**₂), 7.06 (s, br, 1H, N**H**₂), 6.13 (t, J=6.7 Hz, 1H, ¹⁵N-C**H**-O), 5.68 (dd, J=7.4, 3.5 Hz, 1H, ¹⁵N-CH=C**H**-C), 5.17 (d, J=4.1 Hz, 1H, CH₂-CH-O**H**), 4.94 (t, J=5.0 Hz, 1H, CH-CH₂-O**H**), 4.19 – 4.14 (m, 1H, CH₂-C**H**-OH), 3.76 – 3.71 (m, 1H, OH-CH₂-C**H**-O), 3.57 – 3.46 (m, 2H, C-C**H**₂-OH), 2.11 – 2.03 (m, 1H, O-CH-C**H**₂-CH-¹⁵N), 1.94 – 1.86 (m, 1H, O-CH-C**H**₂-CH-¹⁵N). **¹³C-NMR** (100 MHz, [D₆]-DMSO, ppm): δ = 165.9 (dd, J=5.2, 1.9 Hz, ¹⁵N-(**C**=O)-¹⁵N), 163.7 (**C**-NH₂), 141.4 (d, J=12.5 Hz, C=**C**-¹⁵N), 94.3 (**C**=C-¹⁵N), 87.6 (OH-CH₂-**C**H), 85.3 (d, J=11.6 Hz, ¹⁵N-**C**H-O), 70.9 (CH₂-**C**H-OH), 61.8 (**C**H₂-OH), 40.8 (N-CH-**C**H₂). **¹⁵N-NMR** (40 MHz, [D₆]-DMSO, ppm): δ = -223.9, -168.6. **HRMS** (ESI+): calc. for $C_9H_{14}N^{15}N_2O_4^+$ [M+H]⁺: 230.0920, found: 230.0921. **UV** (H₂O, 74 µM): λ_{max} = 271 nm, ε = 9490 mol⁻¹ L cm⁻¹.

5-(Trideuteromethyl)-3′,5′-(*tert*-butyl-dimethylsilyl)-2′-deoxyuridine (3)
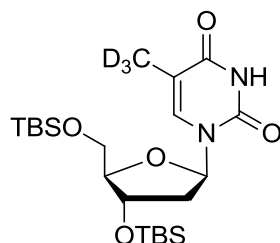
The synthesis was based on the method used for the synthesis of [D₃]-mdC.[1] 5-Iodo-3′,5′-bis-(*tert*-butyl-dimethylsilyl)-2′-deoxyuridine **2**[8] (2.00 g, 3.43 mmol, 1.0 eq.) was dissolved in THF (25 mL), cooled to 0 °C and CD₃MgI (1.0 M in Et₂O, 3.8 mL, 3.8 mmol, 1.1 eq.) were added. The reaction mixture was stirred at 0 °C for 30 min. Subsequently, Pd(PPh₃)₄ (397 mg, 0.34 mmol, 0.1 eq.), CD₃MgI (1.0 M in Et₂O, 18.9 mL, 18.9 mmol, 5.0 eq) and CuBr (2.46 g, 17.2 mmol, 5.0 eq.) were added and the reaction mixture stirred overnight at room temperature. Sat. NH₄Cl (2 mL) was added to the reaction mixture and the emulsion then extracted with EtOAc (100 mL). The organic layers were washed with NH₄Cl (100 mL), dried over MgSO₄ and concentrated *in vacuo*. The crude product was purified by column chromatography (DCM/MeOH 100:1) to yield an inseparable mixture of **3** and TBS-dU (1:1.7) as a pale yellow oil (combined yield: 65 %). The product was directly used in the next reaction.

HRMS (ESI-): calc. for $C_{22}H_{38}D_3N_2O_5Si_2^-$ [M-H]⁻: 472.2748, found: 472.2746.

5-(Trideuteromethyl)-2′-deoxyuridine ([D₃]-dT)
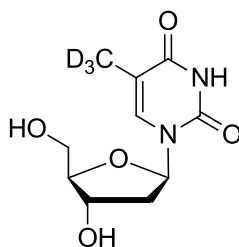
In a polypropylene tube the mixture of **3** and TBS-dU (65 mg, 0.14 mmol, 1.0 eq.) was dissolved in EtOAc (6 mL), then pyridine (150 μL, 1.85 mmol, 13.2 eq.) and HF-pyridine (70% HF, 150 μL, 5.77 mmol, 41.2 eq.) were added and the reaction mixture was stirred at room temperature for 21 h. During this time a white precipitate was formed. Methoxytrimethylsilane (1 mL, 7.25 mmol, 145.1 eq.) was added and the reaction mixture was stirred for another 30 min. Subsequently the solid was collected by centrifugation (6000 rpm, 15 min) and dried *in vacuo*. The product was purified in two steps by RP-HPLC (0% to 10% MeCN in water in 45 min) to give **[D₃]-dT** as a white solid (2.1 mg, 0.009 mmol, 17% based on pure **3**).

¹H-NMR (600 MHz, D₂O, ppm): δ = 7.67 (s, 1H, N-C**H**=C), 6.32 (t, *J*=6.8 Hz, 1H, O-C**H**-N), 4.50 – 4.48 (m, 1H, CH-C**H**-O), 4.09 – 4.00 (m, 1H, O-CH-C**H**-CH₂-O), 3.90 – 3.83 (m, 1H, O-C**H₂**-CH-CH-O), 3.83 – 3.75 (m, 1H, O-C**H₂**-CH-CH-O), 2.46 – 2.36 (m, 2H, O-CH-C**H₂**-CH-N). **¹³C-NMR** (150 MHz, D₂O, ppm): δ = 166.7 (**C**=O-C), 151.8 (N-**C**=O-N), 137.4 (N-**C**H=C), 111.3(N-CH=**C**-CD₃), 86.4 (**C**H-CH₂-O), 84.9 (N-**C**H-O), 70.3 (CH-**C**H-O), 61.1 (**C**H₂-O), 38.4 (**C**H₂-CH-N), 10.6 (CH=C-**C**D₃, only detected by HMBC-NMR). **HRMS** (ESI+): calc. for $C_{10}H_{12}D_3N_2O_5^+$ [M+H⁺]⁺: 246.1164, found: 246.1164. **HRMS** (ESI-): calc. for $C_{10}H_{10}D_3N_2O_5^+$ [M-H]⁻: 244.1018, found: 244.1021. **IR** (ATR): ṽ (cm⁻¹) = 3308 (br, w), 2836 (w), 1700 (s), 1654 (s), 1474 (m), 1435 (m), 1276 (m), 1067 (s), 1017 (m), 872 (m), 852 (m), 626 (m). **Melting range**: > 250 °C decomposition. **UV** (H₂O, 369 μM): λ_max = 268 nm.

5-Deuteroformyl-3′,5′-bis-(*tert*-butyl-dimethylsilyl)-2′-deoxy-(*N*¹,*N*³-¹⁵N)-cytidine (5)
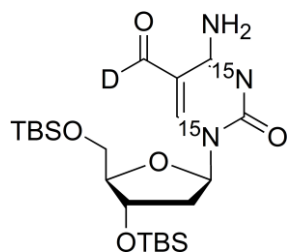
5′-Iodo-3′,5′-bis-(*tert*-butyl-dimethylsilyl)-2′-deoxy-(*N*¹,*N*³-¹⁵N)-cytidine[4] (**4**) (0.20 g, 0.34 mmol, 1.0 eq.), Pd₂(dba)₃·CHCl₃ (56.0 mg, 0.05 mmol, 0.2 eq.) and triphenylphosphine (85.0 mg, 0.32 mmol, 0.9 eq.) were dissolved in [D₈]-toluene (7.8 mL) in a high pressure glas autoclave. The autoclave was flushed with CO to remove residual air and after heating to 60 °C, the CO pressure was set to 3.5 bar. Tributyltin deuteride (0.18 mL, 0.68 mmol, 2.0 eq.) in *d*⁸-toluene (0.82 mL) were added within 10 h and the reaction was stirred for additional 14 h at 60 °C and 3.5 bar CO pressure. After evaporation to dryness, the product was purified by column chromatography (*i*Hex/EtOAc 1:1) to give product **5** as an orange solid (106 mg, 0.22 mmol, 64%).

¹H-NMR (300 MHz, CDCl₃, ppm): δ = 8.54 (s, 1H, N-C**H**=C), 8.18 (s br, 1H, N**H₂**), 6.40 (s br, 1H, N**H₂**), 6.21 (dt, *J*=6.2, 0.9 Hz, 1H, N-C**H**-O), 4.36 – 4.34 (m, 1H, C**H**-O-Si), 4.04 (dd, *J*=5.9, 2.6 Hz, 1H, C**H**-CH₂-O), 3.95 (dd, *J*=11.6, 2.6 Hz, 1H, C**H₂**-O-Si), 3.78 (dd, *J*=11.6, 2.5 Hz, 1H, C**H₂**-O-Si), 2.64 – 2.59 (m, 1H, N-CH-C**H₂**), 2.07 (dt, *J*=6.2, 2.6 Hz, 1H, N-CH-C**H₂**), 0.90 (s, 9H, Si-C-C**H₃**),

0.89 (s, 9H, Si-C-C$\underline{\text{H}}_3$), 0.10 (s, 3H, Si-C$\underline{\text{H}}_3$), 0.09 (s, 3H, Si-C$\underline{\text{H}}_3$), 0.08 (s, 3H, Si-C$\underline{\text{H}}_3$), 0.06 (s, 3H, Si-C$\underline{\text{H}}_3$). ^{13}**C-NMR** (151 MHz, CDCl$_3$, ppm): δ =186.8 (t, J=26.6 Hz, COD), 162.7 (d, J=5.9 Hz, $\underline{\text{C}}$-NH$_2$), 153.3 (dd, J=9.4, 7.2 Hz, $\underline{\text{C}}$=O-N), 153.0 (d, J=14.1, $\underline{\text{C}}$H=C), 104.8 ($\underline{\text{C}}$=CO), 88.6 (O-$\underline{\text{C}}$H-CH$_2$-O), 87.7 (d, J=9.5 Hz, N-$\underline{\text{C}}$H-O), 71.5 ($\underline{\text{C}}$H-O-Si), 62.6 ($\underline{\text{C}}$H$_2$-O-Si), 42.9 (N-CH-$\underline{\text{C}}$H$_2$), 25.9 (Si-C-$\underline{\text{C}}$H$_3$), 25.7 (Si-C-$\underline{\text{C}}$H$_3$), 18.4 (Si-$\underline{\text{C}}$), 18.0 (Si-$\underline{\text{C}}$), -4.5 (Si-$\underline{\text{C}}$H$_3$), -4.9 (Si-$\underline{\text{C}}$H$_3$), -5.3 (Si-$\underline{\text{C}}$H$_3$), -5.3 (Si-$\underline{\text{C}}$H$_3$). ^{15}**N-NMR** (40 MHz, CDCl$_3$, ppm): δ =-172.7, -207.3. **HRMS** (ESI+): calc. for C$_{22}$H$_{41}$^{2}HN^{15}N$_2$O$_5$Si$_2$ [M+H]$^+$: 487.2660, found: 487.2661. **IR** (ATR): $\tilde{v}$ (cm^{-1}) = 3366 (w), 2954 (m), 2927 (m), 2857 (m), 1639 (s), 1504 (m), 1471 (m), 1462 (m), 1251 (m), 1084 (s), 829 (s), 775 (s). **Melting range:** 137 – 139 °C.

5-Dideuterohydroxymethyl-3′,5′-bis-(*tert*-butyl-dimethylsilyl)-2′-deoxy-(N^1,N^3-^{15}N)-cytidine (6)
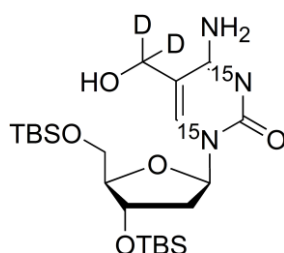
To 5-deuteroformyl-3′,5′-bis-(*tert*-butyl-dimethylsilyl)-2′-deoxy-(N^1,N^3-^{15}N)-cytidine (**5**) (77.0 mg, 0.16 mmol, 1.0 eq.) in CD$_3$OD (5.3 mL) were added sodium borodeuteride (7.5 mg, 0.18 mmol, 1.1 eq.) and cerium(III) chloride (0.12 g, 0.49 mmol, 3.0 eq.). The resulting orange solution was stirred at room temperature for 25 min and then quenched with a saturated ammonium chloride solution (53 mL). The aqueous phase was extracted with EtOAc (53 mL), the resulting organic phase was washed with a saturated ammonium chloride solution (2 × 53 mL) and dried over magnesium sulfate. The solvent was evaporated to dryness and the product purified by column chromatography (DCM/MeOH 20:1) to obtain **6** as a colorless solid (31.0 mg, 0.06 mmol, 40%).

^{1}**H-NMR** (600 MHz, CDCl$_3$, ppm): δ = 7.59 (d, J=1.1 Hz, 1H, N-C$\underline{\text{H}}$=C), 6.12 (t, J=6.4 Hz, 1H, N-C$\underline{\text{H}}$-O), 4.30 (dd, J=6.6, 3.5 Hz, 1H, C$\underline{\text{H}}$-O-Si), 3.91 (q, J=3.1 Hz, 1H, C$\underline{\text{H}}$-CH$_2$-O), 3.82 (dd, J=11.2, 3.2 Hz, 1H, C$\underline{\text{H}}_2$-O-Si), 3.72 (dd, J=11.2, 3.0 Hz, 1H, C$\underline{\text{H}}_2$-O-Si), 2.40 – 2.36 (m, 1H, N-CH-C$\underline{\text{H}}_2$), 1.96 – 1.91 (m, 1H, N-CH-C$\underline{\text{H}}_2$), 0.89 (s, 9H, Si-C-C$\underline{\text{H}}_3$), 0.87 (s, 9H, Si-C-C$\underline{\text{H}}_3$), 0.08 (s, 3H, Si-C$\underline{\text{H}}_3$), 0.08 (s, 3H, Si-C$\underline{\text{H}}_3$), 0.05 (s, 3H, Si-C$\underline{\text{H}}_3$), 0.05 (s, 3H, Si-C$\underline{\text{H}}_3$). ^{13}**C-NMR** (151 MHz, CDCl$_3$, ppm): δ = 165.3 (d, J=0.7 Hz, $\underline{\text{C}}$-NH$_2$), 156.3 (dd, J=12.6, 8.1 Hz, $\underline{\text{C}}$=O-N), 138.5 (d, J=13.2 Hz, $\underline{\text{C}}$H=C), 105.9 (d, J=0.7 Hz, $\underline{\text{C}}$=CD$_2$), 87.8 (O-$\underline{\text{C}}$H-CH$_2$-O), 86.2 (d, J=11.2 Hz, N-$\underline{\text{C}}$H-O), 71.7 ($\underline{\text{C}}$H-O-Si), 62.7 ($\underline{\text{C}}$H$_2$-O-Si), 58.8 (s br, $\underline{\text{C}}$D$_2$), 42.2 (N-CH-$\underline{\text{C}}$H$_2$), 25.9 (Si-C-$\underline{\text{C}}$H$_3$), 25.8 (Si-C-$\underline{\text{C}}$H$_3$), 18.4 (Si-$\underline{\text{C}}$), 18.0 (Si-$\underline{\text{C}}$), -4.6 (Si-$\underline{\text{C}}$H$_3$), -4.9 (Si-$\underline{\text{C}}$H$_3$), -5.3 (Si-$\underline{\text{C}}$H$_3$), -5.4 (Si-$\underline{\text{C}}$H$_3$). ^{15}**N-NMR** (40 MHz, CDCl$_3$, ppm): δ =-173.6, -223.3. **HRMS** (ESI+): calc. for C$_{22}$H$_{42}$^{2}H$_2$N^{15}N$_2$O$_5$Si$_2$ [M+H]$^+$: 490.2880, found: 490.2878. **IR** (ATR): $\tilde{v}$ (cm^{-1}) = 3202 (w), 2928 (w), 2856 (w), 1657 (m), 1604 (m), 1471 (m), 1253 (m), 1075 (m), 1029 (m), 832 (s), 775 (s). **Melting range:** 95 – 97 °C.

5-Dideuterohydroxymethyl-2′-deoxy-(N^1,N^3-^{15}N)-cytidine ([D$_2$,^{15}N$_2$]-hmdC)
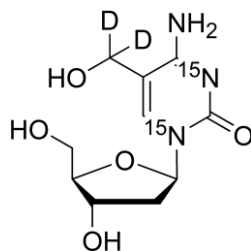
To a solution of 5-dideuterohydroxymethyl-3′,5′-bis-(*tert*-butyl-dimethylsilyl)-2′-deoxy-(N^1,N^3-^{15}N)-cytidine (**6**) (25.0 mg, 0.05 mmol, 1.0 eq.) in EtOAc (0.8 mL) were added pyridine (19.8 µL, 0.24 mmol, 4.9 eq.) and HF-pyridine (20.6 µL, 0.23 mmol, 4.6 eq.) and the reaction mixture stirred at room temperature for 21 h. After quenching with methoxytrimethylsilane (0.25 mL, 1.81 mmol, 36.3 eq.) and evaporating to dryness the crude product was purified by RP-HPLC (0% to 25% MeCN in water in 45 min) to give **[D$_2$,^{15}N$_2$]-hmdC** as a white solid (6.1 mg, 0.02 mmol, 40%).

^{1}H-NMR (600 MHz, D$_2$O, ppm): δ = 7.90 (s, 1H, N-C**H**=C), 6.27 – 6.26 (m, N-C**H**-O), 4.45 – 4.44 (m, 1H, C**H**-OH), 4.07 – 4.06 (m, 1H, C**H**-CH$_2$-O), 3.88 – 3.85 (m, 1H, C**H$_2$**-OH), 3.79 – 3.76 (m, 1H, C**H$_2$**-OH), 2.47 – 2.44 (m, 1H, N-CH-C**H$_2$**), 2.34 – 2.30 (m, 1H, N-CH-C**H$_2$**). **^{13}C-NMR** (100 MHz, D$_2$O, ppm): δ = 164.9 (dd, *J*=6.3, 1.7 Hz, **C**-NH$_2$), 157.2 (dd, *J*=13.5, 8.8 Hz, **C**=O-N), 140.4 (d, *J*=13.2 Hz, **C**H=C), 106.4 (**C**-CD$_2$), 86.5 (O-CH$_2$-**C**H), 86.0 (d, *J*=11.1 Hz, N-**C**H-O), 70.2 (**C**H-OH), 61.0 (**C**H$_2$-OH), 57.0 (qi, *J*=22.5 Hz, **C**D$_2$), 39.3 (N-CH-**C**H$_2$). **^{15}N-NMR** (40 MHz, CDCl$_3$, ppm): δ =-178.0, -223.2. **HRMS** (ESI+): calc. for C$_{10}$H$_{13}$^{2}H$_2$N^{15}N$_2$O$_5$Na [M+Na]$^+$: 284.0971, found: 284.0974. **IR** (ATR): υ̃ (cm^{-1}) = 2361 (w), 1648 (s), 1614 (s), 1515 (s), 1483 (s), 1043 (s), 928 (s), 783 (s). **Melting range:** > 250 °C (decomposition). **UV** (H$_2$O, 255 µM): λ$_{max}$ = 274 nm, ε = 8700 mol^{-1} L cm^{-1}.

5-Formyl-2′-deoxy-(N^1,N^3-^{15}N)-cytidine ([^{15}N$_2$]-fdC)
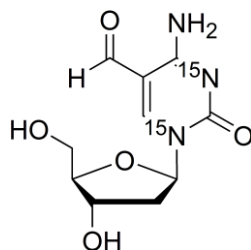
Compound **4**[4] (1.0 g, 1.71 mmol, 1.0 eq.), triphenylphosphine (0.27 g, 1.03 mmol, 0.6 eq.) and Pd$_2$(dba)$_3$ (0.17 g, 0.17 mmol, 0.1 eq.) were dissolved in toluene (40 mL) in a high pressure glas autoclave. The autoclave was flushed with CO to remove residual air and after heating to 60 °C, the CO pressure was set to 3.5 bar. Bu$_3$SnH (0.55 mL, 2.06 mmol, 1.2 eq.) was added dropwise (0.1 mL/h). The reaction mixture was stirred at 60 °C for 15 h. After evaporation to dryness, the crude product was purified by column chromatography (*i*Hex/EtOAc 4:1) to give 0.88 g (96 %) of **7** as a light brown solid.

To a solution of **7** (49.0 mg, 0.10 mmol, 1.0 eq.) in EtOAc (1.4 mL) was added pyridine (40 µL, 0.49 mmol, 5.0 eq.) and HF-pyridine (30 µL, 0.47 mmol, 15.0 eq.). The reaction mixture was stirred at

room temperature for 17 h. The reaction was stopped by addition of TMSOMe (0.284 mL, 2.45 mmol, 25.0 eq.) at 0 °C. After stirring for 30 min the white solid was collected by centrifugation (4000 rpm) and dried *in vacuo*. The solid was dissolved in water and purified by RP-HPLC (0 to 40 % MeCN in water in 45 min) which afforded **[^{15}N$_2$]-fdC** (14 mg, 0.05 mmol, 50%) as a white solid.

^{1}H-NMR (200 MHz, D$_2$O/CD$_3$OD, ppm): δ = 9.54 (s, 1H, C-(C=O)-**H**), 8.98 (s, 1H, N-C**H**=C), 6.17 (t, *J*=6.2 Hz, 1H, N-C**H**-O), 4.40 (dd, *J*=11.2, 4.5 Hz, 1H, O-C**H**-CH$_2$-OH), 4.08 − 4.02 (m, 1H, O-CH-C**H**-OH), 3.92 (dd, *J*=12.3, 3.1 Hz, 1H, HO-C**H$_2$**-CH-O), 3.78 (dd, *J*=12.3, 3.8 Hz, 1H, HO-C**H$_2$**-CH-O), 2.61 − 2.48 (m, 1H, HO-CH-C**H$_2$**-CH-N), 2.32 − 2.18 (m, 1H, HO-CH-C**H$_2$**-CH-N). **^{13}C-NMR** (150 MHz, D$_2$O/CD$_3$OD, ppm): δ =198.5 (d, *J*=2.0 Hz, H-**C**=O-C), 162.6 (dd, *J*=5.9, 1.2 Hz, N=**C**-NH$_2$), 154.4 (d, *J*=14.0 Hz, N-**C**H=C), 154.3 (N-(**C**=O)-N), 105.4 (C=**C**-C=N), 87.6 (OH-CH-**C**H-O), 87.4 (d, *J*=9.5 Hz, N-**C**H-O), 69.7 (OH-**C**H-CH-O), 60.6 (OH-**C**H$_2$-CH-O), 40.8 (N-CH-**C**H$_2$-CH). **^{15}N-NMR** (40 MHz, D$_2$O/CD$_3$OD, ppm): δ = -181.7, -212.16. **HRMS** (ESI+): calc. for C$_{10}$H$_{14}$N$_{15}$N$_2$O$_5$ ([M+H]$^+$): 258.0869, found: 258.0867. **UV** (H$_2$O, 152 µM): λ$_{max}$ = 284 nm, ε = 11270 mol^{-1} L cm^{-1}.

5-Carboxymethyl-2′-deoxy-(N^1,N^3-^{15}N)-cytidine (9)
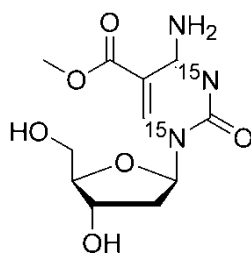
To a solution of 3′,5′-bis-(*tert*-butyl-dimethylsilyl)-5-carboxymethyl-2′-deoxy-(N^1,N^3-^{15}N)-cytidine[4] (**8**) (0.10 g, 0.19 mmol, 1.0 eq.) in EtOAc (6.3 mL) and pyridine (0.08 mL, 1.00 mmol, 5.3 eq.) was added HF-pyridine (0.08 mL, 0.89 mmol, 4.67 eq.). The reaction mixture was stirred at room temperature for 22 h. Then fresh HF-pyridine (0.05 mL, 0.55 mmol, 2.86 eq.) and pyridine (0.05 mL, 0.62 mmol, 3.19 eq.) were added and stirring was continued for 20 h. The reaction was quenched by the addition of methoxytrimethylsilane (1.55 mL, 11.24 mmol, 57.96 eq.) After centrifugation the white solid was purified by RP-HPLC (0% to 70% MeCN in water in 45 min). Product **9** was obtained as a colorless powder (17.8 mg, 32%).

^{1}H-NMR (200 MHz, D$_2$O, ppm): δ = 8.96 (s, 1H, N-C**H**=C), 6.16 (t, *J*=5.8 Hz, 1H, N-C**H**-O), 4.45 − 4.36 (m, 1H, C**H**-OH), 4.08 − 4.02 (m, 1H, C**H**-CH$_2$-OH), 3.94 − 3.71 (m, 2H, C**H$_2$**-OH), 3.83 (s, 3H, OC**H$_3$**), 2.58 − 2.43 (m, 1H, O-CH-C**H$_2$**), 2.41 − 2.26 (m, 1H, O-CH-C**H$_2$**). **^{13}C-NMR** (101 MHz, D$_2$O, ppm): δ =166.0 (d, *J*=1.6 Hz, O-**C**O), 163.5 (dd, *J*=6.2, 1.4 Hz, **C**-NH$_2$), 155.8 (dd, *J*=11.8, 8.3 Hz, **C**=O-N), 148.8 (d, *J*=14.0 Hz, **C**H=C), 96.8 (d, *J*=1.0 Hz, **C**=CO), 86.9 (d, *J*=9.7 Hz, N-**C**H-CH$_2$), 86.7 (**C**H-CH$_2$-O), 69.2 (**C**H-OH), 60.1 (**C**H$_2$-OH), 52.3 (O**C**H$_3$), 39.9 (O-**C**H-CH$_2$). **^{15}N-NMR** (40 MHz, D$_2$O, ppm): δ = -179.7, -212.1. **HRMS** (ESI+): calc. for C$_{11}$H$_{16}$N^{15}N$_2$O$_6$ [M+H]$^+$: 288.0974, found: 288.0977. **IR** (ATR): $\tilde{\upsilon}$ (cm^{-1}) = 2546 (w), 1713 (m), 1635 (s), 1480 (m), 1321 (s), 1241 (m),

1097 (s), 1063 (s), 787 (s). **Melting range:** 169 – 171 °C.

5-Carboxyl-2′-deoxy-(N^1,N^3-^{15}N)-cytidine ([^{15}N$_2$]-cadC)
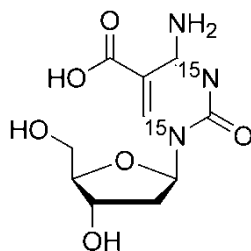
To 5-carboxymethyl-2′-deoxy-(N^1,N^3-^{15}N)-cytidine (**9**) (2.0 mg, 6.96 μmol, 1.0 eq.) dissolved in acetonitrile (1.76 mL) and water (0.52 mL) was added lithium hydroxide (1.99 mg, 83.10 μmol, 11.9 eq.). After stirring for 2 d the solvent was removed. The solid was dissolved in water and purified by RP-HPLC (0 to 25% MeCN in water in 45 min) to obtain **[^{15}N$_2$]-cadC** as a colorless solid (2.62 mg, quant.).

^{1}H-NMR (400 MHz, D$_2$O): δ (ppm) = 8.79 (s, 1H, N-C**H**=C), 6.29 (dt, J=6.5, 1.3 Hz, 1H, N-C**H**-O), 4.53–4.49 (m, 1H, C**H**-OH), 4.16 (dd, J=8.1, 4.2 Hz, 1H, C**H**-CH$_2$-O), 3.92 (dd, J=12.4, 3.5 Hz, 1H, C**H$_2$**-O), 3.83 (dd, J=12.5, 5.3 Hz, 1H, C**H$_2$**-O), 2.59–2.53 (m, 1H, N-CH-C**H$_2$**), 2.46–2.38 (m, 1H, N-CH-C**H$_2$**). **^{13}C-NMR** (101 MHz, D$_2$O, ppm): δ =168.0 (**C**O-O), 159.8 (d, J=13.3 Hz, **C**-NH$_2$), 150.0 (N-**C**=O), 148.0 (d, J=13.1 Hz, **C**H=C), 101.3 (**C**-CO), 87.2 (**C**H-CH$_2$-O), 86.9 (d, J=10.2 Hz, N-**C**H-CH$_2$), 70.1 (**C**H-OH), 60.8 (**C**H$_2$-OH), 39.5 (O-**C**H-CH$_2$). **^{15}N-NMR** (41 MHz, [D$_6$]-DMSO, ppm): δ = -171.6, -212.1. **HRMS** (ESI+): calc. for C$_{10}$H$_{14}$N^{15}N$_2$O$_6^+$ [M+H]$^+$: 274.0818, found: 274.0818. **IR** (ATR): ṽ (cm^{-1}) = 3402 (m), 3274 (br, m), 1661 (s), 1627 (m), 1448 (m), 1291 (m), 1202 (m), 1093 (s), 1003 (s), 816 (m). **Melting range**: > 250 °C decomposition. **UV** (H$_2$O, 108 μM): λ$_{max}$ = 283 nm, ε = 7060 mol^{-1} L cm^{-1}.

2′-Deoxy-(N^1,N^3-^{15}N)-uridine ([^{15}N$_2$]-dU)
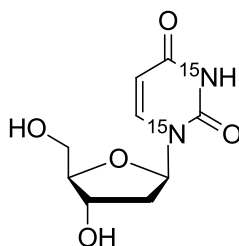
The precursor 3´,5´-(toluoyl)-2´-deoxy-(N^1,N^3-^{15}N)-uridine (**1**) was synthesized according to *Schiesser et al.*[4] In a round bottom flask **1** (76 mg, 0.16 mmol, 1.0 eq.) was dissolved in dry MeOH (2.1 mL) and K$_2$CO$_3$ (49 mg, 0.35 mmol, 2.2 eq.) was added. The suspension was stirred at 40 °C for 6 h. The solvent was removed by rotary evaporation. The residue was then suspended in H$_2$O (5 mL) and extracted with DCM (5 mL). The aqueous layer was then concentrated to dryness, redissolved in H$_2$O and subjected to RP-HPLC (0% to 20% MeCN in water in 45 min, 5 mL/min) to give **[^{15}N$_2$]-dU** as a white solid (31 mg, 0.13 mmol, 84%).

¹H-NMR (400 MHz, [D₆]-DMSO, ppm): δ = 11.27 (d, J=89.3 Hz, 1H, ¹⁵N-**H**), 7.84 (dd, J=8.1, 2.0 Hz, 1H, ¹⁵N-C**H**), 6.15 (t, J=6.7 Hz, 1H, O-C**H**-N), 5.62 (ddd, J=7.9, 4.5, 2.5 Hz, 1H, C**H**-C=O), 5.23 (s, 1H, OH), 5.00 (s, 1H, OH), 4.27 – 4.20 (m, 1H, O-CH₂-CH-C**H**-O), 3.79 – 3.76 (m, 1H, O-CH₂-C**H**-O), 3.59 – 3.51 (m, 2H, O-C**H₂**-CH-O), 2.13 – 2.02 (m, 2H, N-CH-C**H₂**). **¹³C-NMR** (100 MHz, [D₆]-DMSO, ppm): δ = 163.1 (d, J=9.2 Hz, **C**=O-C), 150.4 (dd, J=18.1, 17.6 Hz, ¹⁵N-**C**=O-¹⁵N), 140.5 (d, J=12.3 Hz, ¹⁵N-**C**H=CH), 101.7 (d, J=6.4 Hz, N-CH=**C**H), 87.4 (**C**H-CH₂-O), 84.1 (d, J=12.1 Hz, N-**C**H-O), 70.4 (d, J=0.9 Hz, CH-**C**H-O), 61.3 (**C**H₂-O), 40.4 (**C**H₂-CH-N). **¹⁵N-NMR** (40 MHz, [D₆]-DMSO, ppm): δ = -233.1, -222.4. **HRMS** (ESI-): calc. for C₉H₁₁¹⁵N₂O₅⁻ [M-H]⁻: 229.0614, found: 229.0616. **IR** (ATR): ṽ (cm⁻¹) = 3266 (w), 3127 (w), 3001 (w), 2936 (w), 2795 (w), 1660 (s), 1450 (m), 1388 (m), 1367 (m), 1250 (s), 1179 (m), 1098 (s), 1046 (s), 998 (s), 958 (s), 860 (s), 760 (s), 626 (s). **Melting range**: > 250 °C decomposition. **UV** (H₂O, 161 μM): λ_max = 262 nm.

5-Hydroxymethyl-2′-deoxyuridine (hmdU)
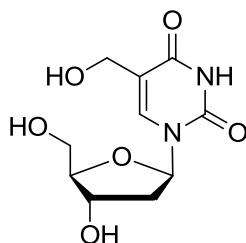
The synthesis was performed with slight modifications according to *Globisch and Münzel et al.*[2]

In a polypropylene tube 5-(hydroxymethyl)-3′,5′-(*tert*-butyl-dimethylsilyl)-2′-deoxyuridine (**10**) (182 mg, 0.37 mmol, 1.0 eq.) was dissolved in THF (20 mL) and EtOAc (10 mL), then pyridine (310 μL, 3.83 mmol, 10.4 eq.) and HF-pyridine (70% HF, 284 μL, 10.9 mmol, 29.5 eq.) were added and the reaction mixture was stirred at room temperature for 21 h. An additional amount of pyridine (310 μL, 3.83 mmol, 10.4 eq.) and HF-pyridine (70% HF, 284 μL, 10.9 mmol, 29.5 eq.) were added and the reaction mixture was further stirred at room temperature for 21 h. Methoxytrimethylsilane (3.6 mL, 26.1 mmol, 70.6 eq.) was added and the reaction mixture was stirred for another 30 min and then concentrated *in vacuo*. The residue was suspended in EtOAc (2.5 mL), collected by centrifugation (6000 rpm, 15 min) and washed another two times by repeating this procedure. The product was then purified by RP-HPLC (0% to 20% MeCN in water in 45 min) to give **hmdU** as a white solid (49 mg, 0.19 mmol, 52%).

¹H-NMR (400 MHz, D₂O, ppm): δ = 7.93 (s, 1H, N-C**H**=C-CH₂), 6.33 (t, J=6.6 Hz, 1H, O-C**H**-N), 4.53 – 4.50 (m, 1H, O-C**H**-CH₂-CH-N), 4.41 (s, 2H, C-C**H₂**-OH), 4.10 – 4.07 (m, 1H, O-CH-C**H**-CH₂-O), 3.90 (dd, J=12.5, 3.3 Hz, 1H, O-C**H₂**-CH-O), 3.81 (dd, J=12.5, 4.9 Hz, 1H, O-C**H₂**-CH-O), 2.53 – 2.37 (m, 2H, O-CH-C**H₂**-CH-N). **¹³C-NMR** (100 MHz, D₂O, ppm): δ = 164.9 (**C**=O-C), 151.4 (N-**C**=O-N), 139.5 (N-**C**H=C), 113.5 (N-CH=**C**-CH2), 86.6 (**C**H-CH₂-O), 85.3 (N-**C**H-O), 70.3 (CH-**C**H-O), 61.0 (**C**H₂-O), 56.5 (CH=C-**C**H₂-OH), 38.7 (**C**H₂-CH-N). **HRMS** (ESI-): calc. for C₁₀H₁₃N₂O₆⁻ [M-H]⁻: 257.0779, found: 257.0781. **IR** (ATR): ṽ (cm⁻¹) = 3433 (w), 3272 (br, w), 3032 (w), 2932 (w),

2836 (w), 1683 (s), 1467 (m), 1274 (m), 1078 (m), 1000 (s). **Melting range**: > 250 °C decomposition. **UV** (H_2O, 173 μM): λ_{max} = 264 nm, ε = 8600 mol^{-1} L cm^{-1}.

5-(Dideuterohydroxymethyl)-3′,5′-(*tert*-butyl-dimethylsilyl)-2′-deoxyuridine (11)
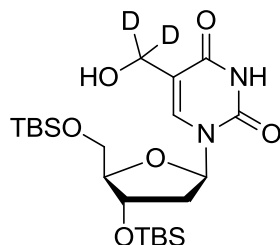
The mixture of **3** and TBS-dU (430 mg, 0.93 mmol, 1.0 eq.), AIBN (30 mg, 0.18 mmol, 0.2 eq.) and NBS (367 mg, 2.06 mmol, 2.2 eq.) were suspended in benzene (10 mL) and directly put into a preheated oil bath (100 °C). The solution was heated to reflux for 1 h and subsequently concentrated *in vacuo*. The crude product was dissolved in THF (8 mL) and DIPEA (0.3 mL, 1.72 mmol, 1.9 eq.) and H_2O (0.4 mL, 22.2 mmol, 23.9 eq.) were added. The solution was stirred for 1 h at 50 °C, subsequently diluted with EtOAc, extracted twice with diluted NH_4Cl and once with H_2O, dried over $MgSO_4$ and concentrated *in vacuo*. The crude product was purified by column chromatography (DCM/MeOH 99:1 → 50:1 → 10:1) to yield **11** (74 mg, 0.15 mmol, 45% based on pure **3**) as a pale yellow oil.

^{1}H-NMR (600 MHz, CDCl$_3$, ppm): δ = 7.66 (s, 1H, N-C**H**=C-CD$_2$), 6.28 (dd, *J*=7.7, 5.9 Hz, 1H, O-C**H**-N), 4.38 (dt, *J*=5.9, 2.5 Hz, 1H, O-C**H**-CH$_2$-CH-N), 3.94 – 3.92 (m, 1H, O-CH-C**H**-CH$_2$-O), 3.83 (dd, *J*=11.4, 2.8 Hz, 1H, O-C**H**$_2$-CH-O), 3.74 (dd, *J*=11.4, 2.6 Hz, 1H, O-C**H**$_2$-CH-O), 2.29 – 2.25 (m, 1H, O-CH-C**H**$_2$-CH-N), 2.03 – 1.96 (m, 1H, O-CH-C**H**$_2$-CH-N), 0.90 (s, 9 H, Si-C(C**H**$_3$)$_3$), 0.88 (s, 9H, Si-C(C**H**$_3$)$_3$), 0.09 (s, 6H, Si-C**H**$_3$), 0.06 (s, 3H, Si-C**H**$_3$), 0.06 (s, 3H, Si-C**H**$_3$). **^{13}C-NMR** (150 MHz, CDCl$_3$, ppm): δ = 163.6 (**C**=O-C), 150.0 (N-**C**=O-N), 137.5 (N-**C**H=C), 113.7 (N-CH=**C**-CD$_2$), 88.3 (**C**H-CH$_2$-O), 85.5 (N-**C**H-O), 72.4 (CH-**C**H-O), 63.2 (**C**H$_2$-O), 58.2 (CH=C-**C**D$_2$-OH, only detected by HMBC-NMR), 41.8 (**C**H$_2$-CH-N), 26.2 (Si-C-**C**H$_3$), 26.0 (Si-C-**C**H$_3$), 18.7 (Si-**C**), 18.2 (Si-**C**), -4.4 (Si-**C**H$_3$), -4.6 (Si-**C**H$_3$), -5.1 (Si-**C**H$_3$), -5.2 (Si-**C**H$_3$). **HRMS** (ESI-): calc. for C$_{22}$H$_{39}$D$_2$N$_2$O$_6$Si$_2$$^-$ [M-H]$^-$: 487.2634, found: 487.2631. **IR** (ATR): $\tilde{v}$ (cm^{-1}) = 3203 (br, w), 2954 (m), 2928 (m), 2857 (m), 1706 (s), 1693 (s), 1472 (m), 1464 (m), 1255 (m), 1103 (m), 1030 (m), 837 (s), 778 (s).

5-(Dideuterohydroxymethyl)-2′-deoxyuridine ([D₂]-hmdU)

In a polypropylene tube **11** (54 mg, 0.11 mmol, 1.0 eq.) was dissolved in EtOAc (3 mL), then pyridine (60 μL, 0.74 mmol, 6.7 eq.) and HF-pyridine (70% HF, 60 μL, 2.31 mmol, 21.0 eq.) were added and the reaction mixture was stirred at room temperature for 21 h. During this time a precipitate was formed. Methoxytrimethylsilane (400 μL, 2.90 mmol, 15.3 eq.) was added and the reaction mixture was stirred for another 30 min. Subsequently the solid was collected by centrifugation (6000 rpm, 15 min) and dried *in vacuo*. The product was dissolved in DMSO (0.5 mL), diluted with water (4.5 mL) and then purified by RP-HPLC (0% to 25% MeCN in water in 45 min) to give **[D₂]-hmdU** as a white solid (5.2 mg, 0.02 mmol, 18%).

¹H-NMR (600 MHz, [D₆]-DMSO, ppm): δ = 11.31 (s, 1H, N-**H**), 7.73 (s, 1H, N-C**H**=C-CD₂), 6.19 (t, *J*=6.9, 1H, O-C**H**-N), 5.24 (d, *J*=4.2, 1H, CH-O**H**), 4.96 (t, *J*=5.1, 1H, CH₂-O**H**), 4.85 (s, 1H, CD₂-O**H**), 4.24 – 4.21 (m, 1H, O-C**H**-CH₂-CH-N), 3.78 – 3.76 (m, 1H, **CH**-CH₂-O), 3.58 – 3.51 (m, 2H, O-C**H₂**-CH-CH-O), 2.11 – 2.05 (m, 2H, O-CH-C**H₂**-CH-N). **¹³C-NMR** (150 MHz, [D₆]-DMSO, ppm): δ = 163.1 (**C**=O-C), 150.8 (N-**C**=O-N), 137.3 (N-**C**H=C), 114.6 (N-CH=**C**-CD₂), 87.7 (**C**H-CH₂-O), 84.3 (N-**C**H-O), 71.0 (CH-**C**H-O), 61.9 (**C**H₂-O), 55.6 (CH=C-**C**D₂-OH, only detected by HMBC-NMR), 40.9 (**C**H₂-CH-N). **HRMS** (ESI-): calc. for C₁₀H₁₁D₂N₂O₆⁻ [M-H]⁻: 259.0905, found: 259.0908. **IR** (ATR): ṽ (cm⁻¹) = 3272 (br, m), 1676 (s). **Melting range**: > 250 °C decomposition. **UV** (H₂O, 801 μM): λ_max = 264 nm.

5-Formyl-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxyuridine (12)

A mixture of 5-Hydroxymethyl-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxyuridine **10**[1] (95 mg, 0.20 mmol, 1.0 eq.) and ruthenium(IV) oxide monohydrate (304 mg, 2.01 mmol, 10.1 eq.) in 1,4-dioxane (1.84 mL) was stirred at 110 °C for 14 h. After filtration, evaporation to dryness and column chromatography (DCM → DCM/MeOH 200:1) **12** was obtained as a colourless solid (45 mg, 0.09 mmol, 45%).

¹H-NMR (600 MHz, CDCl₃, ppm): δ = 10.00 (s, 1H, C**H**O), 8.90 (s br, 1H, N**H**), 8.54 (s, 1H, N-C**H**=C), 6.23 (dd, *J*=7.7, 5.7 Hz, 1H, N-C**H**-O), 4.42 – 4.41 (m, 1H, C**H**-OSi), 4.07 – 4.06 (m, 1H,

C**H**-CH$_2$-O), 3.88 (dd, *J*=11.5, 2.9 Hz, 1H, CH-C**H$_2$**-O), 3.78 (dd, *J*=11.5, 2.6 Hz, 1H, CH-C**H$_2$**-O), 2.44 (ddd, *J*=13.2, 5.7, 2.1 Hz, 1H, C**H$_2$**-CH-OSi), 2.05 (ddd, *J*=13.3, 7.7, 5.7 Hz, 1H, C**H$_2$**-CH-OSi), 0.90 (s, 9H, Si-C-C**H$_3$**), 0.89 (s, 9H, Si-C-C**H$_3$**), 0.11 (s, 3H, Si-C**H$_3$**), 0.11 (s, 3H, Si-C**H$_3$**), 0.09 (s, 3H, Si-C**H$_3$**), 0.08 (s, 3H, Si-C**H$_3$**). **^{13}C-NMR** (150 MHz, CDCl$_3$, ppm): δ = 185.6 (**C**HO), 161.8 (**C**=O-C), 149.1 (**C**=O-N), 145.4 (N-**C**H-C), 111.0 (**C**-CHO), 89.2 (**C**H-CH$_2$-O), 87.2 (N-**C**H-O), 72.9 (**C**H-O), 63.0 (**C**H$_2$-O), 42.5 (**C**H$_2$-CH-N), 25.9 (Si-C-**C**H$_3$), 25.7 (Si-C-**C**H$_3$), 18.4 (Si-**C**-CH$_3$), 18.0 (Si-**C**-CH$_3$), -4.7 (Si-**C**H$_3$), -4.8 (Si-**C**H$_3$), -5.5 (Si-**C**H$_3$), -5.7 (Si-**C**H$_3$). **HRMS** (ESI+): calc. for C$_{22}$H$_{40}$N$_2$NaO$_6$Si$_2$$^+$ [M+Na]$^+$: 507.2317, found: 507.2315. **IR** (ATR): ṽ (cm^{-1}) = 3190 (w), 3052 (w), 1725 (m), 1694 (m), 1600 (m), 1462 (m), 1470 (m), 1253 (m), 1097 (m), 1067 (m), 829 (s), 774 (s). **Melting range**: 79 – 80 °C.
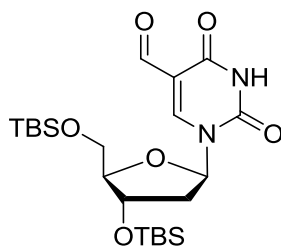
5-Formyl-2´-deoxyuridine (fdU)
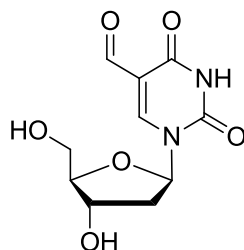
A mixture of 5-Formyl-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxyuridine **12** (45 mg, 0.09 mmol, 1.0 eq.), HF-pyridine (40.6 μL, 1.56 mmol, 17.4 eq.) and pyridine (39.5 μL, 0.49 mmol, 5.4 eq.) in EtOAc (3.3 mL) was stirred at room temperature for 18 h. After addition of further HF-pyridine (40.6 μL, 1.56 mmol, 17.4 eq.) and pyridine (39.5 μL, 0.49 mmol, 5.4 eq.) to ensure complete reaction stirring was continued for additional 44 h. The reaction was quenched by addition of methoxytrimethylsilane (0.99 mL, 7.18 mmol, 79.8 eq.). After evaporation to dryness the product was purified by RP-HPLC (0% buffer B to 15% buffer B in 45 min) to obtain **fdU** as a colourless solid (20.9 mg, 0.08 mmol, 88%).

^{1}H-NMR (600 MHz, D$_2$O, ppm): δ = 9.67 (s, 1H, C**H**O), 8.80 (s, 1H, N-C**H**=C), 6.28 – 6.26 (m, 1H, N-C**H**-O), 4.49 – 4.48 (m, 1H, C**H**-OH), 4.14 – 4.12 (m, 1H, C**H**-CH$_2$-O), 3.92 (d br, *J*=12.8 Hz, 1H, CH-C**H$_2$**-O), 3.82 – 3.79 (m, 1H, CH-C**H$_2$**-O), 2.57 – 2.53 (m, 1H, C**H$_2$**-CH-OH), 2.47 – 2.43 (m, 1H, C**H$_2$**-CH-OH). **^{13}C-NMR** (101 MHz, D$_2$O, ppm): δ = 188.9 (**C**HO), 163.0 (**C**=O-C), 151.1 (**C**=O-N), 150.4 (N-**C**H-C), 111.2 (**C**-CHO), 87.2 (**C**H-CH$_2$-O), 86.9 (N-**C**H-O), 69.8 (**C**H-O), 60.6 (**C**H$_2$-O), 39.5 (**C**H$_2$-CH-N). **HRMS** (ESI+): calc. for C$_{10}$H$_{12}$N$_2$NaO$_6$$^+$ [M+Na]$^+$: 279.0588, found: 279.0585. **IR** (ATR): ṽ (cm^{-1}) = 3345 (w), 1706 (m), 1475 (m), 1072 (s), 959 (m), 787 (m). **Melting range**: > 270 °C (decomposition). **UV** (H$_2$O, 136 μM): λ$_{max}$ = 282 nm, ε = 9110 mol^{-1} L cm^{-1}.

5-Formyl-2′-deoxy-(N^1,N^3-^{15}N)-uridine ([^{15}N$_2$]-fdU)
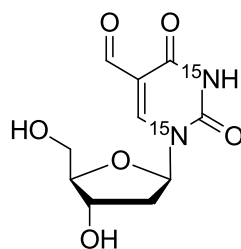
The synthesis was performed on a μmol-scale based on a method published by *Taghizadeh et al.* and references therein.[9] To an aqueous solution of **[^{15}N$_2$]-fdC** (30 mL, 152 μM, 4.56 μmol) aqueous NaNO$_2$ (3.6 mL, 2 M) and aqueous NaOAc (3.6 mL, 3 M, pH 4.4 with HOAc) were added and the mixture was stirred at 50 °C for 4 h. Afterwards, the mixture was neutralized with KOH, desalted by C18-Sep-Pak cartridges (Waters, WAT051910), dried *in vacuo* and redissolved in water. The product was purified by HPLC using a Nucleosil C18 column (250 mm × 4 mm, 3 μm, 120 Å, *Macherey Nagel*) with an elution performed by a linear gradient of 0% to 5% MeCN in water (0.5 mL/min). Fractions containing the product **[^{15}N$_2$]-fdU** were confirmed by LC-MS/MS. The reaction turnover was estimated by HPLC-UV to be 3%.

HRMS (ESI-): calc. for $C_{10}H_{11}{}^{15}N_2O_6^-$ [M-H]$^-$: 257.0563, found: 257.0565. **LC-MS/MS**: $C_{10}H_{11}{}^{15}N_2O_6^-$ [M-H]$^-$ → $C_9H_{10}{}^{15}NO_5^-$ [M-H^{15}NCO-H]$^-$: 257.1 → 213.1. **UV** (H$_2$O, 45 μM): λ_{max} = 282 nm.

2′-Deoxy-cytidine-5-bis-(*β*-hydroxyethyl)-dithioacetal (13)
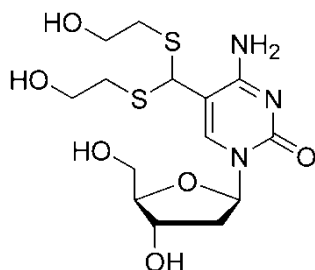
A solution of 5-formyl-3′,5′-bis-(*tert*-butyl-dimethylsilyl)-2′-deoxy-cytidine[10] (50 mg, 0.10 mmol, 1.0 eq.), triethyl orthoformate (0.16 mL, 0.96 mmol, 9.6 eq.) and tetrabutylammonium tribromide (41 mg, 0.09 mmol, 0.9 eq.) in 2-mercaptoethanol (0.25 mL, 3.56 mmol, 35.6 eq.) was stirred at room temperature for 4.5 h. The slightly yellow solution was aduste to pH 7.5 via addition of a saturated NaHSO$_4$-solution (1 mL) and extracted with EtOAc (2 × 10 mL). RP-HPLC-purification (0% to 15% MeCN in 45 min) afforded **13** as a white solid (18 mg, 0.05 mmol, 50%).

^{1}H-NMR (600 MHz, CD$_3$OD, ppm): δ = 8.39 (d, *J*=0.6 Hz, 1H, N-C**H**=C), 6.28 (t, *J*=6.4 Hz, 1H, N-C**H**-O), 5.19 (d, *J*=0.5 Hz, 1H, C**H**-S), 4.38 (dt, *J*=6.2, 3.6 Hz, 1H, C**H**-O), 3.97 (q, *J*=3.7 Hz, 1H, C**H**-CH$_2$-O), 3.83 – 3.69 (m, 6H, CH-C**H$_2$**-O, 2 × C**H$_2$**-OH), 2.84 – 2.78 (m, 2H, S-C**H$_2$**), 2.70 – 2.65 (m, 2H, S-C**H$_2$**), 2.41 (ddd, *J*=13.6, 6.2, 3.9 Hz, 1H, C**H$_2$**-CH-O), 2.19 – 2.14 (m, 1H, C**H$_2$**-CH-O).

^{13}C-NMR (101 MHz, CD$_3$OD, ppm): δ = 165.1 (**C**-NH$_2$), 157.6 (**C**=O-N), 142.3 (N-**C**H-C), 106.5 (**C**-

CH-S), 89.1 (<u>C</u>H-CH$_2$-O), 88.0 (N-<u>C</u>H-O), 72.3 (<u>C</u>H-O), 63.0 (<u>C</u>H$_2$), 62.8 (<u>C</u>H$_2$), 62.8 (<u>C</u>H$_2$), 48.4 (<u>C</u>H-S), 42.3 (<u>C</u>H$_2$-CH-N), 35.9 (S-<u>C</u>H$_2$), 35.8 (S-<u>C</u>H$_2$). **HRMS** (ESI-): calc. for C$_{14}$H$_{22}$N$_3$O$_6$S$_2^-$ [M-]$^-$: 392.0956, found: 392.0959. **IR** (ATR): $\tilde{\upsilon}$ (cm^{-1}) = 3286 (w), 1650 (m), 1479 (m), 1067 (s), 958 (m), 958 (m), 790 (m). **Melting range**: > 300 °C (decomposition).

LC-ESI-MS/MS method development and analysis

LC-ESI-MS/MS analysis was performed using an *Agilent* 1290 UHPLC system and an *Agilent* 6490 triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. DNA samples were digested to give a nucleoside mixture and spiked with specific amounts of the corresponding isotopically labeled standards before LC-MS/MS analysis (see below). The nucleosides were analyzed in the positive as well as in the negative ion selected reaction monitoring mode (SRM). In the positive ion mode [M+H]$^+$ species and in the negative ion mode [M-H]$^-$ species were measured. The specific MS/MS transitions which gave highest intensities during our method development are summarized in Table S2. MS/MS fragmentation patterns of these compounds were partly earlier reported by *Cao et al.*[11], *Wang et al.*[12] and cited references therein.

The general source-dependent parameters were optimized by the direct injection of 150 fmol of fdC (positive ion mode) and 200 fmol of hmdU (negative ion mode) in 1 µL H$_2$O without using a column at a flow rate of 0.4 mL min^{-1}. The mobile phase consisted of H$_2$O, 10% (v/v) MeCN and 0.01% (v/v) formic acid. The optimized general source-dependent parameters were as follows: Gas Temp 50 °C, Gas Flow 15 L/min (N$_2$), Nebulizer 30 psi, Sheath Gas Heater 275 °C, Sheath Gas Flow 11 L/min (N$_2$), Capillary Voltage 2500 V in the positive ion mode, Capillary Voltage -2250 V in the negative ion mode and Nozzle Voltage 500 V. The unusually low gas temperature of 50 °C minimized unwanted fragmentations of the nucleosides during the ionization process. Especially the fdC nucleoside is prone to de-glycosylation in the ion source.

For compound-dependent parameters see Table S2 – S7. The compound-dependent parameters were optimized to achieve linearity in each experiment, in which either highest sensitivity in the quantification of trace compounds (fmol range) or lower sensitivity in the quantification of highly abundant nucleosides (pmol range) was necessary.

For the analysis we used a Zorbax RRHD SB-C8 column from *Agilent* (1.8 µm, 2.1 mm × 150 mm). The compounds were separated by a gradient using water and MeCN, each containing 0.0075% (v/v) formic acid: 0 → 5 min; 0 → 3.5% MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 7.2 min; 5 → 80% MeCN; 7.2 → 10 min; 80% MeCN; 10 → 10.3 min; 80 → 0% MeCN; 10.3 → 12 min; 0% MeCN. The presence of 0.0075% formic acid in the mobile phase was important for a sensitive cadC-detection and avoided unwanted double peaks during chromatography. The column temperature was maintained at 30 °C. The flow rate was 0.4 mL min^{-1}, and the injection volume amounted to 10 µL. The effluent up to 1.5 min and after 9 min (total run time of 12 min) was diverted to waste by a Valco valve in order to protect the mass spectrometer. The auto-sampler was cooled to 10 °C.

Table S2. Compound-dependent LC-MS/MS-parameters used for the analysis of nucleoside mixtures obtained after treatment of oligonucleotides. The fragmentor voltage was 380 V. Delta EMV was set to 300 (positive mode) and 800 (negative mode). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. β-ME derivative of hmdC: 5-((2″-hydroxyethyl)thio)methyl-dC.

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1 (1.5−4.1 min)								
[^{15}N$_2$]-cadC	274.08	enhanced	158.03	enhanced	40	2	5	Positive
cadC	272.09	enhanced	156.04	enhanced	40	2	5	Positive
[^{15}N$_2$,D$_2$]-hmdC	262.12	enhanced	146.07	enhanced	40	1	1	Positive
hmdC	258.11	enhanced	142.06	enhanced	40	1	1	Positive
[^{15}N$_2$]-dC	230.1	wide	114.1	wide	40	5	5	Positive
dC	228.1	wide	112.1	wide	40	5	5	Positive
time segment 2 (4.1−9 min)								
[^{15}N$_5$]-8-oxo-dG	289.08	Wide	173.04	wide	120	9	7	Positive
8-oxo-dG	284.1	Wide	168.05	wide	120	9	7	Positive
cadU	273.07	wide	157.02	wide	20	5	5	Positive
cadU	271.06	wide	227.07	wide	20	7	5	Negative
[^{15}N$_2$]-fdC	258.09	enhanced	142.04	enhanced	20	2	5	Positive
fdC	256.09	enhanced	140.05	enhanced	20	2	5	Positive
[D$_3$]-dT	246.12	enhanced	130.07	enhanced	20	3	3	Positive
dT	243.1	enhanced	127.05	enhanced	20	3	3	Positive
[D$_2$]-hmdU	259.09	wide	216.08	wide	40	7	5	Negative
hmdU	257.08	wide	214.07	wide	40	7	5	Negative
[^{15}N$_2$]-fdU	257.06	wide	213.05	wide	40	6	5	Negative
fdU	255.06	wide	212.06	wide	40	6	5	Negative
[^{15}N$_2$]-dU	229.06	wide	185.06	wide	20	5	5	Negative
dU	227.07	wide	184.06	wide	20	5	5	Negative
13	394.11	wide	278.06	wide	20	5	5	Positive
13	394.11	wide	140.05	wide	20	5	5	Positive
β-ME derivative of hmdC	318.11	wide	202.06	wide	10	6	5	Positive

Table S3. Compound-dependent LC-MS/MS-parameters used for the analysis of the deamination kinetics of the dC nucleoside at 60 °C and 80 °C in buffered H$_2$O (pH 7.4). The fragmentor voltage was 380 V. Delta EMV was set to 300 (positive mode) and 800 (negative mode). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage.

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1 (1.5−4.1 min)								
[^{15}N$_2$]-dC	230.1	enhanced	114.1	enhanced	240	30	5	Positive
dC	228.1	enhanced	112.1	enhanced	240	30	5	Positive
time segment 2 (4.1−9 min)								
[^{15}N$_2$]-dU	229.06	wide	185.06	wide	240	5	5	Negative
dU	227.07	wide	184.06	wide	240	5	5	Negative

Table S4. Compound-dependent LC-MS/MS-parameters used for the analysis of the deamination and oxidation kinetics of the mdC nucleoside at 60 °C and 80 °C in buffered H_2O (pH 7.4). The fragmentor voltage was 380 V. Delta EMV was set to 500 (positive mode). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage.

compound	Precursor Ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1 (1.5−4.1 min)								
[$^{15}N_2$]-cadC	274.08	wide	158.03	wide	60	5	5	Positive
cadC	272.09	wide	156.04	wide	60	5	5	Positive
[$^{15}N_2,D_2$]-hmdC	262.12	wide	146.07	wide	60	5	5	Positive
hmdC	258.11	wide	142.06	wide	60	5	5	Positive
[D_3]-mdC	245.13	enhanced	129.09	enhanced	50	60	1	Positive
mdC	242.11	enhanced	126.07	enhanced	50	60	1	Positive
time segment 2 (4.1−9 min)								
[$^{15}N_2$]-fdC	258.09	wide	142.04	wide	120	5	5	Positive
fdC	256.09	wide	140.05	wide	120	5	5	Positive
[D_3]-dT	246.12	wide	130.07	wide	120	5	5	Positive
dT	243.1	wide	127.05	wide	120	5	5	Positive

Table S5. Compound-dependent LC-MS/MS-parameters used for the analysis of the deamination, oxidation and C-C bond cleavage kinetics of the hmdC nucleoside at 60 °C and 80 °C in buffered H_2O (pH 7.4). The fragmentor voltage was 380 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage.

compound	Precursor Ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1 (1.5−4.1 min)								
[$^{15}N_2$]-cadC	274.08	wide	158.03	wide	80	5	5	Positive
cadC	272.09	wide	156.04	wide	80	5	5	Positive
[$^{15}N_2,D_2$]-hmdC	262.12	enhanced	146.07	enhanced	80	30	3	Positive
hmdC	258.11	enhanced	142.06	enhanced	80	30	3	Positive
[$^{15}N_2$]-dC	230.1	wide	114.1	wide	80	5	5	Positive
dC	228.1	wide	112.1	wide	80	5	5	Positive
time segment 2 (4.1−9 min)								
[$^{15}N_2$]-fdC	258.09	wide	142.04	wide	40	5	5	Positive
fdC	256.09	wide	140.05	wide	40	5	5	Positive
[D_2]-hmdU	259.09	wide	216.08	wide	40	7	5	Negative
hmdU	257.08	wide	214.07	wide	40	7	5	Negative
[$^{15}N_2$]-fdU	257.06	wide	213.05	wide	40	6	5	Negative
fdU	255.06	wide	212.06	wide	40	6	5	Negative

Table S6. Compound-dependent LC-MS/MS-parameters used for the analysis of the deamination, oxidation and C-C bond cleavage kinetics of the fdC nucleoside at 60 °C and 80 °C in buffered H_2O (pH 7.4). The fragmentor voltage was 380 V. Delta EMV was set to 500 and 300 (positive mode) in time segment 1 and 2, respectively, and 800 (negative mode). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage.

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1 (1.5−4.1 min)								
[$^{15}N_2$]-cadC	274.08	wide	158.03	wide	80	5	5	Positive
cadC	272.09	wide	156.04	wide	80	5	5	Positive
[$^{15}N_2,D_2$]-hmdC	262.12	wide	146.07	wide	80	5	5	Positive
hmdC	258.11	wide	142.06	wide	80	5	5	Positive
[$^{15}N_2$]-dC	230.1	wide	114.1	wide	80	5	5	Positive
dC	228.1	wide	112.1	wide	80	5	5	Positive
time segment 2 (4.1−9 min)								
[$^{15}N_2$]-fdC	258.09	enhanced	142.04	enhanced	40	2	5	Positive
fdC	256.09	enhanced	140.05	enhanced	40	2	5	Positive
[$^{15}N_2$]-fdU	257.06	wide	213.05	wide	40	6	5	Negative
fdU	255.06	wide	212.06	wide	40	6	5	Negative
[$^{15}N_2$]-dU	229.06	wide	185.06	wide	40	5	5	Negative
dU	227.07	wide	184.06	wide	40	5	5	Negative

Table S7. Compound-dependent LC-MS/MS-parameters used for the analysis of the deamination and C-C bond cleavage kinetics of the cadC nucleoside at 60 °C and 80 °C in buffered H_2O (pH 7.4). The fragmentor voltage was 380 V. Delta EMV was set to 300 and 500 (positive mode) in time segment 1 and 2, respectively, and 800 (negative mode). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage.

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1 (1.5−4.1 min)								
[$^{15}N_2$]-cadC	274.08	enhanced	158.03	enhanced	120	2	5	Positive
cadC	272.09	enhanced	156.04	enhanced	120	2	5	Positive
[$^{15}N_2$]-dC	230.1	wide	114.1	wide	120	5	5	Positive
dC	228.1	wide	112.1	wide	120	5	5	Positive
time segment 2 (4.1−9 min)								
cadU	273.07	wide	157.02	wide	80	5	5	Positive
cadU	271.06	wide	227.07	wide	80	7	5	Negative
[$^{15}N_2$]-dU	229.06	wide	185.06	wide	80	5	5	Negative
dU	227.07	wide	184.06	wide	80	5	5	Negative

LC-MS/MS calibration curves and representative chromatograms
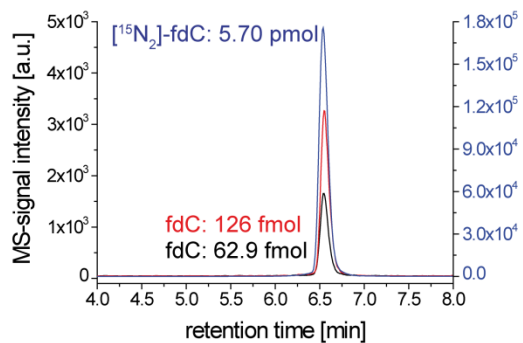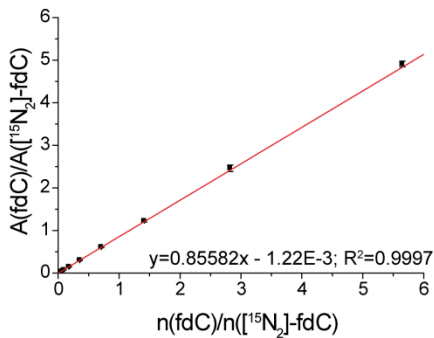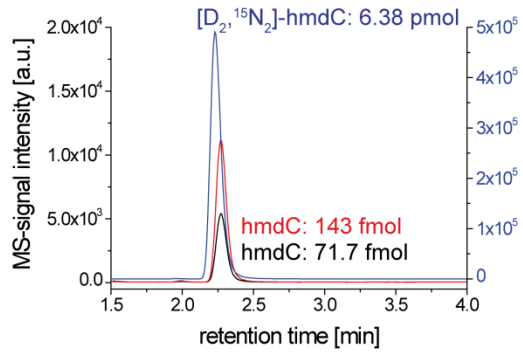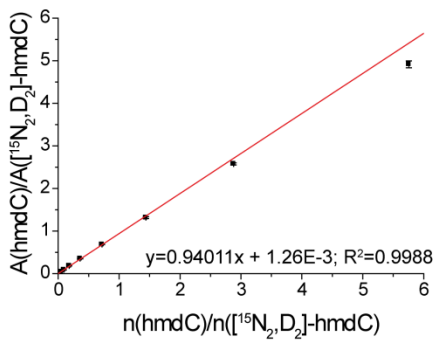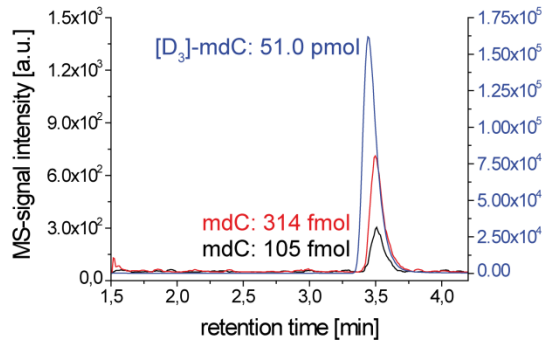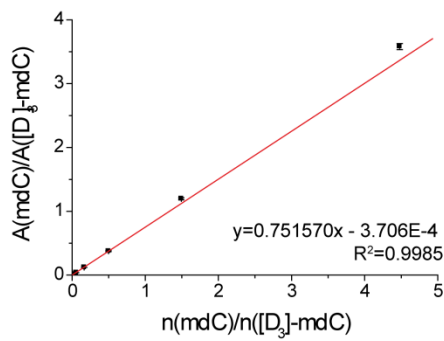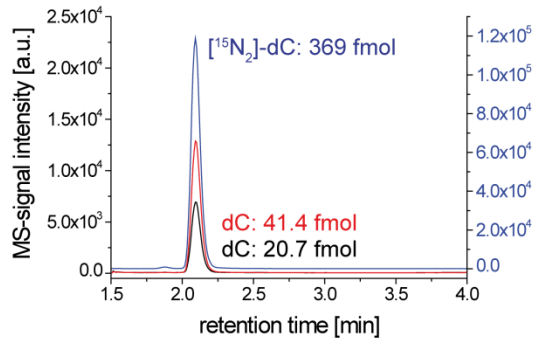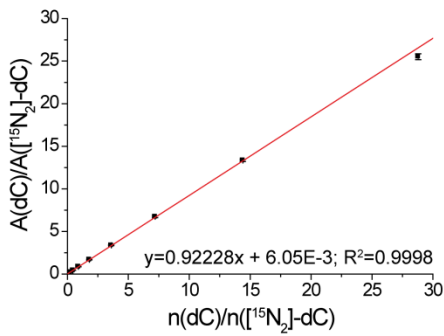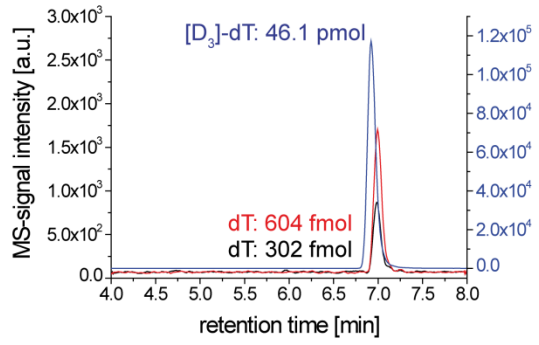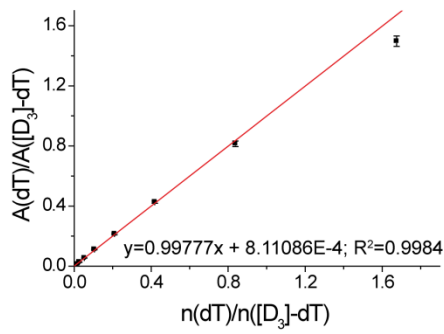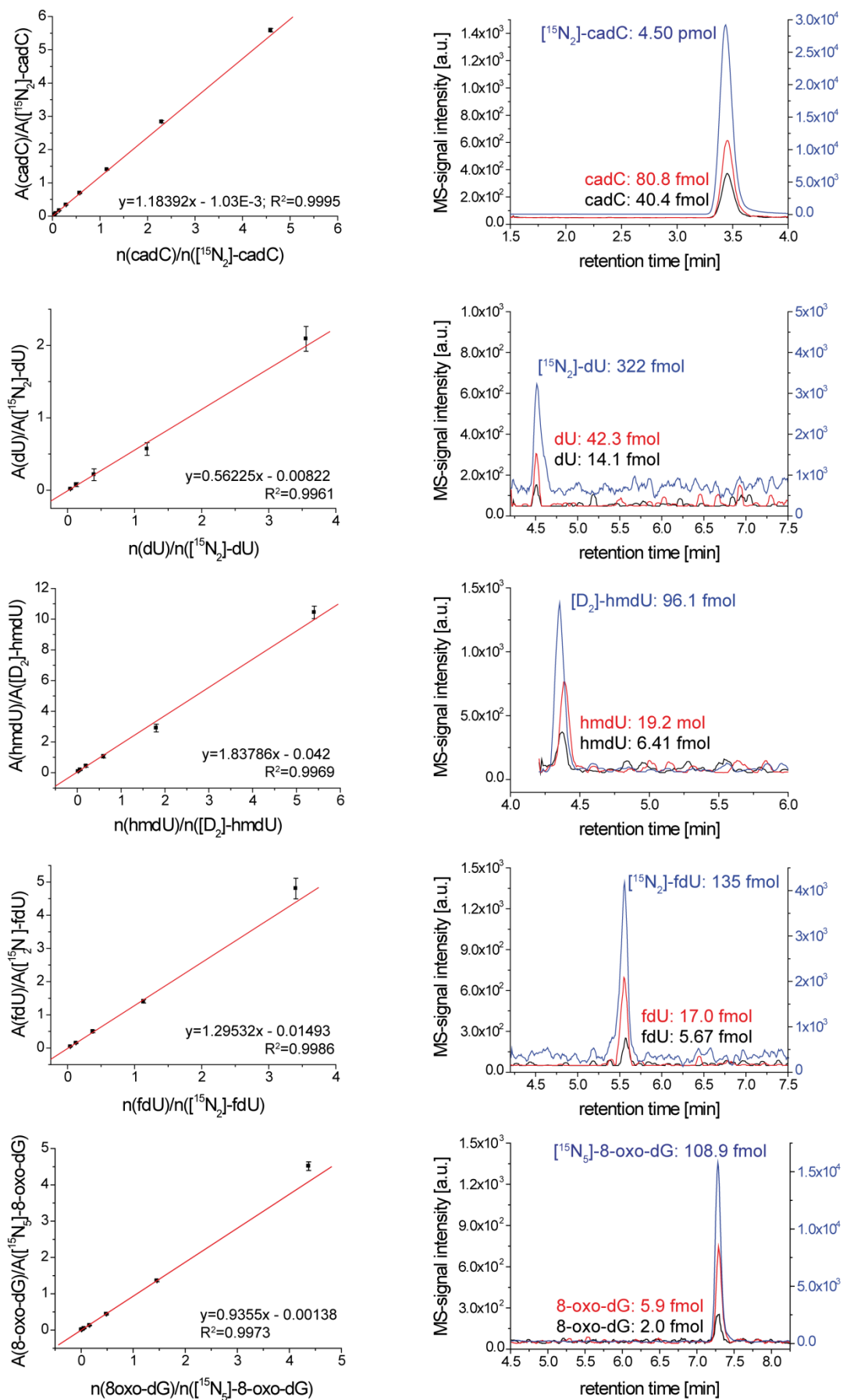
Figure S1. Calibration curves and representative chromatograms of dT/[D$_3$]-dT, dC/[^{15}N$_2$]-dC, mdC/[D$_3$]-mdC, hmdC/[^{15}N$_2$,D$_2$]-hmdC, fdC/[^{15}N$_2$]-fdC, cadC/[^{15}N$_2$]-cadC, dU/[^{15}N$_2$]-dU, hmdU/[D$_2$]-hmdU, fdU/[^{15}N$_2$]-fdU,

8-oxo-dG/[^{15}N$_5$]8-oxo-dG. These were obtained by applying the compound-dependent parameters summarized in Table S2. Depicted are the means of five replicates of one sample batch. Error bars reflect SD. Linearity was given across the following compound amounts in 10 µL injection volume: 302 fmol – 154 pmol dT; 20.7 fmol – 10.6 pmol dC; 105 fmol – 229 pmol mdC; 71.7 fmol – 36.7 pmol hmdC; 62.9 fmol – 32.2 pmol fdC; 40.4 fmol – 20.7 pmol cadC; 14.1 fmol – 1.15 pmol dU; 6.41 – 519 fmol hmdU; 5.67 – 460 fmol fdU; 2.0 – 475.7 fmol 8-oxo-dG. The amounts of the labeled internal standards in 10 µL injection volume were as follows: 46.1 pmol [D$_3$]-dT; 369 fmol [^{15}N$_2$]-dC; 51.0 pmol [D$_3$]-mdC; 6.38 pmol [^{15}N$_2$,D$_2$]-hmdC; 5.70 pmol [^{15}N$_2$]-fdC; 4.51 pmol [^{15}N$_2$]-cadC; 322 fmol [^{15}N$_2$]-dU; 96.1 fmol [D$_2$]-hmdU; 135 fmol [^{15}N$_2$]-fdU; 108.9 fmol [^{15}N$_5$]-8-oxo-dG.

Validation of the quantification method:

Validation for the established LC-MS/MS quantification method was based on three different series (i.e., calibration functions and quality control samples) accomplished on different days. Each calibration standard (5-8 standard concentrations) was technically analyzed five times. Each validation experiment was complemented by matrix blanks (analyzed in triplicates) to ensure selectivity and specificity of the method. Linear regression was applied by ORIGIN® 6.0 (*Microcal™*) to obtain calibration curves. Therefore, the ratio of the area under the curve (A/A*) of the unlabeled nucleoside to the labeled internal standard was plotted against the ratio of the amount of substance (n/n*) of the unlabeled nucleoside to the internal standard (see Figure S1). Calibration functions were calculated without weighting. Additionally, acceptable accuracy (80–120%) as well as precision (<20% RSD) was required. Accuracy was proven by computing the amount of substance n from the obtained A/A* ratios of the calibration standards using the respective calibration function. Here, accuracy was defined as the ratio of the used amount of substance to the calculated amount of substance in percent and had to be between 80–120% for each standard concentration. Precision was defined as follows: technical replicates of A/A* ratios for each calibration standard had to have relative standard deviations (RSD) smaller than 20%. The lower limit of quantification (LLOQ) was defined as the lowest concentration fulfilling the requirements of accuracy and precision and achieving a response of at least three times the response compared with the blank response. The Long-term stability of aqueous solutions of the labeled and unlabeled nucleosides at a storage temperature of −20 °C was investigated over two months including several freeze and thaw cycles by analyzing the MS/MS-responses with each batch.

Reaction of hmdC-, fdC- and cadC-containing DNA with β-mercaptoethanol

A solution containing 50 µM DNA (sequence 5′-AXG TAX GTT TTT AXG TAX GT-3′; hairpin 1: X = hmdC, hairpin 2: X = fdC and hairpin 3: X = cadC or 5′-XAG TXA GTT TTT AXT GAX TG-3′; hairpin 4: X = hmdC, hairpin 5 = fdC and hairpin 6 X = cadC) and 1 M NaCl was hybridized by heating to 95 °C for 5 min followed by gradual cooling to room temperature (1 °C/min). 1 µL of this solution was added to 9 µL of a 12 M β-mercaptoethanol, 1 M imidazole solution (pH 5.0) to obtain a 80% v/v thiol solution. In order to obtain different thiol concentrations the corresponding volume of

the 12 M β-mercaptoethanol, 1 M imidazole solution was diluted with water (pH 5.0) to a final volume of 9 μL prior to addition of the DNA solution. The solutions were incubated at 50 °C for 2 d in a *Mastercycler personal* (*Eppendorf*) with a heatable lid (95 °C). After evaporation to dryness (*Lyophille Alpha 2-4 LD plus*, *Christ*) the DNA was dissolved in 10 μL deionized H₂O. 5 μL of this solution were desalted by dialysis via a 0.025 μm filter membrane (*Merck Millipore*) for 1 h and again evaporated to dryness. The resulting DNA was then digested by the following procedure: The DNA was dissolved in an aqueous solution (32.5 μL) of 74 μM ZnSO₄, containing 5.2 units antarctic phosphatase, 84 units Nuclease S1 and specific amounts of labeled internal standards (see below). The mixture was incubated at 37 °C for 3 h in a *Thermomixer comfort* (*Eppendorf*). After addition of 7.5 μL of a 0.35 mM [Na]₂-EDTA solution containing 0.1 units snake venom phosphodiesterase I the sample was incubated for another 3 h at 37 °C. The sample was then kept at -20°C until the day of analysis. After filtration via a AcroPrep^TM Advance 96 Filter Plate (0.45 μm Supor®, *Pall Life Sciences*) the sample was analyzed by UHPLC-MS/MS according to the above mentioned procedure. The injection volume was 10 μL. The sample batch was cooled to 10 °C until injection.

For all tables in this section is valid: A is the area under the curve (AUC) of the respective peak.

* marks the labeled compound, which was added as an internal standard. The amount of substance n is calculated from the ratio of the AUC of the unlabeled nucleoside to the AUC of the spiked internal standard based on the respective calibration curve (see Figure S1). The quantification was carried out with the following amounts of internal standards: 184.0 pmol [D₃]-dT, 2.951 pmol [¹⁵N₂]-dC, 25.51 pmol [¹⁵N₂,D₂]-hmdC, 12.15 pmol [¹⁵N₂]-fdC, 16.13 pmol [¹⁵N₂]-cadC, 966.1 fmol [¹⁵N₂]-dU, 480.3 fmol [D₂]-hmdU, 567.4 fmol [¹⁵N₂]-fdU and 90.72 fmol [¹⁵N₂]-8-oxo-dG. The calculated amounts of the reaction products (Table S8–S22) were normalized to the calculated amount of dT (yield of nucleoside in %, see main figures).

Table S8. Quantification of the oxidation of hmdC to fdC and cadC in hairpin 1 in the presence of different concentrations of β-mercaptoethanol (β–ME) at pH 5.0 and 50 °C.

c(β-ME) [%]	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	yield of fdC [%]
80	0.9326	171.9	1.434	38.89	1.153	16.38	19.07
60	1.321	243.5	3.260	88.46	0.2106	3.008	2.470
40	1.455	268.1	3.394	92.09	0.0577	0.8371	0.6244
20	1.444	266.2	3.851	104.5	0.0273	0.4049	0.3043
15	1.706	314.5	4.435	120.3	0.0794	1.145	0.7280
10	2.145	395.4	5.305	144.0	0.1069	1.536	0.7767

Table S9. Detection of 5-((2″-hydroxyethyl)thio)methyl-dC as a 5-β-ME adduct of hmdC and quantification of the oxidation of dG to 8-oxo-dG in hairpin 1 in the presence of different concentrations of β-mercaptoethanol (β-ME) at pH 5.0 and 50 °C.

$c(\beta\text{-ME})$ [%]	$A_{5\text{-((2″-hydroxyethyl)thio)methyl-dC}}$	$A_{8\text{-oxo-dG}}/A_{8\text{-oxo-dG*}}$	$n_{8\text{-oxo-dG}}$ [pmol]	yield of 8-oxo-dG [%]
80	1.02×10^7	12.05	1.168	1.360
60	1.06×10^7	1.753	0.170	0.1397
40	5.31×10^6	0.8768	0.085	0.0635
20	1.12×10^6	1.076	0.1045	0.0785
15	6.89×10^5	1.258	0.1221	0.0777
10	4.40×10^5	1.910	0.1854	0.0938

Table S10. Quantification of the oxidation of fdC to cadC in hairpin 2 in the presence of different concentrations of β-mercaptoethanol (β−ME) at pH 5.0 and 50 °C.

$c(\beta\text{-ME})$ [%]	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{cadC}/A_{cadC*}	n_{cadC} [pmol]	yield of cadC [%]
80	0.2116	38.88	0.1006	1.385	7.123
60	0.7446	137.2	0.1018	1.401	2.042
40	0.6897	127.1	0.0473	0.6579	1.036
20	0.6370	117.3	0.0285	0.4028	0.6865
15	0.7293	134.4	0.0409	0.5716	0.8509
10	0.7240	133.4	0.0425	0.5926	0.8885

Table S11. Quantification of the C-C-bond cleavage reaction of hmdC to dC in hairpin 1 in the presence of different concentrations of β-mercaptoethanol (β−ME) at pH 5.0 and 50 °C. Yields of dC ≤ 0.15% are below the lower limit of quantification.

$c(\beta\text{-ME})$ [%]	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]
80	0.9326	171.9	0.1514	0.4652	0.5414
60	1.321	243.5	0.0787	0.2324	0.1909
40	1.454	268.1	0.0764	0.2250	0.1678
20	1.385	255.4	0.0861	0.2560	0.2005
15	1.706	314.5	0.0724	0.2123	0.1350
10	2.145	395.4	0.1000	0.3006	0.1520
5	1.850	341.0	0.0746	0.2192	0.1286
0	1.784	328.9	0.0678	0.1977	0.1202

Table S12. Quantification of the C-C-bond cleavage reaction of fdC to dC in hairpin 2 in the presence of different concentrations of β-mercaptoethanol (β−ME) at pH 5.0 and 50 °C. Yields of dC ≤ 0.36% are below the lower limit of quantification.

c(β-ME) [%]	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]
80	0.2116	38.88	0.1753	0.5415	2.786
60	0.7446	137.2	0.2268	0.7065	1.030
40	0.6897	127.1	0.1293	0.3944	0.6208
20	0.6370	117.3	0.0727	0.2132	0.3633
15	0.7293	134.4	0.0776	0.2291	0.3410
10	0.7240	133.4	0.0646	0.1872	0.2807
5	0.6793	125.1	0.1170	0.3549	0.5671
0	0.7561	139.3	0.0694	0.2028	0.2911

Table S13. Quantification of the C-C-bond cleavage reaction of cadC to dC in hairpin 3 in the presence of different concentrations of β-mercaptoethanol (β−ME) at pH 5.0 and 50 °C.

c(β-ME) [%]	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]
80	0.7810	143.9	6.355	20.32	28.24
60	1.250	230.3	5.349	17.10	14.85
40	1.276	235.1	2.999	9.576	8.145
20	1.414	260.7	0.9074	2.884	2.213
15	1.756	323.8	0.9878	3.142	1.941
10	1.851	341.2	0.4745	1.499	0.8786
5	1.706	314.4	0.2234	0.6953	0.4423
0	1.572	289.8	0.0809	0.2395	0.1653

Table S14. Quantification of the C-C-bond cleavage reaction of cadC to dC in hairpin 3 in the presence of different concentrations of β-mercaptoethanol (β−ME) at pH 5.0 and 37 °C.

c(β-ME) [%]	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]
80	1.623	299.3	3.225	10.30	6.883
60	1.397	257.5	0.9114	2.897	2.250
40	1.414	260.7	0.4006	1.262	0.9684
20	1.429	263.4	0.1554	0.4779	0.3629
10	1.902	350.6	0.2083	0.6473	0.3693

Table S15. Quantification of the C-C-bond cleavage reaction of cadC to dC in hairpin 3 in the presence of different concentrations of β-mercaptoethanol (β−ME) at pH 7.4 and 50 °C. Yields of dC ≤ 0.16% are below the lower limit of quantification.

c(β-ME) [%]	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]
80	1.254	231.1	0.7616	2.418	2.092
60	1.198	220.8	0.7717	2.450	2.220
40	0.8226	151.6	0.7376	2.341	3.089
20	1.588	292.8	0.2513	0.7849	0.5361
10	1.454	268.0	0.0711	0.2083	0.1554

Table S16. Quantification of the C-C-bond cleavage reaction of cadC to dC in hairpin 3 in the presence of different concentrations of β-mercaptoethanol (β−ME) at pH 7.4 and 37 °C. Yields of dC below 0.16% are below the lower limit of quantification.

c(β-ME) [%]	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]
80	1.802	332.2	0.4435	1.400	0.8428
60	1.541	284.1	0.2580	0.8061	0.5675
40	2.005	369.6	0.1347	0.4117	0.2228
20	1.672	308.2	0.0748	0.2199	0.1427
10	1.776	327.4	0.4019	1.267	0.7738

Table S17. Quantification of the C-C bond cleavage of hmdC to dC and of the deamination of hmdC to hmdU in hairpin 1 and hairpin 4 in the presence of 80 % β-mercaptoethanol at pH 5.0 and 50 °C. Every measurement was confirmed by two independent replicates.

	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]	A_{hmdU}/A_{hmdU*}	n_{hmdU} [pmol]	yield of hmdU [%]
hairpin 1	0.9084	167.4	0.1473	0.4520	0.5400	0.9360	0.2320	0.2771
hairpin 1	1.182	217.9	0.1903	0.5894	0.5410	1.397	0.3529	0.3239
hairpin 1	0.9326	171.9	0.1514	0.4652	0.5414	1.034	0.2577	0.2999
hairpin 4	0.2299	42.26	0.1131	0.3425	1.621	0.4049	0.0927	0.4386
hairpin 4	0.7571	139.5	0.1868	0.5785	0.8294	1.080	0.2697	0.3866
hairpin 4	0.8385	154.5	0.1685	0.5197	0.6727	1.095	0.2736	0.3542

Table S18: Mean value (MV) and standard deviation (SD) of the quantification of the C-C bond cleavage of hmdC to dC and of the deamination of hmdC to hmdU in hairpin 1 and hairpin 4 in the presence of 80 % β-mercaptoethanol at pH 5.0 and 50 °C.

	MV yield of dC [%]	SD yield of dC [%]	MV yield of hmdU [%]	SD yield of hmdU [%]
hairpin 1	0.5408	0.0007	0.3003	0.0234
hairpin 4	1.041	0.5082	0.3931	0.0426

Table S19. Quantification of the C-C bond cleavage of fdC to dC and of the deamination of fdC to fdU in hairpin 2 and hairpin 5 in the presence of 80 % β-mercaptoethanol at pH 5.0 and 50 °C. Every measurement was confirmed by one or two independent replicates.

	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]	A_{fdU}/A_{fdU*}	n_{fdU} [pmol]	yield of fdU [%]
hairpin 2	0.5927	109.2	0.1296	0.3974	2.294	1.252	0.0633	0.1160
hairpin 2	0.2116	38.88	0.0572	0.1753	2.786	0.5415	0.0316	0.1626
hairpin 5	0.2898	53.30	0.2745	0.8589	3.223	0.2144	0.1004	0.3769
hairpin 5	0.4387	80.76	0.3570	1.123	2.781	0.2369	0.1103	0.2732
hairpin 5	0.2880	52.96	0.3890	1.225	4.627	0.2253	0.1052	0.3973

Table S20: Mean value (MV) and standard deviation (SD) of the quantification of the C-C bond cleavage of fdC to dC and of the deamination of fdC to fdU in hairpin 2 and hairpin 5 in the presence of 80 % β-mercaptoethanol at pH 5.0 and 50 °C.

	MV yield of dC [%]	SD yield of dC [%]	MV yield of fdU [%]	SD yield of fdU [%]
hairpin 2	2.540	0.3478	0.1393	0.0329
hairpin 5	3.544	0.9641	0.3491	0.0666

Table S21. Quantification of the C-C bond cleavage of cadC to in hairpin 3 and hairpin 6 in the presence of 80% β-mercaptoethanol at pH 5.0 and 50 °C. Every measurement was confirmed by two independent replicates. The deamination product cadU was not detected.

	A_{dT}/A_{dT*}	n_{dT} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]	yield of dC [%]
hairpin 3	0.5671	104.4	4.989	15.94	30.53
hairpin 3	0.6504	119.8	5.664	18.10	30.22
hairpin 3	0.7810	143.9	6.355	20.32	28.24
hairpin 6	0.3797	69.89	3.400	10.86	31.08

hairpin 6	0.5791	106.7	5.856	18.72	35.10
hairpin 6	0.5183	95.44	5.189	16.58	34.75

Table S22: Mean value (MV) and standard deviation (SD) of the quantification of the C-C bond cleavage of cadC to dC in hairpin 3 and hairpin 6 in the presence of 80 % β-mercaptoethanol at pH 5.0 and 50 °C.

	MV yield of dC [%]	SD yield of dC [%]
hairpin 3	29.66	1.246
hairpin 6	33.64	2.227

Oxidation, deamination and C-C bond cleavage kinetics of dC, mdC, hmdC, fdC and cadC nucleosides and determination of rate constants

10 μM aqueous nucleoside solutions (dC, mdC, hmdC and fdC) in 18 mM sodium phosphate buffer (pH 7.4) were incubated at the given temperatures in a *Mastercycler personal* (*Eppendorf*) with a heatable lid (95 °C). At given time points the reaction mixtures were immediately cooled to room temperature and samples of 10 μL (100 pmol) were collected. An aqueous solution of internal standards (15 μL) were added to each sample: Reactions of dC were quantified with 22.13 pmol $[^{15}N_2]$-dC and 966.1 fmol $[^{15}N_2]$-dU; reactions of mdC were quantified with 25.51 pmol $[D_3]$-mdC, 510.3 fmol $[^{15}N_2,D_2]$-hmdC, 607.6 fmol $[^{15}N_2]$-fdC and 736.1 fmol $[D_3]$-dT; reactions of hmdC were quantified with 737.8 fmol $[^{15}N_2]$-dC, 25.52 pmol $[^{15}N_2,D_2]$-hmdC, 607.6 fmol $[^{15}N_2]$-fdC, 430.1 fmol $[^{15}N_2]$-cadC and 480.3 fmol $[D_2]$-hmdU; reactions of fdC were quantified with 737.8 fmol $[^{15}N_2]$-dC, 30.38 pmol $[^{15}N_2]$-fdC, 430.1 fmol $[^{15}N_2]$-cadC and 567.4 fmol $[^{15}N_2]$-fdU; reactions of cadC were quantified with 737.8 fmol $[^{15}N_2]$-dC and 21.51 pmol $[^{15}N_2]$-cadC.

Samples were kept at -20°C until the day of analysis. After filtration via a AcroPrep™ Advance 96 Filter Plate (0.45 μm Supor®, *Pall Life Sciences*), LC-ESI-MS/MS analysis was performed using the above described methods. The injection volume was 10 μL. The sample batch was cooled to 10 °C until injection.

For all tables in this section is valid: A is the area under the curve (AUC) of the respective peak. * marks the internal standard. The amount of substance n (Table S23 – S48) is calculated from the ratio of the AUC of the unlabeled nucleoside to the AUC of the spiked internal standard based on the respective calibration curve (see Figure S1). The amount of each quantified reaction product was normalized to the amount of the quantified starting material in order to take changes in the concentration or in the sample volume due to evaporation into account. For each time point three independently prepared samples were analyzed. The mean values and standard deviations of each triplica were calculated and, after subtraction of the normalized amount of the respective nucleoside at t=0 h (Table S49 – S58, S66 – S81 and S87 – S89), were plotted against time (see Figure 1, 3, S2, S6

253

and S9). After calculation of the natural logarithm (Table S49 – S58, S66 – S81 and S87 – S89) and plotting it against the reaction time (Figures S3, S4, S7 and S10) the reactions were found to be pseudo first order and the rate constants were obtained from the slope of the straight according to the equation:

$$k = t^{-1} \cdot \ln [(n_0 - n_\infty) / (n_t - n_\infty)]$$

In order to determine the rate constants at 37 °C the rate constants at elevated temperatures were linearly extrapolated (ln(k) versus T^{-1}, Table S64 and S85, Figure S5 and S8) to 37.0 °C using the Arrhenius equation:

$$\ln(k) = -E_A/R \times T^{-1} + \ln(A)$$

with k being the rate constant, R being the gas constant and A being the preexponential factor. The activation energy E_A was calculated from the slope of this straight.

Table S23. Quantification of the deamination of dC to dU at 60.0 °C.

time [h]	A_{dC}/A_{dC*}	n_{dC} [pmol]	A_{dU}/A_{dU*}	n_{dU} [pmol]
0	4.281	102.6	0.0417	0.0858
0	4.047	96.99	0.0283	0.0628
0	4.429	106.1	0.0291	0.0642
11.5	4.368	104.7	0.0847	0.1596
11.5	4.184	100.3	0.0597	0.1167
11.5	4.673	112.0	0.1039	0.1927
24	4.492	107.7	0.1174	0.2159
24	4.204	100.7	0.1137	0.2095
24	4.664	111.8	0.1627	0.2936
36	4.664	111.8	0.1784	0.3207
36	4.450	106.7	0.1773	0.3187
36	4.825	115.7	0.1978	0.3539
48	4.950	118.6	0.2532	0.4491
48	4.586	109.9	0.1978	0.3539
48	5.151	123.5	0.2511	0.4457
60	5.451	130.7	0.3117	0.5496
60	5.199	124.6	0.3059	0.5398
60	5.662	135.7	0.3122	0.5505

Table S24. Quantification of the deamination of dC to dU at 67.5 °C.

time [h]	A_{dC}/A_{dC*}	n_{dC} [pmol]	A_{dU}/A_{dU*}	n_{dU} [pmol]
0	4.306	103.2	0.0296	0.0650
0	4.445	106.5	0.0147	0.0393
0	4.521	108.4	0.0300	0.0657
11.5	4.586	109.9	0.1120	0.2066
11.5	4.710	112.9	0.1266	0.2317
11.5	4.751	113.9	0.1544	0.2794
24	4.741	113.6	0.2425	0.4308
24	4.887	117.1	0.2742	0.4852
24	4.872	116.8	0.2401	0.4268
36	5.052	121.1	0.3507	0.6167
36	5.089	122.0	0.4518	0.7905
36	5.256	126.0	0.3549	0.6239
48	5.611	134.5	0.5985	1.0425
48	5.368	128.7	0.5465	0.9531
48	5.917	141.9	0.5756	1.003
60	5.881	141.0	0.7532	1.308
60	6.143	147.3	0.6923	1.204
60	6.028	144.5	0.7599	1.320

Table S25. Quantification of the deamination of dC to dU at 75.0 °C.

time [h]	A_{dC}/A_{dC*}	n_{dC} [pmol]	A_{dU}/A_{dU*}	n_{dU} [pmol]
0	4.451	106.7	0.0708	0.1357
0	4.249	101.8	0.0154	0.0406
0	4.391	105.2	0.0226	0.0530
11.5	4.744	113.7	0.3212	0.5661
11.5	4.546	109.0	0.2904	0.5130
11.5	4.618	110.7	0.2570	0.4557
24	4.984	119.5	0.6145	1.070
24	4.913	117.8	0.5097	0.8900
24	5.080	121.8	0.5970	1.040
36	5.744	137.7	1.037	1.796
36	5.373	128.8	0.8923	1.547
36	5.624	134.8	1.0034	1.738
48	7.235	173.5	1.733	2.991

48	6.778	162.5	1.566	2.704
48	6.951	166.7	1.492	2.578
60	3.557	85.21	0.9872	1.710
60	3.476	83.28	0.9838	1.705
60	3.311	79.32	0.9946	1.723

Table S26. Quantification of the deamination of dC to dU at 82.5 °C.

time [h]	A_{dC}/A_{dC*}	n_{dC} [pmol]	A_{dU}/A_{dU*}	n_{dU} [pmol]
0	4.376	104.9	0.0311	0.0676
0	4.513	108.2	0.0279	0.0621
0	4.636	111.1	0.0347	0.0737
11.5	4.830	115.8	0.5649	0.9847
11.5	5.009	120.1	0.6051	1.054
11.5	4.972	119.2	0.6007	1.046
24	5.395	129.3	1.460	2.523
24	5.587	133.9	1.453	2.511
24	5.608	134.4	1.518	2.622
36	6.481	155.4	2.520	4.345
36	5.613	134.6	2.041	3.520
36	6.981	167.4	2.323	4.006
48	8.314	199.4	4.186	7.207
48	9.979	239.4	5.088	8.757
48	8.786	210.7	4.640	7.987

Table S27. Quantification of the deamination of mdC to dT at 60.0 °C.

time [h]	A_{mdC}/A_{mdC*}	n_{mdC} [pmol]	A_{dT}/A_{dT*}	n_{dT} [pmol]
0	4.789	162.7	0.3248	0.2390
0	4.402	149.6	0.3051	0.2245
0	4.684	159.1	0.2991	0.2200
11.5	4.674	158.8	0.4413	0.3250
11.5	4.654	158.1	0.4499	0.3313
11.5	4.616	156.8	0.4749	0.3498
24	4.931	167.5	0.5881	0.4333
24	4.750	161.4	0.5739	0.4228
24	4.918	167.1	0.5872	0.4327
36	5.326	180.9	0.7157	0.5274

36	4.860	165.1	0.6883	0.5072
36	5.058	171.8	0.6871	0.5063
48	5.360	182.1	0.8471	0.6243
48	5.101	173.3	0.8262	0.6089
48	5.387	183.0	0.8430	0.6213
60	6.041	205.2	1.141	0.8409
60	5.633	191.4	1.134	0.8359
60	5.829	198.0	1.085	0.7996

Table S28. Quantification of the oxidation of mdC to hmdC at 60.0 °C.

time [h]	A_{mdC}/A_{mdC^*}	n_{mdC} [pmol]	A_{hmdC}/A_{hmdC^*}	n_{hmdC} [pmol]
0	4.573	155.3	0.0112	0.0054
0	4.454	151.2	0.0109	0.0052
0	4.702	159.7	0.0115	0.0056
12	4.825	163.8	0.0121	0.0059
12	4.595	156.0	0.0127	0.0062
12	4.709	159.9	0.0117	0.0057
23.5	5.031	170.8	0.0134	0.0066
23.5	4.853	164.8	0.0127	0.0062
23.5	4.785	162.5	0.0125	0.0061
35	5.207	176.8	0.0143	0.0071
35	5.025	170.6	0.0146	0.0072
35	4.768	161.9	0.0134	0.0066
46.5	5.336	181.2	0.0159	0.0079
46.5	5.278	179.2	0.0153	0.0076
46.5	5.304	180.1	0.0152	0.0076

Table S29. Quantification of the deamination of mdC to dT at 67.5 °C.

time [h]	A_{mdC}/A_{mdC^*}	n_{mdC} [pmol]	A_{dT}/A_{dT^*}	n_{dT} [pmol]
0	4.528	153.8	0.2918	0.2147
0	4.507	153.1	0.3138	0.2309
0	4.641	157.7	0.3260	0.2399
11.5	4.638	157.6	0.6478	0.4773
11.5	4.688	159.3	0.5917	0.4359
11.5	4.886	166.0	0.5983	0.4408
24	4.953	168.3	0.8877	0.6543

24	5.045	171.4	0.8468	0.6242
24	5.137	174.5	0.9067	0.6683
36	5.329	181.0	1.122	0.8275
36	5.404	183.6	1.127	0.8306
36	5.387	183.0	1.152	0.8491
48	5.623	191.0	1.566	1.155
48	5.968	202.7	1.674	1.234
48	5.865	199.2	1.605	1.184
60	6.214	211.1	2.130	1.571
60	5.923	201.2	2.182	1.610
60	6.306	214.2	1.971	1.453

Table S30. Quantification of the deamination of mdC to dT at 75.0 °C.

time [h]	A_{mdC}/A_{mdC*}	n_{mdC} [pmol]	A_{dT}/A_{dT*}	n_{dT} [pmol]
0	4.669	158.6	0.3251	0.2393
0	4.526	153.8	0.3263	0.2402
0	4.709	160.0	0.2793	0.2055
11.5	4.999	169.8	1.020	0.7515
11.5	4.910	166.8	1.018	0.7506
11.5	5.123	174.1	1.025	0.7555
24	5.187	176.2	1.479	1.090
24	5.270	179.0	1.789	1.319
24	5.510	187.2	1.567	1.156
36	5.772	196.1	2.424	1.788
36	5.989	203.5	2.653	1.957
36	5.998	203.7	2.485	1.832
48	7.128	242.1	3.769	2.780
48	7.390	251.0	4.107	3.030
48	7.134	242.3	4.127	3.044
60	4.123	140.1	2.993	2.207
60	3.489	118.6	2.495	1.840
60	4.081	138.7	3.071	2.265

Table S31. Quantification of the oxidation of mdC to hmdC at 80.0 °C.

time [h]	A_{mdC}/A_{mdC*}	n_{mdC} [pmol]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]
0	1.969	66.86	0.0061	0.0026
0	2.054	69.76	0.0066	0.0029

258

0	2.304	78.22	0.0069	0.0031
12	4.641	157.6	0.0114	0.0055
12	4.474	151.9	0.0111	0.0053
12	4.826	163.9	0.0115	0.0055
23.5	5.681	192.9	0.0164	0.0082
23.5	4.968	168.7	0.0148	0.0074
23.5	5.006	170.0	0.0156	0.0078
35	5.605	190.3	0.0179	0.0090
35	5.493	186.5	0.0191	0.0097
35	5.690	193.2	0.0198	0.0101
46.5	6.366	216.1	0.0232	0.0119
46.5	6.615	224.6	0.0242	0.0124
46.5	6.301	213.9	0.0256	0.0132

Table S32. Quantification of the deamination of mdC to dT at 82.5 °C.

time [h]	A_{mdC}/A_{mdC*}	n_{mdC} [pmol]	A_{dT}/A_{dT*}	n_{dT} [pmol]
0	4.593	156.1	0.3266	0.2403
0	4.611	156.7	0.3152	0.2320
0	4.523	153.7	0.3207	0.2360
11.5	5.042	171.3	1.767	1.303
11.5	5.002	169.9	2.025	1.493
11.5	5.149	174.9	1.812	1.336
24	5.720	194.3	3.306	2.438
24	5.704	193.8	3.552	2.620
24	5.685	193.1	3.376	2.490
36	6.885	233.9	5.723	4.221
36	7.077	240.4	6.291	4.641
36	6.712	228.0	6.555	4.835
48	9.074	308.2	10.34	7.628
48	9.279	315.1	10.22	7.538
48	8.826	299.7	10.09	7.447

Table S33. Quantification of the oxidation of hmdC to fdC and the deamination of hmdC at 60.0 °C.

time [h]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{hmdU}/A_{hmdU*}	n_{hmdU} [pmol]
0	3.266	88.63	0.1240	0.0889	1.010	0.2514
0	3.295	89.40	0.1291	0.0925	1.031	0.2569

time [h]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{hmdU}/A_{hmdU*}	n_{hmdU} [pmol]
0	3.489	94.66	0.1307	0.0936	1.024	0.2551
11.5	3.369	91.41	0.2050	0.1464	1.114	0.2787
11.5	3.392	92.02	0.1916	0.1369	1.116	0.2791
11.5	3.335	90.48	0.2147	0.1533	1.164	0.2918
24	3.547	96.25	0.2857	0.2037	1.357	0.3423
24	3.665	99.44	0.2946	0.2100	1.316	0.3315
24	3558	96.53	0.3094	0.2205	1.582	0.4014
36	3.569	96.85	0.3702	0.2637	1.553	0.3939
36	3.588	97.35	0.3531	0.2515	1.592	0.4040
36	3.717	100.9	0.4045	0.2880	1.678	0.4266
48	3.772	102.4	0.4562	0.3248	1.854	1.854
48	3.877	105.2	0.4468	0.3181	1.740	1.740
48	3.861	104.8	0.5039	0.3586	1.820	1.820
60	4.112	111.6	0.6031	0.4291	2.058	0.5263
60	4.209	114.2	0.5560	0.3956	2.140	0.5476
60	4.228	114.7	0.6683	0.4753	2.016	0.5152

Table S34. Quantification of the oxidation of hmdC to fdC and the deamination of hmdC to hmdU at 67.5 °C.

time [h]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{hmdU}/A_{hmdU*}	n_{hmdU} [pmol]
0	3.255	88.31	0.1219	0.0874	0.8955	0.2213
0	3.417	92.71	0.1318	0.0945	1.015	0.2527
0	3.283	89.09	0.1469	0.1052	0.9535	0.2365
11.5	3.393	92.07	0.2824	0.2013	1.435	0.3629
11.5	3.415	92.66	0.3013	0.2148	1.419	0.3585
11.5	3.392	92.03	0.3559	0.2535	1.215	0.3051
24	3.612	98.01	0.4188	0.2982	1.673	0.4251
24	3.478	94.36	0.4362	0.3106	1.677	0.4262
24	3.496	94.87	0.5355	0.3811	1.744	0.4439
36	3.797	103.0	0.5298	0.3770	2.197	0.5628
36	3.8468	104.4	0.6011	0.4276	2.279	0.5841
36	3.7689	102.3	0.7424	0.5279	2.211	0.5663
48	4.150	112.6	0.7288	0.5182	2.655	0.6828
48	4.265	115.7	0.8037	0.5715	2.881	0.7420
48	4.199	114.0	1.007	0.7155	2.687	0.6911
60	4.712	127.9	0.9892	0.7032	3.708	0.9588
60	4.553	123.6	1.033	0.7342	3.545	0.9162
60	4.538	123.1	1.285	0.9129	3.523	0.9103

Table S35. Quantification of the oxidation of hmdC to fdC and the deamination of hmdC to hmdU at 75.0 °C.

time [h]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{hmdU}/A_{hmdU*}	n_{hmdU} [pmol]
0	3.290	89.28	0.1287	0.0923	0.9406	0.2332
0	3.295	89.39	0.1303	0.0934	0.8981	0.2220
0	3.332	90.41	0.1464	0.1048	0.8958	0.2214
11.5	3.443	93.42	0.2932	0.2090	1.821	0.4640
11.5	3.649	99.01	0.3163	0.2254	1.922	0.4904
11.5	3.577	97.04	0.4491	0.3197	1.816	0.4628
24	3.830	103.9	0.4496	0.3200	2.765	0.7115
24	3.852	104.5	0.5421	0.3857	2.864	0.7375
24	3.898	105.8	0.7140	0.5078	2.839	0.7311
36	4.227	114.7	0.6484	0.4612	4.128	1.069
36	5.024	136.3	0.9085	0.6458	9.334	2.434
36	4.285	116.3	1.012	0.7195	4.136	1.071
48	4.971	134.9	0.9144	0.6500	5.874	1.527
48	5.205	141.3	1.253	0.8902	6.510	1.694
48	5.276	143.2	1.456	1.0347	5.923	1.540
60	3.044	82.59	0.6076	0.4322	3.936	1.019
60	2.900	78.67	0.7671	0.5454	3.761	0.9727
60	2.840	77.05	0.9384	0.6671	3.817	0.9876

Table S36. Quantification of the oxidation of hmdC to fdC at 80.0 °C.

time [h]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]
0	3.203	86.90	0.1261	0.0904
0	3.149	85.45	0.1350	0.0967
0	3.257	88.36	0.1357	0.0972
12	3.426	92.95	0.3389	0.2415
12	3.336	90.51	0.3623	0.2581
12	3.620	98.23	0.3819	0.2720
23.5	3.913	106.2	0.5395	0.3839
23.5	3.800	103.1	0.5508	0.3919
23.5	3.824	103.8	0.6629	0.4715
35	4.492	121.9	0.7604	0.5407
35	4.406	119.6	0.8502	0.6044
35	4.374	118.7	0.9515	0.6764
46.5	6.297	170.9	1.333	0.9470
46.5	6.216	168.7	1.565	1.112

| **46.5** | 5.955 | 161.6 | 1.556 | 1.105 |

Table S37. Quantification of the deamination of hmdC to hmdU at 82.5 °C.

time [h]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{hmdU}/A_{hmdU*}	n_{hmdU} [pmol]
0	3.449	93.57	1.074	0.2682
0	3.339	90.61	1.017	0.2531
0	3.311	89.84	0.9627	0.2390
11.5	3.688	100.1	2.792	0.7186
11.5	3.617	98.15	2.763	0.7110
11.5	3.677	99.76	2.727	0.7016
24	4.251	115.4	5.439	1.413
24	4.095	111.1	5.342	1.387
24	4.202	114.0	5.295	1.375
36	4.350	118.0	4.100	1.062
36	4.981	135.2	8.985	2.343
36	4.990	135.4	8.442	2.200
48	6.299	170.9	13.47	3.520
48	6.444	174.8	14.36	3.752
48	6.427	174.4	14.10	3.685

Table S38. Quantification of the oxidation of fdC to cadC and the deamination of fdC to fdU at 60.0 °C.

time [h]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{cadC}/A_{cadC*}	n_{cadC} [pmol]	A_{fdU}/A_{fdU*}	n_{fdU} [pmol]
0	2.817	100.03	0.0514	0.0191	0.0136	0.0714
0	2.895	102.8	0.0451	0.0168	0.0158	0.0723
0	2.794	99.22	0.0541	0.0200	0.0106	0.0700
11.5	2.851	101.2	0.0977	0.0359	0.2237	0.1634
11.5	2.898	102.9	0.0902	0.0332	0.2222	0.1627
11.5	2.838	100.8	0.0926	0.0340	0.2255	0.1642
24	3.012	107.0	0.1195	0.0438	0.4198	0.2493
24	2.947	104.6	0.1181	0.0433	0.4284	0.2531
24	2.883	102.4	0.1130	0.0414	0.3994	0.2403
36	3.023	107.3	0.1452	0.0531	0.5911	0.3243
36	3.073	109.1	0.1327	0.0486	0.5841	0.3213
36	3.032	107.7	0.1196	0.0438	0.5868	0.3225
48	3.233	114.8	0.1596	0.0584	0.8131	0.4216
48	3.195	113.5	0.1601	0.0585	0.7809	0.4075

48	3.080	109.4	0.1395	0.0511	0.7732	0.4041
60	3.528	125.3	0.1768	0.0646	1.027	0.5153
60	3.455	122.7	0.1800	0.0658	0.9481	0.4808
60	3.262	115.8	0.1594	0.0583	0.9564	0.4844

Table S39. Quantification of the oxidation of fdC to cadC and the deamination of fdC to fdU at 67.5 °C.

time [h]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{cadC}/A_{cadC*}	n_{cadC} [pmol]	A_{fdU}/A_{fdU*}	n_{fdU} [pmol]
0	2.806	99.66	0.0415	0.0154	0.0172	0.0729
0	2.789	99.03	0.0387	0.0144	0.0148	0.0719
0	2.807	99.67	0.0360	0.0134	0.0150	0.0720
11.5	2.845	101.0	0.0928	0.0341	0.4943	0.2819
11.5	2.866	101.8	0.0949	0.0348	0.4879	0.2791
11.5	2.829	100.5	0.0896	0.0329	0.5310	0.2980
24	2.919	103.7	0.1140	0.0418	0.8733	0.4480
24	2.889	102.6	0.2522	0.0920	0.8269	0.4277
24	2.977	105.7	0.1147	0.0421	0.8492	0.4374
36	3.056	108.5	0.1336	0.0489	1.187	0.5852
36	3.027	107.5	0.1312	0.0480	1.178	0.5813
36	3.052	108.4	0.1386	0.0507	1.137	0.5636
48	3.358	119.2	0.1542	0.0564	1.597	0.7648
48	3.225	114.5	0.1528	0.0559	1.578	0.7565
48	3.262	115.8	0.1557	0.0569	1.578	0.7567
60	3.295	117.0	0.1654	0.0605	1.992	0.9380
60	3.488	123.9	0.1811	0.0662	1.929	0.9103
60	3.379	112.0	0.1803	0.0659	1.956	0.9221

Table S40. Quantification of the oxidation of fdC to cadC and the deamination of fdC to fdU at 75.0 °C.

time [h]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{cadC}/A_{cadC*}	n_{cadC} [pmol]	A_{fdU}/A_{fdU*}	n_{fdU} [pmol]
0	2.873	102.0	0.0397	0.0148	0.0114	0.0704
0	2.755	97.85	0.0392	0.0146	0.0281	0.0777
0	2.790	99.06	0.0461	0.0171	0.0150	0.0720
11.5	2.842	100.9	0.1393	0.0510	1.072	0.5349
11.5	2.819	100.1	0.0971	0.0356	1.027	0.5152
11.5	2.811	99.83	0.1292	0.0473	1.031	0.5170
24	2.858	101.5	0.1701	0.0622	1.677	0.7998
24	2.869	101.9	0.1294	0.0474	1.792	0.8506

24	2.851	101.3	0.1583	0.0579	1.775	0.8429
36	3.055	108.5	0.2133	0.0778	2.497	1.159
36	3.022	107.3	0.1592	0.0582	2.380	1.108
36	3.040	108.0	0.1982	0.0724	2.564	1.189
48	3.365	119.5	0.2601	0.0949	3.486	1.593
48	3.439	122.1	0.2093	0.0764	3.435	1.570
48	3.423	121.5	0.2523	0.0920	3.464	1.583
60	1.656	58.83	0.1643	0.0601	2.060	0.9678
60	1.718	61.02	0.1226	0.0449	2.338	1.090
60	1.758	62.44	0.1465	0.0536	2.318	1.081

Table S41. Quantification of the oxidation of fdC to cadC at 80.0 °C.

time [h]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{cadC}/A_{cadC*}	n_{cadC} [pmol]
0	2.743	97.42	0.0348	0.0130
0	2.845	101.0	0.0442	0.0164
0	3.013	107.0	0.0326	0.0122
12	3.523	125.1	0.1218	0.0446
12	3.185	113.1	0.1376	0.0504
12	3.125	111.0	0.1049	0.0385
23.5	3.514	124.8	0.1438	0.0526
23.5	3.458	122.8	0.1705	0.0623
23.5	3.467	123.1	0.1329	0.0487
35	4.001	142.1	0.2009	0.0734
35	3.986	141.5	0.2301	0.0840
35	3.874	137.5	0.1791	0.0654
46.5	5.416	192.3	0.3042	0.1109
46.5	4.901	174.0	0.3206	0.1168
46.5	5.084	180.5	0.2666	0.0972

Table S42. Quantification of the deamination of fdC to fdU at 82.5 °C.

time [h]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{fdU}/A_{fdU*}	n_{fdU} [pmol]
0	2.836	100.7	0.0141	0.0716
0	2.853	101.3	0.0147	0.0718
0	2.815	99.95	0.0127	0.0710
11.5	2.629	93.37	2.106	0.9880
11.5	2.569	91.23	2.028	0.9539
11.5	2.662	94.52	2.047	0.9623
24	2.543	90.30	3.478	1.589
24	2.461	87.39	3.297	1.510
24	2.583	91.72	3.347	1.532
36	2.761	98.04	4.790	2.164
36	2.611	92.71	4.558	2.062
36	2.862	101.6	5.129	2.313
48	3.042	108.0	6.994	3.129
48	3.081	109.4	7.358	3.289
48	2.926	103.9	6.436	2.885

Table S43. Quantification of the C-C-bond cleavage-reaction of hmdC to dC at 60 °C.

time [h]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]
0	3.113	84.46	0.0130	0.0056
0	3.167	85.93	0.0133	0.0058
0	3.197	86.75	0.0136	0.0060
12	3.282	89.06	0.0143	0.0066
12	3.264	88.56	0.0134	0.0059
12	3.242	87.98		
23.5	3.319	90.05	0.0148	0.0070
23.5	3.362	91.22	0.0150	0.0072
23.5	3.378	91.66	0.0143	0.0066
35	3.499	94.93	0.0146	0.0068
35	3.392	92.03	0.0150	0.0071
35	3.473	94.23	0.0145	0.0068
46.5	3.616	98.11	0.0153	0.0074
46.5	3.552	96.36	0.0146	0.0069
46.5	3.624	98.32	0.0151	0.0073
58	3.926	106.5	0.0164	0.0083

58	3.914	106.2	0.0163	0.0082
58	4.094	111.1	0.0161	0.0081

Table S44. Quantification of the C-C-bond cleavage-reaction of hmdC to dC at 80 °C.

time [h]	A_{hmdC}/A_{hmdC*}	n_{hmdC} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]
0	3.203	86.90	0.0136	0.0061
0	3.149	85.45	0.0132	0.0057
0	3.257	88.36	0.0134	0.0058
12	3.426	92.95	0.0196	0.0109
12	3.336	90.51	0.0211	0.0121
12	3.620	98.23	0.0201	0.0112
23.5	3.913	106.2	0.0282	0.0177
23.5	3.800	103.1	0.0269	0.0167
23.5	3.824	103.8	0.0278	0.0174
35	4.492	121.9	0.0381	0.0256
35	4.406	119.6	0.0402	0.0273
35	4.374	118.7	0.0381	0.0257
46.5	6.297	170.9	0.0604	0.0435
46.5	6.216	168.7	0.0605	0.0435
46.5	5.955	161.6	0.0610	0.0440

Table S45. Quantification of the C-C-bond cleavage-reaction of fdC to dC at 60 °C.

time [h]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]
0	3.148	111.8	0.0133	0.0058
0	2.987	106.1	0.0130	0.0056
0	3.023	107.4	0.0128	0.0054
12	3.086	109.6	0.0305	0.0196
12	3.153	112.0	0.0245	0.0147
12	3.137	111.4	0.0256	0.0156
23.5	3.323	118.0	0.0373	0.0250
23.5	3.320	117.9	0.0338	0.0222
23.5	3.202	113.7	0.0325	0.0212
35	3.488	123.9	0.0447	0.0309
35	3.442	122.2	0.0427	0.0293
35	3.435	122.0	0.0392	0.0265
46.5	3.627	128.8	0.0535	0.0379

46.5	3.618	128.5	0.0477	0.0333
46.5	3.475	123.4	0.0461	0.0320
58	4.101	145.6	0.0639	0.0463
58	3.923	139.3	0.0579	0.0415
58	3.978	141.3	0.0570	0.0408

Table S46. Quantification of the C-C-bond cleavage-reaction of fdC to dC at 80 °C.

time [h]	A_{fdC}/A_{fdC*}	n_{fdC} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]
0	2.743	97.42	0.0135	0.0059
0	2.845	101.0	0.0134	0.0059
0	3.013	107.0	0.0133	0.0058
12	3.523	125.1	0.0704	0.0514
12	3.185	113.1	0.0700	0.0512
12	3.125	111.0	0.0686	0.0500
23.5	3.514	124.8	0.1171	0.0888
23.5	3.458	122.8	0.1143	0.0866
23.5	3.467	123.1	0.1132	0.0857
35	4.001	142.1	0.1784	0.1379
35	3.986	141.5	0.1705	0.1315
35	3.874	137.5	0.1727	0.1333
46.5	5.416	192.3	0.3026	0.2373
46.5	4.901	174.0	0.2765	0.2164
46.5	5.084	180.5	0.3002	0.2353

Table S47. Quantification of the C-C-bond cleavage-reaction of cadC to dC at 60 °C.

time [h]	A_{cadC}/A_{cadC*}	n_{cadC} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]
0	4.788	86.98	0.3579	0.2815
0	4.740	86.11	0.3542	0.2785
0	4.755	86.39	0.3546	0.2788
12	4.853	88.18	0.3727	0.2933
12	4.913	89.26	0.3819	0.3006
12	4.881	88.68	0.3828	0.3014
23.5	4.945	89.85	0.3927	0.3093
23.5	5.010	91.03	0.3893	0.3066
23.5	4.989	90.64	0.4031	0.3176
35	5.221	94.85	0.4198	0.3310

35	5.105	92.75	0.4089	0.3223
35	5.018	91.17	0.4126	0.3252
46.5	5.296	96.22	0.4285	0.3379
46.5	5.469	99.36	0.4382	0.3457
46.5	5.247	95.32	0.4269	0.3367
58	5.921	107.6	0.4775	0.3771
58	5.951	108.1	0.4900	0.3871
58	6.0729	110.33	0.4946	0.3908

Table S48. Quantification of the C-C-bond cleavage-reaction of cadC to dC at 80 °C.

time [h]	A_{cadC}/A_{cadC*}	n_{cadC} [pmol]	A_{dC}/A_{dC*}	n_{dC} [pmol]
0	4.802	87.23	0.3623	0.2850
0	4.783	86.90	0.3657	0.2877
0	4.814	87.46	0.3785	0.2979
12	5.213	94.70	0.4355	0.3435
12	5.136	93.32	0.4324	0.3411
12	5.265	95.66	0.4385	0.3459
23.5	5.842	106.1	0.5317	0.4205
23.5	5.678	103.2	0.5244	0.4147
23.5	5.893	107.1	0.5369	0.4247
35	6.906	125.5	0.6489	0.5143
35	6.818	123.9	0.6621	0.5248
35	6.976	126.7	0.6639	0.5262
46.5	9.828	178.5	0.9977	0.7933
46.5	9.639	175.1	0.9945	0.7907
46.5	10.01	181.9	1.026	0.8162

Table S49. Mean value and standard deviation of the normalized amount of hmdC, formed via oxidation of mdC at 60.0 °C. The natural logarithm was calculated in order to determine the rate constant. n_0=0 pmol, n_∞=100 pmol, n_t is the mean value n_{hmdC}/n_{mdC} at the respective time.

time [h]	mean value of n_{hmdC}/n_{mdC}	standard deviation of n_{hmdC}/n_{mdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	2.86×10^{-7}	0
12	2.35×10^{-6}	2.43×10^{-6}	2.35×10^{-6}
23.5	3.19×10^{-6}	5.28×10^{-7}	3.19×10^{-6}
35	6.27×10^{-6}	1.23×10^{-6}	6.27×10^{-6}
46.5	8.07×10^{-6}	9.36×10^{-7}	8.07×10^{-6}
58	9.81×10^{-6}	1.16×10^{-6}	9.81×10^{-6}

Table S50. Mean value and standard deviation of the normalized amount of hmdC, formed via oxidation of mdC at 80.0 °C.

time [h]	mean value of n_{hmdC}/n_{mdC}	standard deviation of n_{hmdC}/n_{mdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	7.02×10^{-7}	0
12	9.43×10^{-6}	1.61×10^{-6}	9.43×10^{-6}
23.5	1.59×10^{-6}	2.61×10^{-6}	1.59×10^{-5}
35	2.28×10^{-5}	0.00	2.28×10^{-5}
46.5	2.88×10^{-5}	0.00	2.88×10^{-5}

Table S51. Mean value and standard deviation of the normalized amount of fdC, formed via oxidation of hmdC at 60.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{fdC}/n_{hmdC} at the respective time.

time [h]	mean value of n_{fdC}/n_{hmdC}	standard deviation of n_{fdC}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	3.17×10^{-5}	0
12	1.06×10^{-3}	3.62×10^{-4}	1.06×10^{-3}
23.5	1.83×10^{-3}	2.53×10^{-4}	1.83×10^{-3}
35	2.60×10^{-3}	3.86×10^{-4}	2.60×10^{-3}
46.5	3.45×10^{-3}	5.46×10^{-4}	3.46×10^{-3}
58	4.23×10^{-3}	7.04×10^{-4}	4.24×10^{-3}

Table S52. Mean value and standard deviation of the normalized amount of fdC, formed via oxidation of hmdC at 67.5 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{fdC}/n_{hmdC} at the respective time.

time [h]	mean value of n_{fdC}/n_{hmdC}	standard deviation of n_{fdC}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	1.03×10^{-4}	0
11.5	1.36×10^{-3}	2.97×10^{-4}	1.36×10^{-3}
24	2.39×10^{-3}	5.06×10^{-4}	2.39×10^{-3}
36	3.24×10^{-3}	7.73×10^{-4}	3.25×10^{-3}
48	4.21×10^{-3}	8.88×10^{-4}	4.22×10^{-3}
60	5.22×10^{-3}	1.00×10^{-3}	5.24×10^{-3}

Table S53. Mean value and standard deviation of the normalized amount of fdC, formed via oxidation of hmdC at 75.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{fdC}/n_{hmdC} at the respective time.

time [h]	mean value of n_{fdC}/n_{hmdC}	standard deviation of n_{fdC}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	6.98×10^{-5}	0
11.5	1.52×10^{-3}	5.99×10^{-4}	1.53×10^{-3}
24	2.78×10^{-3}	8.73×10^{-4}	2.78×10^{-3}
36	3.90×10^{-3}	1.10×10^{-3}	3.91×10^{-3}
48	5.04×10^{-3}	1.21×10^{-3}	5.06×10^{-3}
60	5.86×10^{-3}	1.71×10^{-3}	5.89×10^{-3}

Table S54. Mean value and standard deviation of the normalized amount of fdC, formed via oxidation of hmdC at 80.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{fdC}/n_{hmdC} at the respective time.

time [h]	mean value of n_{fdC}/n_{hmdC}	standard deviation of n_{fdC}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	4.65×10^{-5}	0
12	1.65×10^{-3}	1.29×10^{-4}	1.65×10^{-3}
23.5	2.90×10^{-3}	4.91×10^{-4}	2.90×10^{-3}
35	3.97×10^{-3}	6.32×10^{-4}	3.99×10^{-3}
46.5	5.23×10^{-3}	6.89×10^{-4}	5.25×10^{-3}

Table S55. Mean value and standard deviation of the normalized amount of cadC, formed via oxidation of fdC at 60.0 °C. Due to the non-linearity of the oxidation of fdC to cadC the natural logarithm was not calculated.

time [h]	mean value of n_{cadC}/n_{fdC}	standard deviation of n_{cadC}/n_{fdC}
0	0	1.99×10^{-5}
11.5	1.53×10^{-4}	1.60×10^{-5}
24	2.24×10^{-4}	4.41×10^{-6}
36	2.64×10^{-4}	4.41×10^{-5}
48	3.12×10^{-4}	2.65×10^{-5}
60	3.33×10^{-4}	1.66×10^{-5}

Table S56. Mean value and standard deviation of the normalized amount of cadC, formed via oxidation of fdC at 67.5 °C. Due to the non-linearity of the oxidation of fdC to cadC the natural logarithm was not calculated.

time [h]	mean value of n_{cadC}/n_{fdC}	standard deviation of n_{cadC}/n_{fdC}
0	0	1.01×10^{-5}
11.5	1.91×10^{-4}	7.43×10^{-6}
24	2.55×10^{-4}	3.80×10^{-6}
36	3.10×10^{-4}	1.13×10^{-5}
48	3.39×10^{-4}	9.78×10^{-6}
60	3.88×10^{-4}	1.61×10^{-5}

Table S57. Mean value and standard deviation of the normalized amount of cadC, formed via oxidation of fdC at 75.0 °C. Due to the non-linearity of the oxidation of fdC to cadC the natural logarithm was not calculated.

time [h]	mean value of n_{cadC}/n_{fdC}	standard deviation of n_{cadC}/n_{fdC}
0	0	1.49×10^{-5}
11.5	2.89×10^{-4}	7.86×10^{-5}
24	4.36×10^{-4}	2.87×10^{-5}
36	4.88×10^{-4}	9.07×10^{-5}
48	5.70×10^{-4}	8.86×10^{-5}
60	6.41×10^{-4}	1.43×10^{-4}

Table S58. Mean value and standard deviation of the normalized amount of cadC, formed via oxidation of fdC at 82.5 °C. Due to the non-linearity of the oxidation of fdC to cadC the natural logarithm was not calculated.

time [h]	mean value of n_{cadC}/n_{fdC}	standard deviation of n_{cadC}/n_{fdC}
0	0	7.61×10^{-6}
11.5	4.48×10^{-4}	7.46×10^{-5}
24	6.42×10^{-4}	1.22×10^{-4}
36	8.48×10^{-4}	1.18×10^{-4}
48	1.09×10^{-3}	1.58×10^{-4}
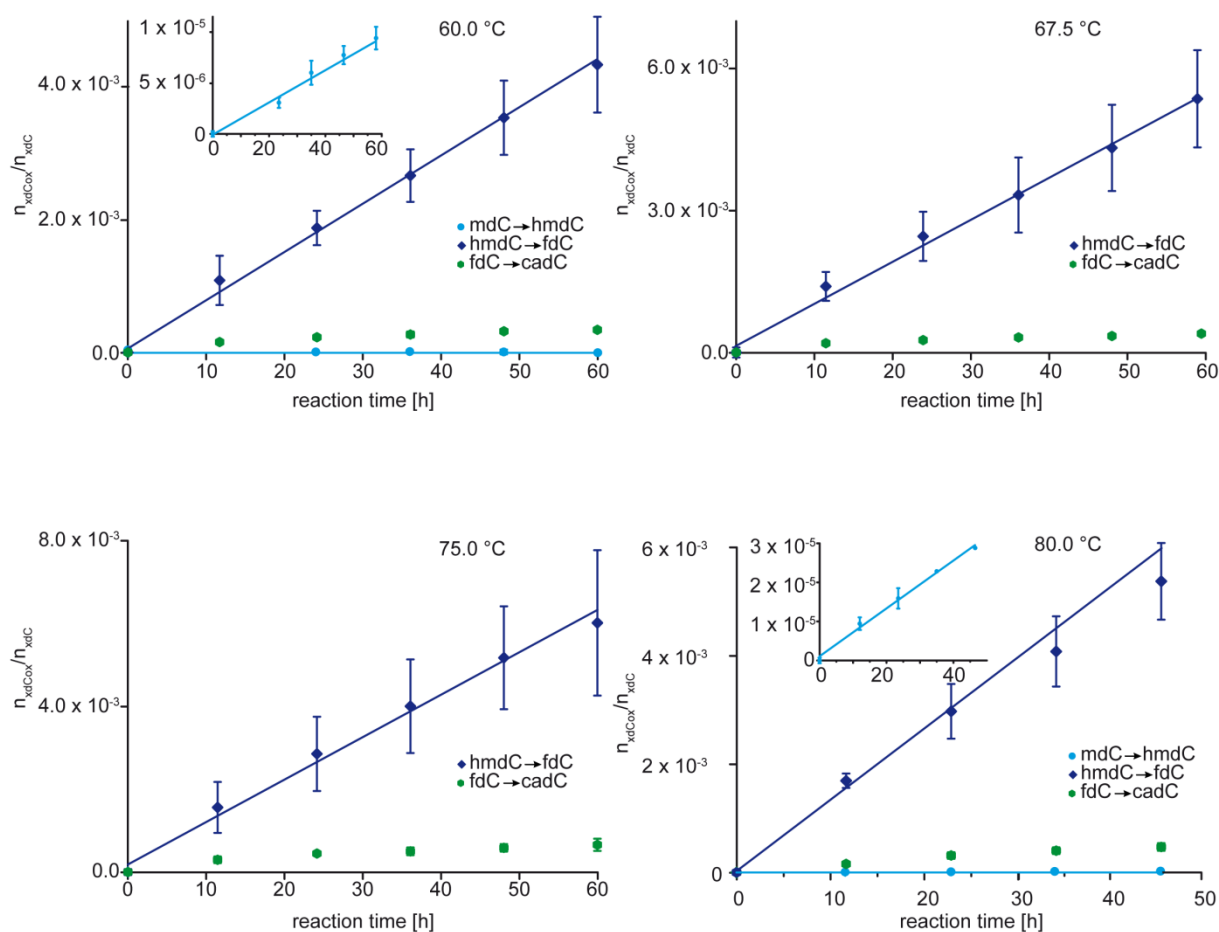
Figure S2. Oxidation kinetics of mdC to hmdC (cyan), hmdC to fdC (blue) and fdC to cadC (green) at four different temperatures.

Table S59. Coefficients of determinations R^2 and linear equations of linear regression analyses of straights in Figure S2 including Figure 1 determined by ORIGIN® 6.0.

reaction	T [°C]	Coefficient of determination R^2	linear equation
mdC➔hmdC	60.0	0.9936	$y = 1.65301 \times 10^{-7} x - 8.49873 \times 10^{-8}$
mdC➔hmdC	80.0	0.9970	$y = 6.12566 \times 10^{-7} x + 1.06006 \times 10^{-6}$
hmdC➔fdC	60.0	0.9992	$y = 7.54166 \times 10^{-5} x + 1.32075 \times 10^{-6}$
hmdC➔fdC	67.5	0.9971	$y = 8.41887 \times 10^{-5} x + 2.18023 \times 10^{-4}$
hmdC➔fdC	75.0	0.9957	$y = 9.71991 \times 10^{-5} x + 2.75459 \times 10^{-4}$
hmdC➔fdC	80.0	0.9955	$y = 1.24922 \times 10^{-4} x + 1.23963 \times 10^{-5}$
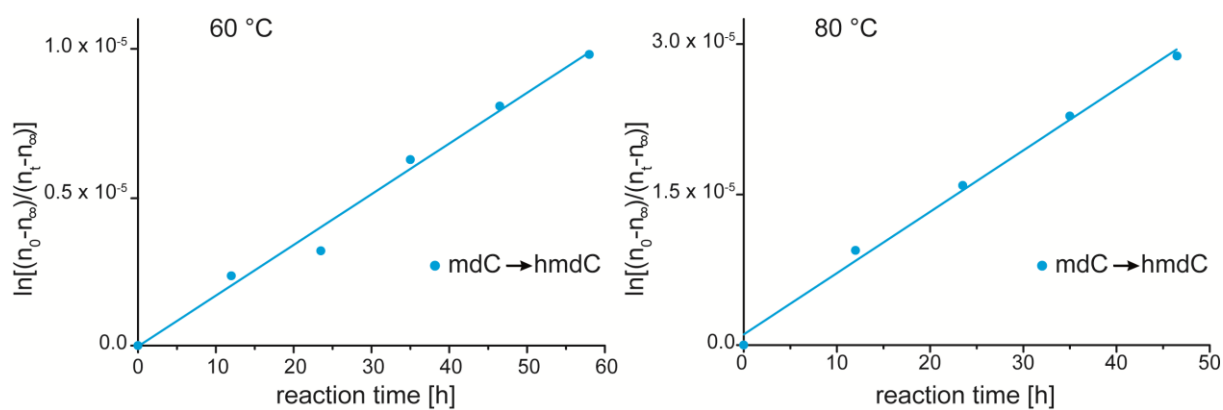
Figure S3. Determination of ψ- first order rate constants of the oxidation of mdC to hmdC at 60 °C (left) and 80 °C (right). n_0 is the normalized mean value of the amount of hmdC at t=0 h, n_t is the normalized mean value of the amount of hmdC at time t, n_∞ is 100 pmol. The rate constant is the slope derived from linear regression by ORIGIN® 6.0.

Table S60. Coefficients of determinations R^2 and linear equations of linear regression analyses of straights in Figure S3 determined by ORIGIN® 6.0.

reaction	T [°C]	Coefficient of determination R^2	linear equation
mdC→hmdC	60.0	0.9937	$y = 1.71034 \times 10^{-7} x - 4.02097 \times 10^{-8}$
mdC→hmdC	80.0	0.9970	$y = 6.12599 \times 10^{-7} x + 1.06002 \times 10^{-6}$

273
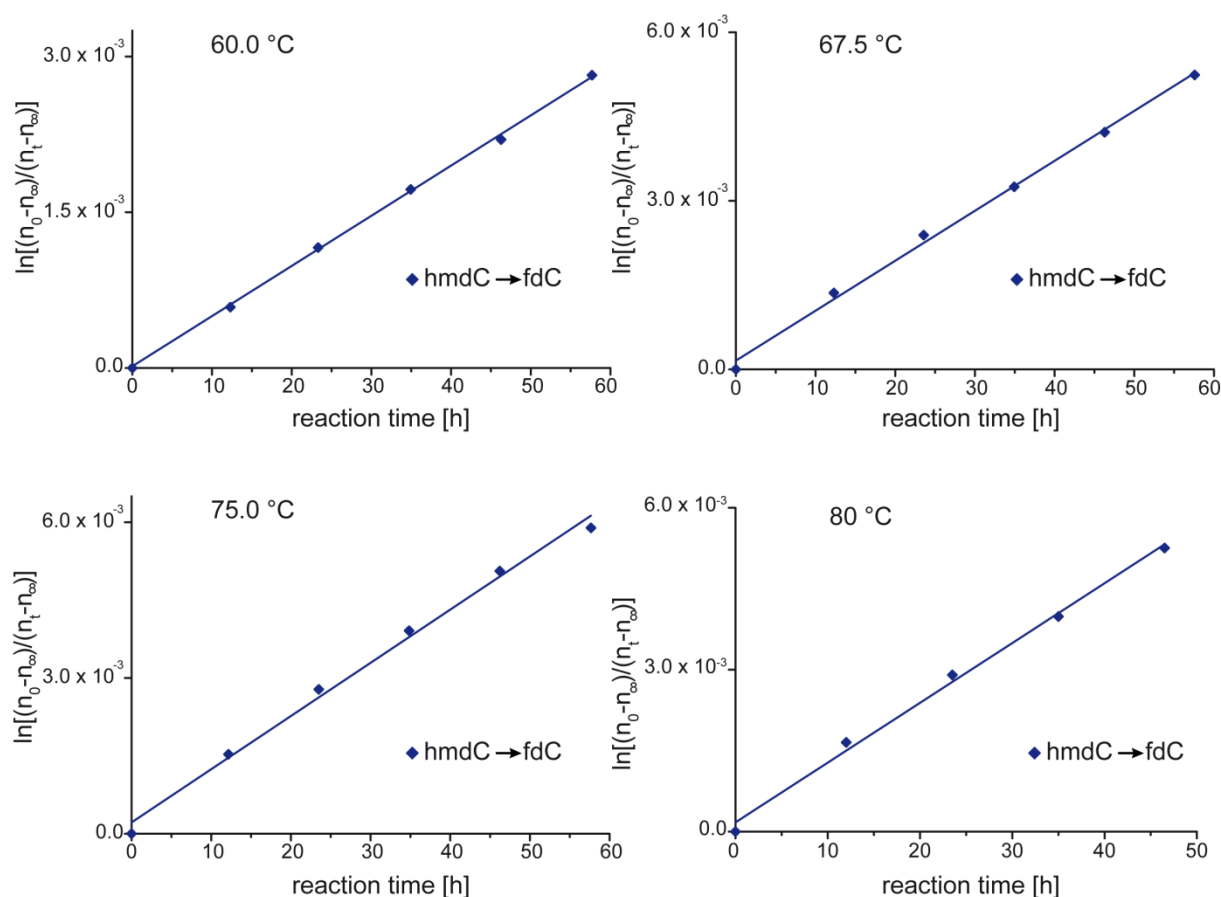
Figure S4. Determination of ψ- first order rate constants of the oxidation of hmdC to fdC at four different temperatures. n_0 is the normalized mean value of the amount of hmdC at t=0 h, n_t is the normalized mean value of the amount of hmdC at time t, n_∞ is 100 pmol. The rate constant is the slope derived from linear regression by ORIGIN® 6.0.

Table S61. Coefficients of determinations R^2 and linear equations of linear regression analyses of straights in Figure S4 determined by ORIGIN® 6.0.

reaction	T [°C]	Coefficient of determination R^2	linear equation
hmdC→fdC	60.0	0.9991	$y = 7.20891 \times 10^{-5} x + 9.88194 \times 10^{-5}$
hmdC→fdC	67.5	0.9984	$y = 8.88895 \times 10^{-5} x + 1.55662 \times 10^{-4}$
hmdC→fdC	75.0	0.9968	$y = 1.02523 \times 10^{-4} x + 2.15271 \times 10^{-4}$
hmdC→fdC	80.0	0.9977	$y = 1.10755 \times 10^{-4} x + 1.67271 \times 10^{-4}$

Table S62. Rate constants of the oxidation of mdC to hmdC at two different temperatures.

	60 °C	80.0 °C
mdC	$4.7 \pm 0.6 \times 10^{-11}$	$1.9 \pm 0.2 \times 10^{-10}$

Table S63. Rate constants of the oxidation of hmdC to fdC at four different temperatures.

	60 °C	67.5 °C	75 °C	80.0 °C
hmdC	$2.2\pm0.2 \times 10^{-8}$	$2.7\pm0.2 \times 10^{-8}$	$3.2\pm0.2 \times 10^{-8}$	$3.4\pm0.3 \times 10^{-8}$

Table S64: Table for the extrapolation of the rate constant of the oxidation of hmdC to fdC to T = 37 °C.

temperature T [K]	1/T [K^{-1}]	ln(k$_T$)
333.15	3.00×10^{-3}	-17.63
340.65	2.94×10^{-3}	-17.42
348.15	2.87×10^{-3}	-17.27
353.15	2.83×10^{-3}	-17.20
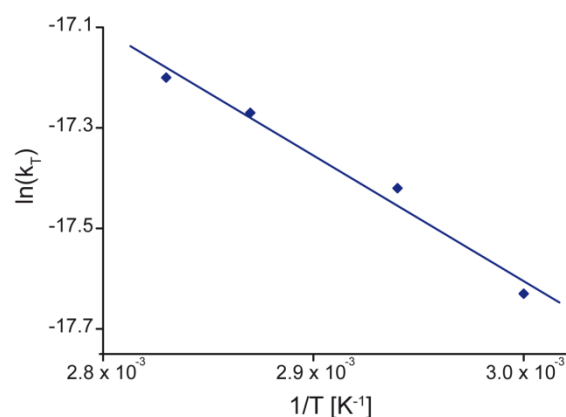
Figure S5: Graph for the extrapolation of the rate constants of the oxidation of hmdC to fdC to T = 37 °C. $R^2 = 0.9878$; $y = -2402.03231 \, x - 10.3968$.

Table S65. Mean value and standard deviation of the normalized amount of dU, formed via deamination of dC at 60.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{dU}/n_{dC} at the respective time.

time [h]	mean value of n_{dU}/n_{dC}	standard deviation of n_{dU}/n_{dC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	1.23×10^{-4}	0
11.5	7.74×10^{-4}	2.82×10^{-4}	7.75×10^{-4}
24	1.54×10^{-3}	3.39×10^{-4}	1.54×10^{-3}
36	2.28×10^{-3}	9.67×10^{-5}	2.28×10^{-3}
48	2.84×10^{-3}	2.89×10^{-4}	2.85×10^{-3}
60	3.50×10^{-3}	1.38×10^{-4}	3.51×10^{-3}

Table S66. Mean value and standard deviation of the normalized amount of dU, formed via deamination of dC at 67.5 °C. The natural logarithm was calculated in order to determine the rate constant. n_0=0 pmol, n_∞=100 pmol, n_t is the mean value n_{dU}/n_{dC} at the respective time.

time [h]	mean value of n_{dU}/n_{dC}	standard deviation of n_{dU}/n_{dC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	1.44×10^{-4}	0
11.5	1.59×10^{-3}	2.94×10^{-4}	1.60×10^{-3}
24	3.33×10^{-3}	2.52×10^{-4}	3.33×10^{-3}
36	4.97×10^{-3}	8.45×10^{-4}	4.99×10^{-3}
48	6.87×10^{-3}	3.39×10^{-4}	6.90×10^{-3}
60	8.33×10^{-3}	6.00×10^{-4}	8.37×10^{-3}

Table S67. Mean value and standard deviation of the normalized amount of dU, formed via deamination of dC at 75.0 °C. The natural logarithm was calculated in order to determine the rate constant. n_0=0 pmol, n_∞=100 pmol, n_t is the mean value n_{dU}/n_{dC} at the respective time.

time [h]	mean value of n_{dU}/n_{dC}	standard deviation of n_{dU}/n_{dC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	4.77×10^{-4}	0
11.5	3.88×10^{-3}	4.40×10^{-4}	3.89×10^{-3}
24	7.63×10^{-3}	7.18×10^{-4}	7.66×10^{-3}
36	1.19×10^{-2}	5.57×10^{-4}	1.20×10^{-2}
48	1.57×10^{-2}	9.01×10^{-4}	1.59×10^{-2}
60	2.00×10^{-2}	8.61×10^{-4}	2.02×10^{-2}

Table S68. Mean value and standard deviation of the normalized amount of dU, formed via deamination of dC at 82.5 °C. The natural logarithm was calculated in order to determine the rate constant. n_0=0 pmol, n_∞=100 pmol, n_t is the mean value n_{dU}/n_{dC} at the respective time.

time [h]	mean value of n_{dU}/n_{dC}	standard deviation of n_{dU}/n_{dC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	4.69×10^{-5}	0
11.5	8.06×10^{-3}	1.57×10^{-4}	8.10×10^{-3}
24	1.86×10^{-2}	4.38×10^{-4}	1.88×10^{-2}
36	2.54×10^{-2}	2.02×10^{-3}	2.57×10^{-2}
48	3.63×10^{-2}	9.17×10^{-4}	3.69×10^{-2}

Table S69. Mean value and standard deviation of the normalized amount of dT, formed via deamination of mdC at 60.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{dT}/n_{mdC} at the respective time.

time [h]	mean value of n_{dT}/n_{mdC}	standard deviation of n_{dT}/n_{mdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	6.11×10^{-5}	0
11.5	6.73×10^{-4}	9.53×10^{-5}	6.74×10^{-4}
24	1.15×10^{-3}	1.85×10^{-5}	1.15×10^{-3}
36	1.53×10^{-3}	8.27×10^{-5}	1.53×10^{-3}
48	2.00×10^{-3}	6.10×10^{-5}	2.00×10^{-3}
60	2.72×10^{-3}	1.76×10^{-4}	2.73×10^{-3}

Table S70. Mean value and standard deviation of the normalized amount of dT, formed via deamination of mdC at 67.5 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{dT}/n_{mdC} at the respective time.

time [h]	mean value of n_{dT}/n_{mdC}	standard deviation of n_{dT}/n_{mdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	6.91×10^{-5}	0
11.5	1.33×10^{-3}	1.96×10^{-4}	1.34×10^{-3}
24	2.31×10^{-3}	1.29×10^{-4}	2.32×10^{-3}
36	3.10×10^{-3}	5.81×10^{-5}	3.11×10^{-3}
48	4.55×10^{-3}	7.64×10^{-5}	4.57×10^{-3}
60	5.93×10^{-3}	6.08×10^{-4}	5.96×10^{-3}

Table S71. Mean value and standard deviation of the normalized amount of dT, formed via deamination of mdC at 75.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{dT}/n_{mdC} at the respective time.

time [h]	mean value of n_{dT}/n_{mdC}	standard deviation of n_{dT}/n_{mdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	1.47×10^{-4}	0
11.5	2.97×10^{-3}	8.00×10^{-5}	2.98×10^{-3}
24	5.13×10^{-3}	6.86×10^{-4}	5.15×10^{-3}
36	7.79×10^{-3}	3.29×10^{-4}	7.83×10^{-3}
48	1.06×10^{-2}	5.40×10^{-4}	1.07×10^{-2}
60	1.44×10^{-2}	4.22×10^{-4}	1.45×10^{-2}

Table S72. Mean value and standard deviation of the normalized amount of dT, formed via deamination of mdC at 82.5 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{dT}/n_{mdC} at the respective time.

time [h]	mean value of n_{dT}/n_{mdC}	standard deviation of n_{dT}/n_{mdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	3.31×10^{-5}	0
11.5	6.49×10^{-3}	6.73×10^{-4}	6.52×10^{-3}
24	1.15×10^{-2}	4.93×10^{-4}	1.16×10^{-2}
36	1.80×10^{-2}	1.59×10^{-3}	1.82×10^{-2}
48	2.30×10^{-2}	5.10×10^{-4}	2.33×10^{-2}

Table S73. Mean value and standard deviation of the normalized amount of hmdU, formed via deamination of hmdC at 60.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{hmdU}/n_{hmdC} at the respective time.

time [h]	mean value of n_{hmdU}/n_{hmdC}	standard deviation of n_{hmdU}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	9.43×10^{-5}	0
11.5	3.01×10^{-4}	1.06×10^{-4}	3.01×10^{-4}
24	8.81×10^{-4}	5.52×10^{-5}	8.84×10^{-4}
36	1.35×10^{-3}	8.16×10^{-5}	1.35×10^{-3}
48	1.62×10^{-3}	2.05×10^{-4}	1.62×10^{-3}
60	1.87×10^{-3}	1.57×10^{-4}	1.87×10^{-3}

Table S74. Mean value and standard deviation of the normalized amount of hmdU, formed via deamination of hmdC at 67.5 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{hmdU}/n_{hmdC} at the respective time.

time [h]	mean value of n_{hmdU}/n_{hmdC}	standard deviation of n_{hmdU}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	1.12×10^{-4}	0
11.5	1.08×10^{-3}	3.43×10^{-4}	1.08×10^{-3}
24	1.88×10^{-3}	1.71×10^{-4}	1.89×10^{-3}
36	2.90×10^{-3}	6.71×10^{5}	2.92×10^{-3}
48	3.55×10^{-3}	2.01×10^{-4}	3.57×10^{-3}
60	4.81×10^{-3}	5.59×10^{-5}	4.83×10^{-3}

Table S75. Mean value and standard deviation of the normalized amount of hmdU, formed via deamination of hmdC at 75.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{hmdU}/n_{hmdC} at the respective time.

time [h]	mean value of n_{hmdU}/n_{hmdC}	standard deviation of n_{hmdU}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	8.57×10^{-5}	0
11.5	2.38×10^{-3}	1.11×10^{-4}	2.39×10^{-3}
24	4.42×10^{-3}	1.07×10^{-4}	4.44×10^{-3}
36	6.75×10^{-3}	7.65×10^{-5}	9.69×10^{-3}
48	8.84×10^{-3}	6.19×10^{-4}	8.90×10^{-3}
60	9.99×10^{-3}	2.70×10^{-4}	1.01×10^{-2}

Table S76. Mean value and standard deviation of the normalized amount of hmdU, formed via deamination of hmdC at 82.5 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{hmdU}/n_{hmdC} at the respective time.

time [h]	mean value of n_{hmdU}/n_{hmdC}	standard deviation of n_{hmdU}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	1.05×10^{-4}	0
11.5	4.38×10^{-3}	1.09×10^{-4}	4.40×10^{-3}
24	9.49×10^{-3}	2.12×10^{-4}	9.56×10^{-3}
36	1.40×10^{-2}	7.65×10^{-4}	1.15×10^{-2}
48	1.83×10^{-2}	4.35×10^{-4}	1.85×10^{-2}

Table S77. Mean value and standard deviation of the normalized amount of fdU, formed via deamination of fdC at 60.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{fdU}/n_{fdC} at the respective time.

time [h]	mean value of n_{fdU}/n_{fdC}	standard deviation of n_{fdU}/n_{fdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	5.13×10^{-6}	0
11.5	9.01×10^{-4}	2.42×10^{-5}	9.02×10^{-4}
24	1.66×10^{-3}	4.63×10^{-5}	1.66×10^{-3}
36	2.28×10^{-3}	3.91×10^{-5}	2.28×10^{-3}
48	2.95×10^{-3}	5.42×10^{-5}	2.95×10^{-3}
60	3.36×10^{-3}	1.37×10^{-4}	3.37×10^{-3}

Table S78. Mean value and standard deviation of the normalized amount of fdU, formed via deamination of fdC at 67.5 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{fdU}/n_{fdC} at the respective time.

time [h]	mean value of n_{fdU}/n_{fdC}	standard deviation of n_{fdU}/n_{fdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	4.89×10^{-6}	0
11.5	2.11×10^{-3}	1.18×10^{-4}	2.11×10^{-3}
24	3.48×10^{-3}	9.82×10^{-5}	3.49×10^{-3}
36	4.61×10^{-3}	1.16×10^{-4}	4.62×10^{-3}
48	5.79×10^{-3}	9.67×10^{-5}	5.81×10^{-3}
60	6.96×10^{-3}	3.33×10^{-4}	6.99×10^{-3}

Table S79. Mean value and standard deviation of the normalized amount of fdU, formed via deamination of fdC at 75.0 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{fdU}/n_{fdC} at the respective time.

time [h]	mean value of n_{fdU}/n_{fdC}	standard deviation of n_{fdU}/n_{fdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	5.27×10^{-5}	0
11.5	4.47×10^{-3}	8.11×10^{-5}	4.48×10^{-3}
24	7.45×10^{-3}	2.65×10^{-4}	7.48×10^{-3}
36	1.01×10^{-2}	2.29×10^{-4}	1.02×10^{-2}
48	1.23×10^{-2}	2.40×10^{-4}	1.24×10^{-2}
60	1.65×10^{-2}	7.09×10^{-4}	1.66×10^{-2}

Table S80. Mean value and standard deviation of the normalized amount of fdU, formed via deamination of fdC at 82.5 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{fdU}/n_{fdC} at the respective time.

time [h]	mean value of n_{fdU}/n_{fdC}	standard deviation of n_{fdU}/n_{fdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	8.89×10^{-7}	0
11.5	9.70×10^{-3}	2.05×10^{-4}	9.75×10^{-3}
24	1.65×10^{-2}	4.57×10^{-4}	1.66×10^{-2}
36	2.16×10^{-2}	3.56×10^{-4}	2.19×10^{-2}
48	2.82×10^{-2}	1.15×10^{-3}	2.86×10^{-2}

Table S81. Mean value and standard deviation of the normalized amount of fdU, formed via deamination of fdC at 80 °C. The natural logarithm was calculated in order to determine the rate constant. n_0=0 pmol, n_∞=100 pmol, n_t is the mean value n_{fdU}/n_{fdC} at the respective time.

time [h]	mean value of n_{fdU}/n_{fdC}	standard deviation of n_{fdU}/n_{fdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	3.47×10^{-5}	0
12	2.44×10^{-3}	1.15×10^{-4}	1.59×10^{-4}
23.5	4.02×10^{-3}	1.73×10^{-4}	2.80×10^{-3}
35	5.23×10^{-3}	4.64×10^{-5}	4.06×10^{-3}
46.5	6.48×10^{-3}	2.49×10^{-4}	4.72×10^{-3}
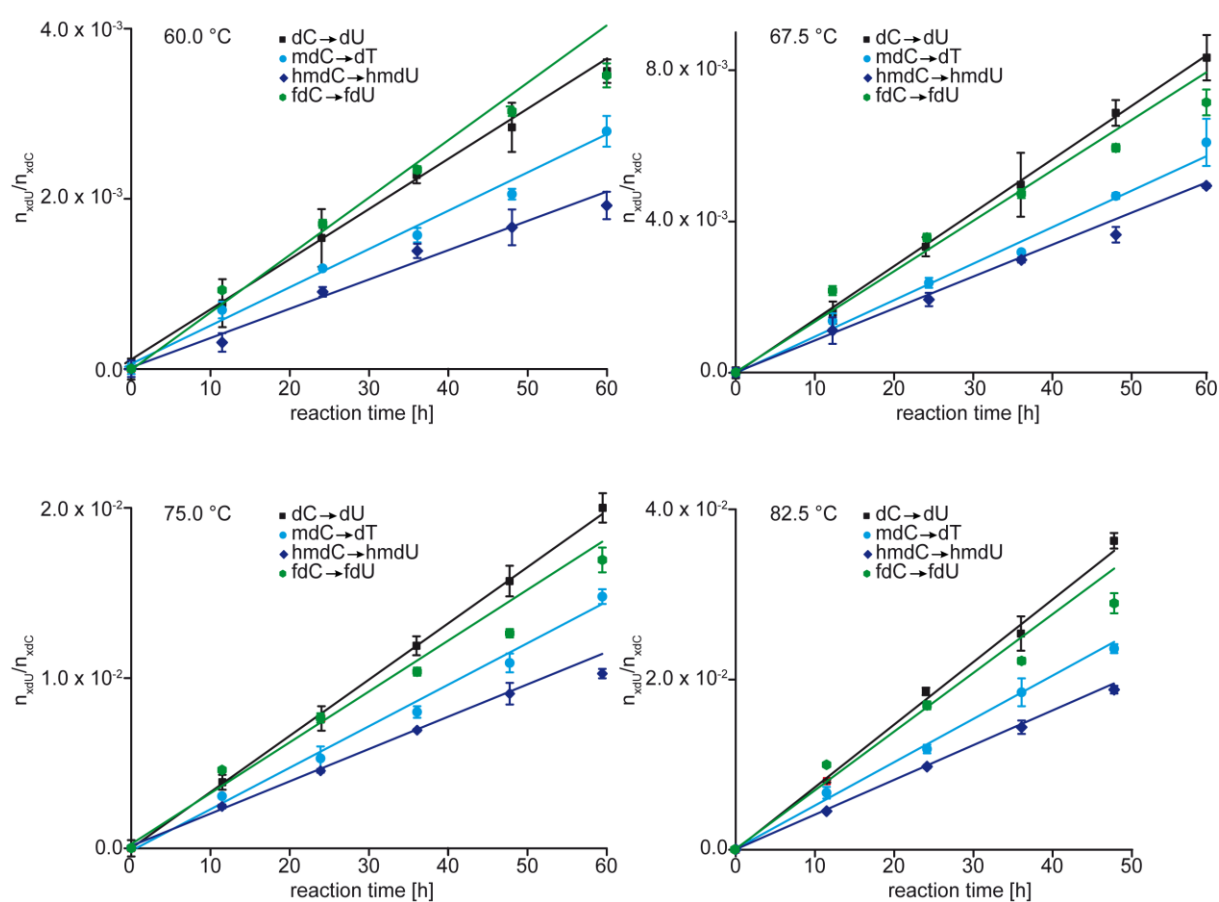
Figure S6. Deamination kinetics of dC to dU (black), mdC to dT (cyan), hmdC to hmdU (blue) and fdC to fdU (green) at four different temperatures.

Table S82. Coefficients of determinations R^2 and linear equations of linear regression analyses of straights in Figure S6 including Figure 3 determined by ORIGIN® 6.0.

reaction	T [°C]	Coefficient of determination R^2	linear equation
dC→dU	60.0	0.9978	$y = 5.89944 \times 10^{-5}\, x + 6.56993 \times 10^{-5}$
dC→dU	67.5	0.9998	$y = 1.41066 \times 10^{-4}\, x - 1.03247 \times 10^{-5}$
dC→dU	75.0	0.9998	$y = 3.3016 \times 10^{-4}\, x - 1.13814 \times 10^{-6}$
dC→dU	82.5	0.9989	$y = 7.34897 \times 10^{-4}\, x - 1.97238 \times 10^{-5}$
mdC→dT	60.0	0.9909	$y = 4.12683 \times 10^{-5}\, x + 1.35455 \times 10^{-4}$
mdC→dT	67.5	0.9958	$y = 9.13478 \times 10^{-5}\, x - 1.56265 \times 10^{-5}$
mdC→dT	75.0	0.9963	$y = 2.2872 \times 10^{-4}\, x + 2.27499 \times 10^{-4}$
mdC→dT	82.5	0.9996	$y = 4.82077 \times 10^{-4}\, x + 1.64584 \times 10^{-6}$
hmdC→hmdU	60.0	0.9881	$y = 3.4428 \times 10^{-5}\, x + 2.33921 \times 10^{-5}$
hmdC→hmdU	67.5	0.9993	$y = 7.98926 \times 10^{-5}\, x + 5.22682 \times 10^{-6}$
hmdC→hmdU	75.0	0.9976	$y = 1.80685 \times 10^{-4}\, x + 1.27125 \times 10^{-4}$
hmdC→hmdU	82.5	0.9997	$y = 3.86935 \times 10^{-4}\, x - 1.30514 \times 10^{-5}$
fdC→fdU	60.0	0.9961	$y = 6.40342 \times 10^{-5}\, x + 6.16821 \times 10^{-6}$
fdC→fdU	67.5	0.9955	$y = 1.26479 \times 10^{-4}\, x + 1.46195 \times 10^{-6}$
fdC→fdU	75.0	0.9832	$y = 2.83593 \times 10^{-4}\, x + 2.95769 \times 10^{-4}$
fdC→fdU	82.5	0.9901	$y = 5.66279 \times 10^{-4}\, x + 1.67000 \times 10^{-3}$
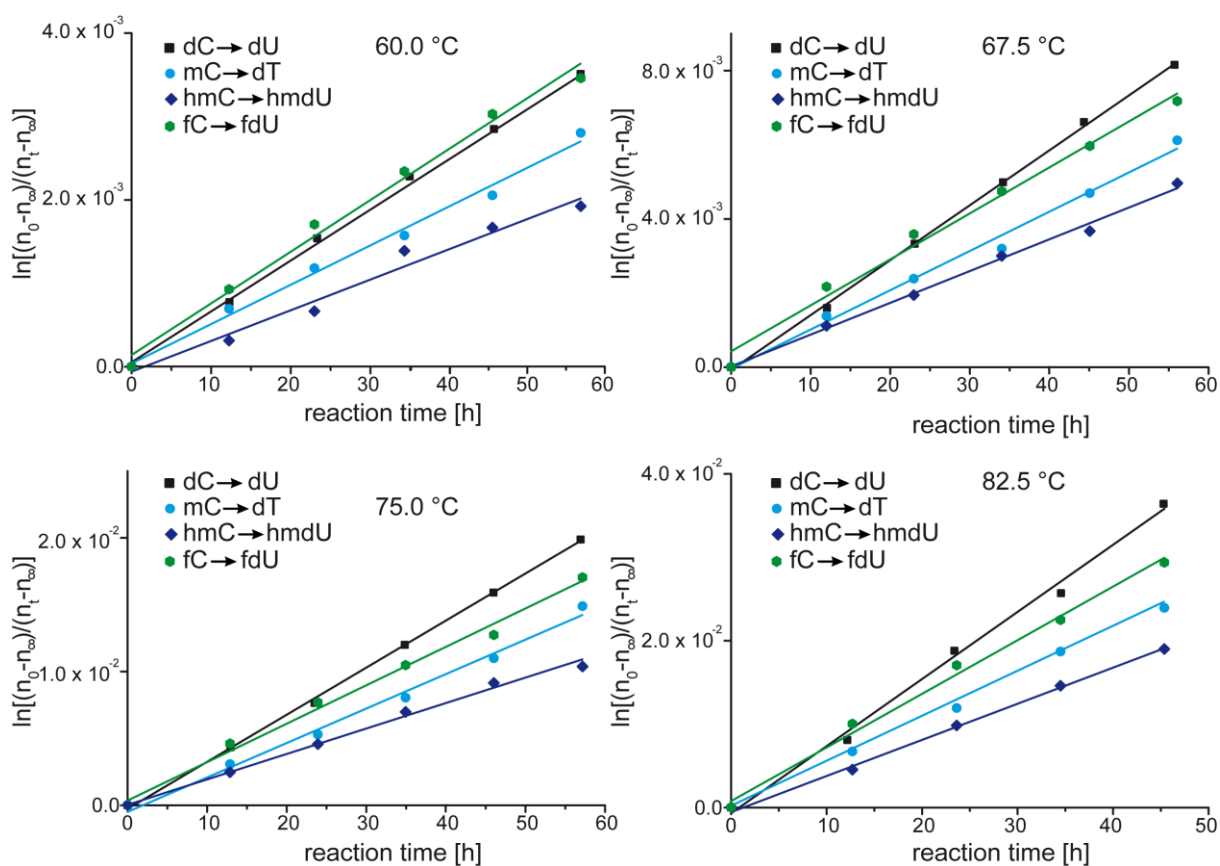
Figure S7. Determination of ψ- first order rate constants of the deamination of dC (black), mdC (cyan), hmdC (blue) and fdC (green) at four different temperatures. n_0 is the normalized mean value of the amount of the respective nucleoside at t=0 h, n_t is the normalized mean value of the amount of the respective nucleoside at time t, n_∞ is 100 pmol. The rate constant is the slope derived from linear regression by ORIGIN® 6.0.

Table S83. Coefficients of determinations R^2 and linear equations of linear regression analyses of straights in Figure S7 determined by ORIGIN® 6.0.

reaction	T [°C]	Coefficient of determination R^2	linear equation
dC→dU	60.0	0.9988	$y = 6.09953 \times 10^{-5} x + 5.17169 \times 10^{-5}$
dC→dU	67.5	0.9994	$y = 1.48024 \times 10^{-4} x - 1.10797 \times 10^{-4}$
dC→dU	75.0	0.9996	$y = 3.51737 \times 10^{-4} x - 2.81179 \times 10^{-4}$
dC→dU	82.5	0.9973	$y = 8.02515 \times 10^{-4} x - 6.88923 \times 10^{-4}$
mdC→dT	60.0	0.9955	$y = 6.09953 \times 10^{-5} x + 5.17169 \times 10^{-5}$
mdC→dT	67.5	0.9957	$y = 1.00375 \times 10^{-4} x - 3.8692 \times 10^{-5}$
mdC→dT	75.0	0.9965	$y = 2.44686 \times 10^{-4} x - 2.51513 \times 10^{-4}$
mdC→dT	82.5	0.9993	$y = 5.12094 \times 10^{-4} x - 6.21963 \times 10^{-5}$
hmdC→hmdU	60.0	0.9896	$y = 3.42504 \times 10^{-5} x + 7.95559 \times 10^{-6}$
hmdC→hmdU	67.5	0.9977	$y = 8.13784 \times 10^{-5} x + 1.26521 \times 10^{-5}$
hmdC→hmdU	75.0	0.9966	$y = 1.80177 \times 10^{-4} x + 2.00011 \times 10^{-4}$

283

hmdC➔hmdU	82.5	0.9994	$y = 4.11047 \times 10^{-4}\, x - 1.89229 \times 10^{-4}$
fdC➔fdU	60.0	0.9948	$y = 5.87829 \times 10^{-5}\, x + 1.50568 \times 10^{-4}$
fdC➔fdU	67.5	0.9942	$y = 1.17756 \times 10^{-4}\, x + 4.08673 \times 10^{-4}$
fdC➔fdU	75.0	0.9953	$y = 2.72685 \times 10^{-4}\, x + 6.01379 \times 10^{-4}$
fdC➔fdU	82.5	0.9955	$y = 6.10289 \times 10^{-4}\, x + 1.23000 \times 10^{-3}$

Table S84. Rate constants of the deamination of dC to dU, mdC to dT, hmdC to hmdU and fdC to fdU at four different temperatures.

	60 °C	**67.5 °C**	**75 °C**	**82.5 °C**
dC	$1.8 \pm 0.1 \times 10^{-8}$	$3.9 \pm 0.2 \times 10^{-8}$	$9.4 \pm 0.4 \times 10^{-8}$	$2.1 \pm 0.2 \times 10^{-7}$
mdC	$1.3 \pm 0.1 \times 10^{-8}$	$2.8 \pm 0.2 \times 10^{-8}$	$6.5 \pm 0.4 \times 10^{-8}$	$1.4 \pm 0.1 \times 10^{-7}$
hmdC	$9.2 \pm 1.3 \times 10^{-9}$	$2.3 \pm 0.2 \times 10^{-8}$	$5.6 \pm 1.1 \times 10^{-8}$	$1.0 \pm 0.1 \times 10^{-7}$
fdC	$1.8 \pm 0.2 \times 10^{-8}$	$3.9 \pm 0.8 \times 10^{-8}$	$8.4 \pm 1.4 \times 10^{-8}$	$1.9 \pm 0.3 \times 10^{-7}$

Table S85. Table for the extrapolation of the rate constant of the deamination reactions to T = 37 °C.

temperature T [K]	1/T [K^{-1}]	dC➔dU ln(k_T)	mdC➔dT ln(k_T)	hmdC➔hmdUln (k_T)	fdC➔fdU ln(k_T)
333.15	3.00×10^{-3}	−17.86	−18.14	−18.50	−17.81
340.65	2.94×10^{-3}	−17.05	−17.40	−17.60	−17.07
348.15	2.87×10^{-3}	−16.18	−16.55	−16.77	−16.29
355.65	2.81×10^{-3}	−15.38	−15.76	−16.04	−15.47
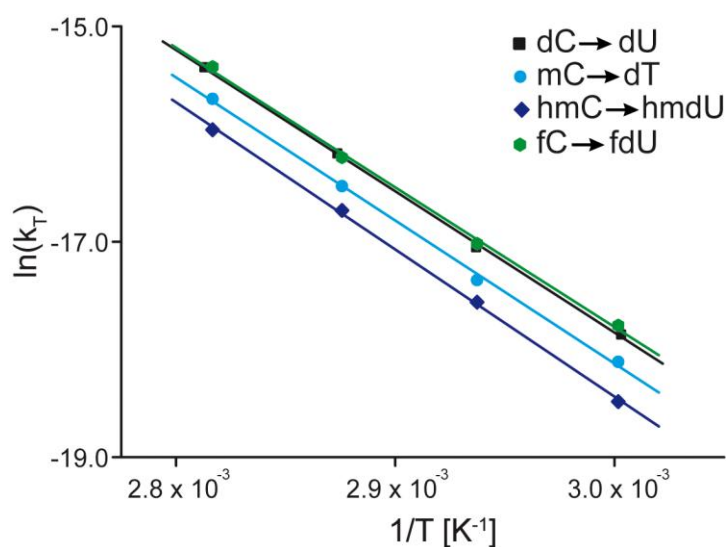
Figure S8: Graph for the extrapolation of the rate constants of the deaminaion reaction to T = 37 °C.

Table S86. Coefficients of determinations R^2 and linear equations of linear regression analyses of straights in Figure S8 determined by ORIGIN® 6.0.

reaction	Coefficient of determination R^2	linear equation
dC→dU	–0.99973	y = –13113.57915 x +21.49831
mdC→dT	–0.99921	y = –12604.48382 x +19.67358
hmdC→hmdU	–0.99963	y = –12966.83026 x +20.46177
fdC→fdU	–0.99910	y = –12302.68123 x +19.09886

Table S87. Mean value and standard deviation of the normalized amount of dC, formed via C-C-bond cleavage of hmdC at 80 °C. The natural logarithm was calculated in order to determine the rate constant. n_0=0 pmol, n_∞=100 pmol, n_t is the mean value n_{dC}/n_{hmdC} at the respective time.

time [h]	mean value of n_{dC}/n_{hmdC}	standard deviation of n_{dC}/n_{hmdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	2.01×10^{-6}	0
12	5.39×10^{-5}	1.03×10^{-5}	5.39×10^{-5}
23.5	0.97×10^{-4}	2.89×10^{-6}	9.78×10^{-5}
35	1.50×10^{-4}	9.17×10^{-6}	1.51×10^{-4}
46.5	1.94×10^{-4}	9.39×10^{-6}	1.94×10^{-4}

Table S88. Mean value and standard deviation of the normalized amount of dC, formed via C-C-bond cleavage of fdC at 80 °C. The natural logarithm was calculated in order to determine the rate constant. n_0=0 pmol, n_∞=100 pmol, n_t is the mean value n_{dC}/n_{fdC} at the respective time.

time [h]	mean value of n_{dC}/n_{fdC}	standard deviation of n_{dC}/n_{fdC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	3.63×10^{-6}	0
12	3.80×10^{-4}	2.33×10^{-5}	3.80×10^{-4}
23.5	6.47×10^{-4}	7.81×10^{-5}	6.47×10^{-4}
35	8.99×10^{-4}	2.35×10^{-5}	8.99×10^{-4}
46.5	1.21×10^{-3}	3.79×10^{-5}	1.20×10^{-3}

Table S89. Mean value and standard deviation of the normalized amount of dC, formed via C-C-bond cleavage of cadC at 80 °C. The natural logarithm was calculated in order to determine the rate constant. $n_0=0$ pmol, $n_\infty=100$ pmol, n_t is the mean value n_{dC}/n_{cadC} at the respective time.

time [h]	mean value of n_{dC}/n_{cadC}	standard deviation of n_{dC}/n_{cadC}	$\ln[(n_0-n_\infty)/(n_t-n_\infty)]$
0	0	7.16×10^{-5}	0
12	3.00×10^{-4}	1.98×10^{-5}	3.06×10^{-4}
23.5	6.50×10^{-4}	3.20×10^{-5}	6.57×10^{-4}
35	8.30×10^{-4}	6.99×10^{-5}	8.38×10^{-4}
46.5	1.15×10^{-3}	3.64×10^{-5}	1.16×10^{-3}
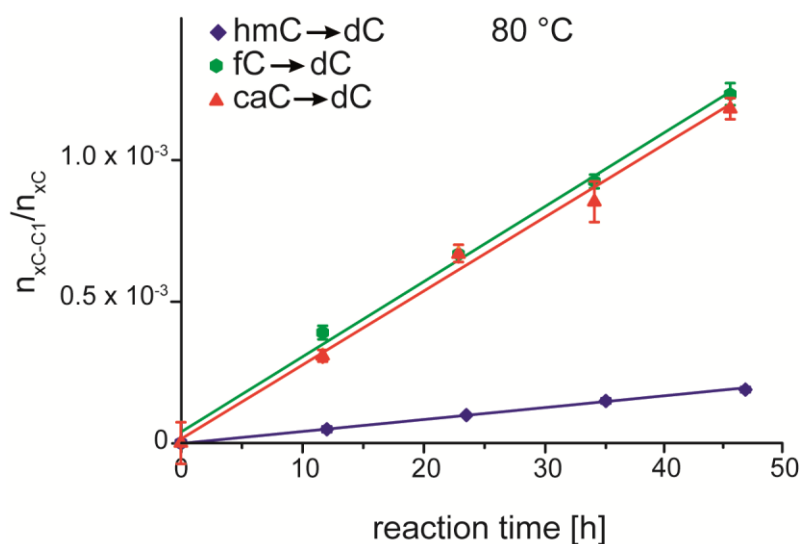
Figure S9. C-C-bond cleavage of hmdC (blue), fdC (green) and cadC (red) at 80 °C in the absence of the thiol/imidazole mixture. At 60 °C no C-C bond cleavage could be observed.

Table S90. Coefficients of determinations R^2 and linear equations of linear regression analyses of straights in Figure S9 determined by ORIGIN® 6.0.

reaction	Coefficient of determination R^2	linear equation
hmdC→dC	0.9998	$y = 4.21209 \times 10^{-6}\,x + 2.01274 \times 10^{-7}$
fdC→dC	0.9990	$y = 2.71274 \times 10^{-5}\,x + 1.78858 \times 10^{-6}$
cadC→dC	0.9962	$y = 2.47637 \times 10^{-5}\,x + 1.38867 \times 10^{-5}$
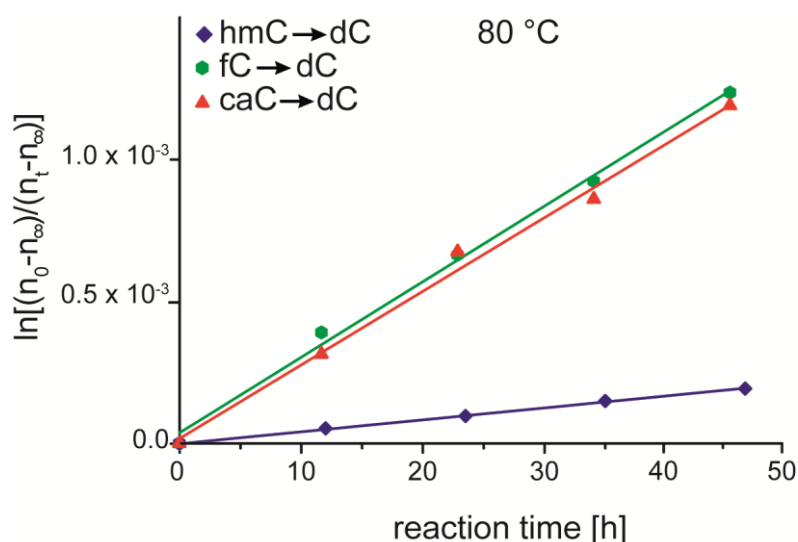
Figure S10. Determination of rate constants of the C-C-bond cleavage of hmdC (blue), fdC (green) and cadC (red) at 80 °C in the absence of the thiol/imidazole mixture. At 60 °C no C-C bond cleavage could be observed.

Table S91. Coefficients of determinations R^2 and linear equations of linear regression analyses of straights in Figure S10 determined by ORIGIN® 6.0.

reaction	Coefficient of determination R^2	linear equation
hmdC→dC	0.9995	$y = 4.17805 \times 10^{-6}\ x + 1.49031 \times 10^{-6}$
fdC→dC	0.9978	$y = 2.52342 \times 10^{-5}\ x + 3.55494 \times 10^{-5}$
cadC→dC	0.9965	$y = 2.45745 \times 10^{-5}\ x + 1.7007 \times 10^{-5}$

Stability of the modified nucleosides

For the kinetic analysis, the nucleosides dC, mdC, hmdC, fdC and cadC (10 μm) were incubated in a phosphate buffer at pH 7.4 for at 60 °C, 67.5 °C, 75 °C and 80 °C. At certain time points samples of each 10 μL were taken, spiked with a specific amount of labeled internal standards and analyzed by LC-MS/MS. Under these conditions, deamination, oxidation and C-C-bond cleavage reactions were observed (see main text). Since large amounts of the starting material were recovered, the glycosidic bonds of the nucleosides seemed to be quite stable, except in the case of fdC. For example, Fig. S11A shows the MS/MS-chromatogram of cadC before and after incubation at 80 °C for 48 h. No significant difference was observed. In the case of fdC (Fig. S11B) three additional peaks with the same MS/MS fragmentation pattern as fdC appeared after 48 h. These are likely the α/β-pyranose and α-furanose isomers of fdC. The formation of these compounds show the weakness of the glycosidic bond of fdC, which can be attributed to the comparatively electron poor nucleobase (-I effect of the aldehyde group). In contrast, cadC is deprotonated at pH 7.4 and therefore more electron rich (+I effect of CO_2^- group). The instability of fdC could be one major reason why the kinetic studies of the fdC to cadC

oxidation deviated from linearity.
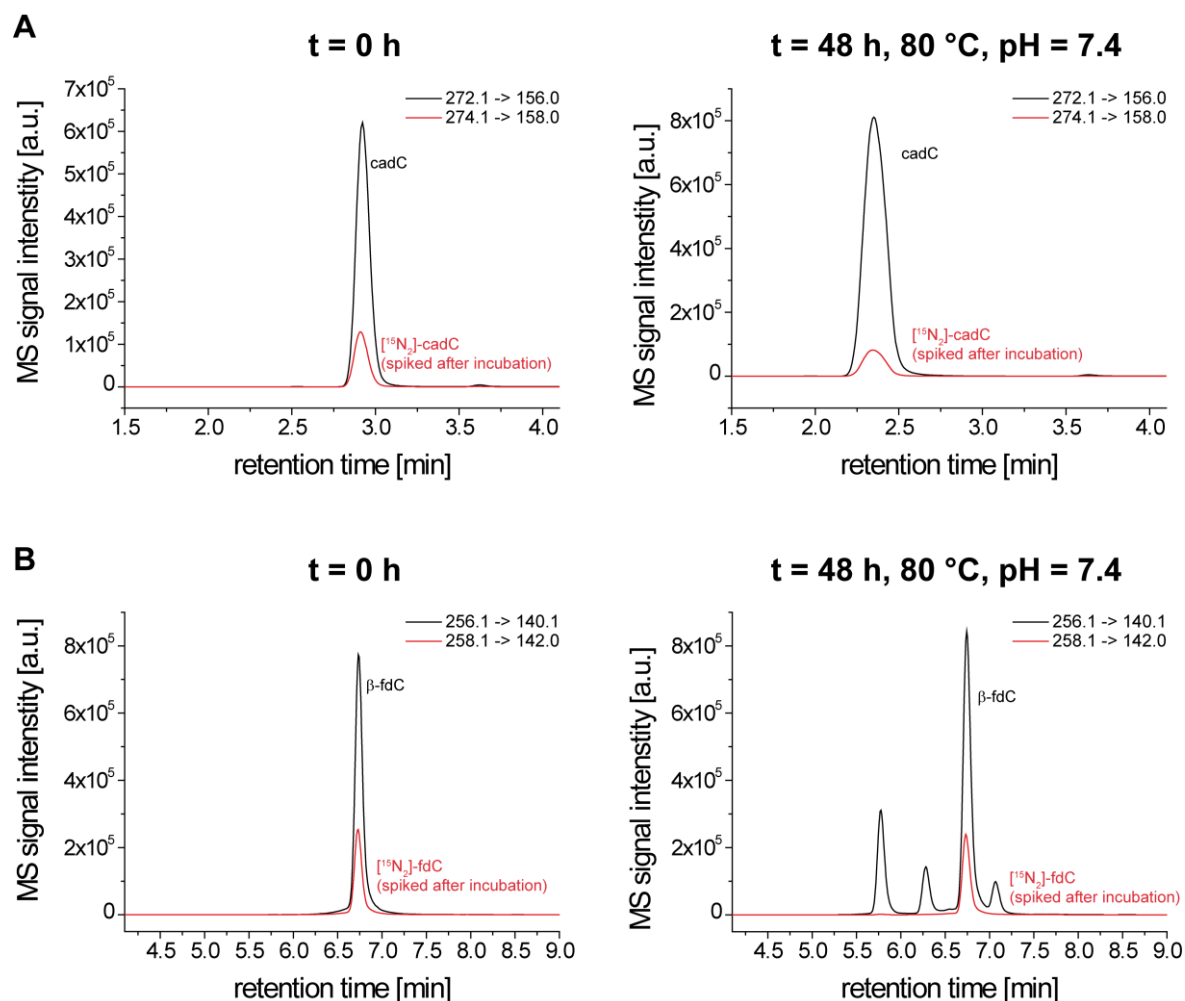
Figure S11. cadC (A) is stable upon heating in a pH 7.4 buffered solution (80 °C, 48 h), while fdC (B) is not. Depicted are the MS/MS chromatograms of cadC and fdC before (t=0 h) and after incubation (t = 48 h), respectively. The β-configurated isotopologues [^{15}N$_2$]-fdC and [^{15}N$_2$]-cadC were added after incubation. The corresponding MS/MS transitions are shown in the top right of each diagramm.

Quantum-chemical calculations

All calculations were carried out using the TURBOMOLE program package.[13] Structure optimization was performed at the DFT level using the B3LYP[14,15] functional and triple-zeta basis set.[16] The empirical dispersion correction, in combination with the Becke and Johnson damping,[17] as proposed by Grimme (DFT-D3),[18] was also taken into account. In order to ensure convergence to a minimum energy structure, the eigenvalues of the hessian were computed. For all structures, the corresponding hessian was found to have no negative eigenvalues. For the energy minimized structures, single point RI-MP2 energy calculations were performed, also employing the triple-zeta basis set.

All nucleobases were capped with a methyl group at the N(1) position. The β-mercaptoethanol used in the experiment was modeled as methyl sulfide for simplicity. The energy of a single proton, used to

for estimating the reaction energy during a protonation, was taken as the difference between the energy of water and H_3O^+. For the cadU (orotat and isoorotat), cadC, cytosine, uracil and both thiol-adducts, an explicit solvation is described by including five water molecules as shown in Fig. S12. In addition, the influence of the second and further solvation shells on the computed reaction energies, were estimated by performing B3LYP-D3 single point energy calculations with each molecule embedded in an implicit solvent cavity using the COSMO-model[19] and the dielectric constant of water ($\varepsilon=80$).

In tables S92 – S95, E_0 and E_0 +ZPE are the DFT absolute energies (in au) with and without zero point energy corrections, respectively. E_0-COSMO (in au) is the energy of the corresponding COSMO calculation using the same geometry, method and basis set. RI-MP2 energies are shown in the last columns.
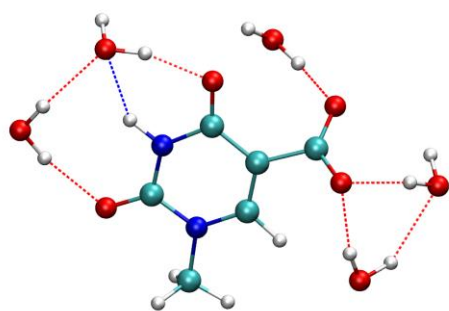
(a) Isoorotate (caU)
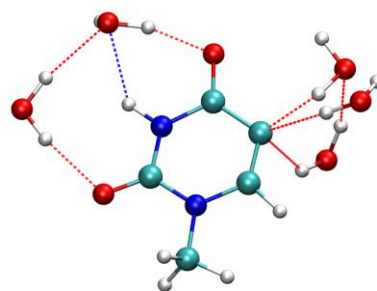
(b) Decarboxy-Isoorotat
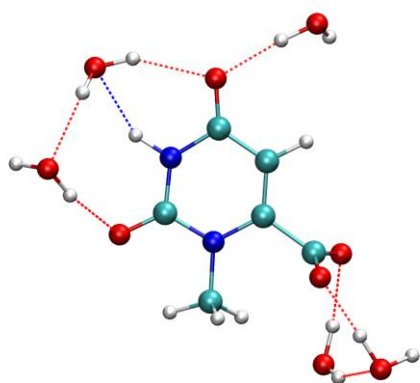
(c) Orotate
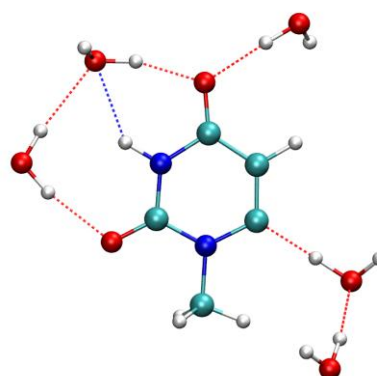
(d) Decarboxy-Orotate
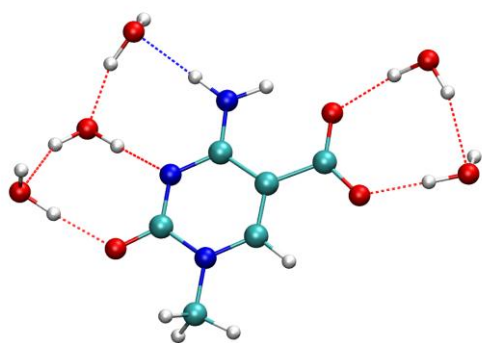
(e) Carboxycytosine (caC)
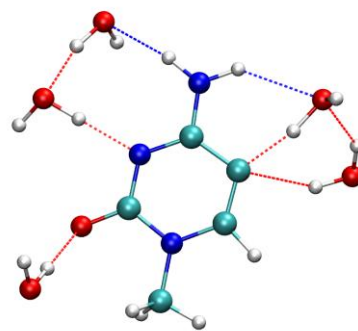
(f) Decarboxycytosine
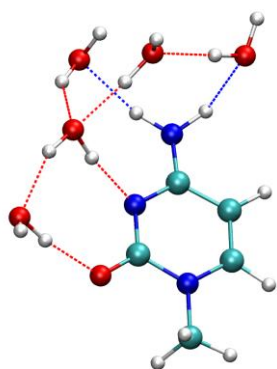
(j) Cytosine (C)
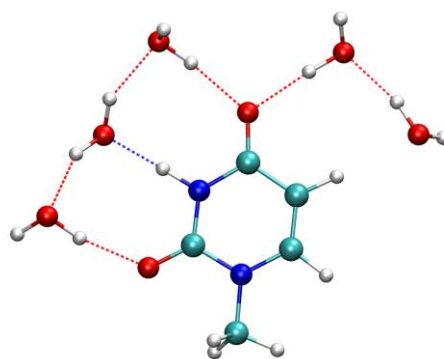
(i) Uracil (U)
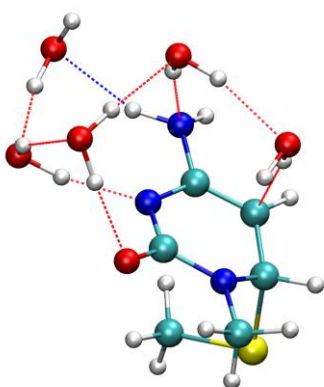
(h) Thio-decarboxycytosine (Thio-caC)
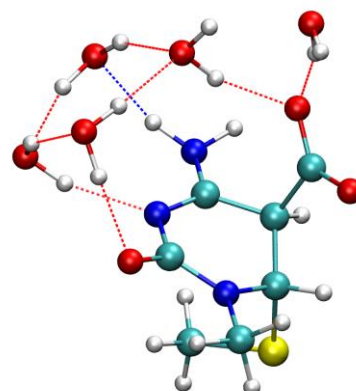
(g) Thio-carboxycytosine (Thio-caC)

Figure S12. Structures of carboxylated and decarboxylated uracil (orotate and isoorotate), cytosine and thiol-adduct, each solvated with five water molecules are shown. Solvated structures of uracil and cytosine are also shown for comparison.

Table S92. Reaction energies ΔE_1 and ΔE_2 computed for decarboxylation of orotat via formation of a carbanion intermediate, respectively, are shown. Following model reactions are used to estimate the reaction energies:

$\Delta E_1 = E \text{ (carbanion} + H_3O^+ + CO_2) - E(caU + H_3O^+)$

$\Delta E_2 = E \text{ (uracil} + H_2O + CO_2) - E(caU + H_3O^+)$

	B3LYP-D3 (au)			RI-MP2 (au)
	E_0	E_0+ZPE	E_0-COSMO	E_0
Orotate	−1024.398185	−1024.157178	−1024.489648	−1023.129545
Carbanion	−835.757867	−835.534580	−835.847069	−834.705704
Uracil	−836.291928	−836.053232	−836.328397	−835.281528
H$_3$O+	−76.705601	−76.671352	−76.849471	−76.617415
CO$_2$	−188.588980	−188.577284	−188.592480	−188.370702
H$_2$O	−76.430163	−76.408870	−76.441529	−76.342331
Reaction energy (kcal/mol)				
ΔE_1	32.2	28.4	31.4	33.3
ΔE_2	−130.0	−132.2	−14.6	−155.2

Table S93. Reaction energies ΔE_1 and ΔE_2 computed for decarboxylation of isoorotat via formation of a carbanion intermediate, respectively, are shown. Following model reactions are used to estimate the reaction energies:

$\Delta E_1 = E \text{ (carbanion} + H_3O^+ + CO_2) - E(caU + H_3O^+)$

$\Delta E_2 = E \text{ (uracil} + H_2O + CO_2) - E(caU + H_3O^+)$

	B3LYP-D3 (au)			RI-MP2 (au)
	E_0	E_0+ZPE	E_0-COSMO	E_0
CaU (Isoorotate)	−1024.391578	−1024.150278	−1024.488619	−1023.123822
Carbanion	−835.743469	−835.519207	−835.835385	−834.696480
Uracil	−836.291928	−836.053232	−836.328397	−835.281528
H$_3$O+	−76.705601	−76.671352	−76.849471	−76.617415
CO$_2$	−188.588980	−188.577284	−188.592480	−188.370702
H$_2$O	−76.430163	−76.408870	−76.441529	−76.342331
Reaction energy (kcal/mol)				
ΔE_1	37.1	33.7	38.1	35.5
ΔE_2	−134.1	−136.5	−15.2	−158.8

Table 94. Reaction energies E_1 and E_2 computed for decarboxylation of 5-carboxycytosine (caC) via formation of a carbanion intermediate, respectively, are shown. Following model reactions are used to estimate the reaction energies:

$\Delta E_1 = E \text{ (Carbanion} + H_3O^+ + CO_2) - E(caC + H_3O^+)$

$\Delta E_2 = E \text{ (cytosine} + H_2O + CO_2) - E(caC + H_3O^+)$

	B3LYP-D3 (au)			RI-MP2 (au)
	E_0	E_0+ZPE	E_0-COSMO	E_0
CaC	−1004.529201	−1004.274631	−1004.624476	−1003.271204
Carbanion	−815.859992	−815.624853	−815.950548	−814.824877
Cytosine	−816.452547	−816.200209	−816.484417	−815.412189
H$_3$O$^+$	−76.705601	−76.671352	−76.849471	−76.617415

CO_2	−188.588980	−188.577284	−188.592480	−188.370702
H_2O	−76.430163	−76.408870	−76.441529	−76.342331
Reaction energy (kcal/mol)				
ΔE_1	50.3	45.5	51.1	47.4
ΔE_2	−148.5	−150.7	−27.9	−148.3

Table 95. Reaction energies E_1, E_2 and E_3 computed for thiol addition, decarboxylation and thiol elinimation of 5-carboxycytosine (caC), respectively. Following model reactions are used to estimate the reaction energies:

$\Delta E_1 = E$ (Thiol-adduct-1 + H_3O^+) - E(caC + methylsulfide + H_3O^+)

$\Delta E_2 = E$ (Thio-adduct-2 + H_3O^+ + CO_2) - E(caC + methylsulfide + H_3O^+)

$\Delta E_3 = E$ (cytosine + methylsulfide + H_2O + CO_2) - E(caC + methylsulfide + H_3O^+)

	B3LYP-D3 (au)			RI-MP2 (au)
	E_0	E_0+ZPE	E_0-COSMO	E_0
CaC	−1004.529201	−1004.274631	−1004.624476	−1003.271204
Thiol adduct 1	−1443.200534	−1442.891614	−1443.288127	−1441.520751
Thiol adduct 2	−1254.576700	−1254.285948	−1254.661434	−1253.116367
Cytosine	−816.452547	−816.200209	−816.484417	−815.412189
H_3O^+	−76.705601	−76.671352	−76.849471	−76.617415
CO_2	−188.588980	−188.577284	−188.592480	−188.370702
H_2O	−76.430163	−76.408870	−76.441529	−76.342331
Methylsulfide	−438.658291	−438.612477	−438.663217	−438.224956
Reaction energy (kcal/mol)				
ΔE_1	−8.2	−2.8	−0.3	−15.4
ΔE_2	13.7	15.0	21.2	5.7
ΔE_3	−148.5	−150.7	−27.9	−148.3

References

(1)   Münzel, M.; Globisch, D.; Brückl, T.; Wagner, M.; Welzmiller, V.; Michalakis, S.; Müller, M.; Biel, M.; Carell, T. *Angew. Chem., Int. Ed.* **2010**, *49*, 5375.
(2)   Globisch, D.; Münzel, M.; Müller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. *PLoS One* **2010**, *5*, e15367, doi:10.1371/journal.pone.0015367.
(3)   Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671.
(4)   Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeier, C.; Truss, M.; Carell, T. *Angew. Chem., Int. Ed.* **2012**, *51*, 6516.
(5)   Hodge, R. P.; Brush, C. K.; Harris, C. M.; Harris, T. M. *J. Org. Chem.* **1991**, *56*, 1553.
(6)   Felczak, K.; Miazga, A.; Poznanski, J.; Bretner, M.; Kulikowski, T.; Dzik, J. M.; Golos, B.; Zielinski, Z.; Ciesla, J.; Rode, W. *J. Med. Chem.* **2000**, *43*, 4647.
(7)   Bischofberger, N. *Tetrahedron Lett.* **1987**, *28*, 2821.
(8)   Münzel , M.; Globisch , D.; Trindler , C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671.
(9)   Taghizadeh, K.; McFaline, J. L.; Pang, B.; Sullivan, M.; Dong, M.; Plummer, E.; Dedon, P. C. *Nat. Protoc.* **2008**, *3*, 1287.
(10)   Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. A.; Karaghiosoff, K.; Carell, T. *Chem.--Eur. J., 17*, 13782.
(11)   Cao, H.; Wang, Y. *J. Am. Soc. Mass. Spectrom.* **2006**, *17*, 1335.
(12)   Wang, J.; Yuan, B.; Guerrero, C.; Bahde, R.; Gupta, S.; Wang, Y. *Anal. Chem. (Washington, DC, U. S.)* **2011**, *83*, 2201.

(13) "TURBOMOLE V6.3 2011.

(14) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(15) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1998**, *37*, 785.

(16) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143.

(17) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2006**, *124*, 174104.

(18) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456.

(19) Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perkin Trans. 2* **1993**, *5*, 799.

8.5 Zusatzmaterialien zu Abschnitt 3.5

Cornelia G. Spruijt, Felix Gnerlich, Arne H. Smits, Toni Pfaffeneder, Pascal W. T. C. Jansen, Christina Bauer, Martin Münzel, Mirko Wagner, Markus Müller, Fariha Khan, H. Christian Eberl, Anneloes Mensinga, Arie B. Brinkman, Konstantin Lephikov, Udo Müller, Jörn Walter, Rolf Boelens, Hugo van Ingen, Heinrich Leonhardt, Thomas Carell und Michiel Vermeulen, *Cell* **2013**, *152*, 1146−1159. *Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives.*

EXTENDED EXPERIMENTAL PROCEDURES

SILAC Labeling of ES Cells

IB10 murine Embryonic stem cells were cultured feeder-free on gelatin coated dishes in medium consisting of 500 ml SILAC Dulbecco's Modified Eagle Medium without arginine, lysine and glutamine (PAA, E15-086), supplemented with 15% MESC serum substitute (Thermo Scientific), Glutamine, Penicillin/Streptomycin, 1x Non-essential amino acids, sodium pyruvate, 73 μg/ml L-Lysine (light/K^0 (Sigma, A6969), medium/K^4 (Sigma, 616192 or Silantes, 211103912) or heavy/K^8 (Sigma, 608041 or Silantes, 211603902)) and 29.4 μg/ml arginine (light/R^0 (Sigma, A6969), medium/R^6 (Sigma, 643440 or Silantes, 201203902) or heavy/R^{10} (Sigma, 608033 or Silantes, 201603902)), LIF (1000 U/ml), β-mercaptoethanol and 2i compounds (CHIR99021 and PD0325901, 3 and 1 μM respectively). Cells were cultured in SILAC medium until labeling efficiency exceeded 95% after which cells were expanded and harvested to generate nuclear extracts.

NPC Culturing

Neuronal progenitor cells were kindly provided by Dr. N. S. Outchkourov. They were cultured in medium consisting of NSA MEM (Euromed EVM0883LD), 1% glutamine, 1x N2 supplement, 10 ng/mL bFGF (RD systems 233-F3) and 10 ng/mL EGF (235-E9) on gelatin-coated dishes. Cells were detached from culture plates using accutase. Nuclear extracts were made as described below.

Mice Brain Nuclear Extracts

Nuclei from adult mouse brain were purified by centrifugation through a sucrose cushion following homogenization, modified from (Lavery and Schibler, 1993). Then nuclei were lysed as described below.

Nuclear Extract Preparation

This protocol is based essentially on Dignam et al. (Dignam et al., 1983). Briefly, cells were trypsinized and washed two times with PBS. Using a hypotonic buffer, the cells were swollen, after which the cells were lysed by dounce homogenizing in the presence of 0.15% NP40 and complete protease inhibitors. After centrifugation, the pellet consisting of nuclei was lysed by 90 min incubation in 2 volumes of nuclear lysis buffer (420 mM NaCl, 20 mM HEPES pH 7.9, 20% v/v glycerol, 2 mM MgCl$_2$, 0.2 mM EDTA, 0.1% NP40, complete protease inhibitor w/o EDTA (Roche) and 0.5 mM DTT). After centrifugation, the supernatant containing the soluble nuclear extract was aliquoted and snap frozen until further usage. Protein concentrations of the nuclear extracts were determined using the Biorad Protein assay.

DNA Synthesis

The synthesis of the oligonucleotides was performed on an ABI 394 DNA/RNA Synthesizer (Applied Biosystems) using typical reagent concentrations (activator: 0.25 M benzylthiotetrazole in MeCN (10 ppm H$_2$O), detritylation: 3% dichloroacetic acid in CH$_2$Cl$_2$, oxidation: 25 mM I$_2$ in MeCN/H$_2$O/2,6-lutidine (11/5/1), capping: Ac$_2$O/2,6-lutidine/MeCN (30 ppm H$_2$O) (20/30/50) and 20% N-methylimidazole in MeCN (10 ppm H$_2$O). The oligonucleotide syntheses were performed on 200 nmol low-volume polystyrene carriers using 0.1 M DNA CE-phosphoramidites: A (Bz-dA), C (Bz-dC), G (iBu-dG), T, mC (Bz-mC) obtained from *Glen Research* or *Link Technologies*. hmC, fC and caC phosphoramidites were synthesized according to literature (Münzel et al., 2010b) and incorporated into DNA using the standard protocol. Benzylthiotetrazole was prepared according to literature (Welz and Muller, 2002). The coupling times for the modified bases were increased to 3 min to ensure maximum coupling efficiency.

   The mC and the unmodified strands were treated with ethanolic ammonia for cleavage of the carrier and removal of the permanent protecting groups. hmC, fC and caC containing DNA was cleaved and deprotected using 0.4 M NaOH in MeOH/H$_2$O 4:1 for 18 hr at room temperature. After addition of 600 μl triethylammonium acetate (1 M) and centrifugation, the supernatant was concentrated to 30% of the original volume in a speedvac. Analysis and purification was performed on a Waters HPLC system (Waters Alliance 2695 with PDA 2996, preparative HPLC: 1525EF with 2482 UV detector) with VP 250/10 Nucleosil 100-7 C 18 columns from *Macherey Nagel* using a gradient of 0.1 M triethylamine/acetic acid in water and 80% acetonitrile. The quality of the strands was determined by MALDI-MS. The forward and reverse oligos were combined and annealed in 10mM Tris pH8; 50 mM NaCl and 1 mM EDTA. Biotin-14-ATP was used to fill in the TT-overhang using Klenow exo-, followed by purification of the DNA on sephadex-G50 columns.

DNA Pull-Downs

For each DNA pull-down, 10 μg of DNA (see Table S4) was immobilized on 75 μl of Dynabeads MyOne C1 (Invitrogen) by incubating for 1 hr at room temperature in a total volume of 350 μl of DNA binding buffer (1M NaCl, 10mM Tris-HCl pH8, 1mM EDTA pH 8 and 0.05% NP40. Coupling of the DNA to the beads was always verified by agarose gel electrophoresis. Beads containing immobilized DNA were then incubated with 400 μg of nuclear extract in a total volume of 600 μl of protein binding buffer (50mM Tris-HCl pH8, 150 mM NaCl, 1mM DTT, 0.25% NP40 and complete protease inhibitors (Roche, EDTA-free)) in the presence of 10 μg poly-dAdT for 2 hr at 4°C. Baits were then washed three times with 0.5 ml of protein binding buffer after which beads containing different DNA modifications and different SILAC labels were combined and loaded on 4%–12% NuPage gradient gels (Invitrogen) (for example, C-beads with light extract were combined with mC beads that were incubated with heavy extract; forward pull-down).

For the label-free analysis, three separate DNA pull-downs with every bait were performed and each of these was loaded on gel separately. For Western blot validation using endogenous antibodies, protein amounts were scaled down by a factor of four.

In-Gel Digestion
Samples were analyzed on 4%–12% precast NuPage gels (Invitrogen) and subsequently stained using colloidal blue staining (Invitrogen). Each lane was cut into 8-12 gel slices and each of these slices was subjected to in-gel trypsin digestion overnight. Tryptic peptides were desalted on Stage-tips (Rappsilber et al., 2003).

Mass Spectrometry
Peptides were separated on an EASY-nLC (Proxeon) connected online to an LTQ-Orbitrap-Velos mass spectrometer. Spectra were recorded in CID mode. A gradient of organic solvent (5%–30% acetonitrile) was applied (120 min) and the top 15 most abundant peptides were fragmented for MS/MS, using an exclusion list of 500 proteins for 45 s.

Data Analysis
Raw data were analyzed using Maxquant version 1.2.2.5 and the integrated Andromeda search engine against protein database ipi.MOUSE.v3.68. Using Perseus, data was filtered for contaminants, reverse hits, number of peptides (>1) and unique peptides (>0). Ratios were logarithmitized (log2) and groups (consisting of forward and reverse) were defined. Proteins were filtered to have at least 2 valid values in one of the groups and missing values were imputed based on a normal distribution (width = 0.2 and shift = 0), after which Significance B was calculated (Benj.Hoch.FDR = 0.05). Scatterplots were made using R. Proteins were defined to be significant when both forward and reverse significance $p < 0.05$ and minimal ratios were > 2 in both experiments. The H/L ratios shown in Figure 2A–C were calculated using the formula (log(forward ratio) – log(reverse ratio))/2.

Label-Free Quantification
LFQ values, based on the summed measured intensities of all tryptic peptides of a single protein, allows for comparing the relative abundance of a protein in different pull-downs. Changes in the LFQ intensity of a protein between pull-downs with different DNA modifications indicate preferential binding of that protein to one modification over another. Raw data were analyzed using Maxquant version 1.2.2.5 and protein database ipi.MOUSE.v3.68.fasta. Settings that were different from SILAC analyses were: multiplicity set at 1 and the options for 'label-free quantification' and 'match between runs' were selected Using Perseus, data were filtered for contaminants, reverse hits, number of peptides (>1) and unique peptides (>0). LFQ intensities were logarithmitized (log2). After defining each triplicate as a group, proteins were filtered to have at least 3 values in a single group, assuming that when a protein binds specifically to one modification, it may only be identified in the three pull-downs with that modification. The missing values were imputed using a normal distribution (with = 0.3, shift = 1.8). Groups were defined and the significant outliers were calculated using ANOVA (FDR = 0.025, $S_0 = 2$ for NPC and $S_0 = 0$ for brain). Correlation based clustering was done in R for the ANOVA-outliers only, using LFQ-values which had been normalized by row-mean-subtraction.

Purification of GFP-Fusion Proteins for EMSA
HEK293T cells were transfected with expression constructs encoding for GFP-Tdg or GFP-Dnmt1. 48 hr after transfection, cells were lysed 30 min on ice in Lysis-Buffer (50 mM NaH_2PO_4, 150 mM NaCl, 10 mM Imidazole, 0.5 mM EDTA, 0.5% Tween, 1 g/l DNaseI, 2 mM $MgCl_2$, 0.5 mM $CaCl_2$, 1 mM PMSF, 1x Protease-Inhibitor-Mix M (SERVA Electrophoresis GmbH)). The lysate was cleared by centrifugation (14 000 rpm, 10 min, 4°C) and incubation of the supernatant with equilibrated Ni-NTA beads (QIAGEN) in IP-buffer (50 mM NaH_2PO_4, 150 mM NaCl, 10 mM Imidazole, 0.5 mM EDTA, 0.05% Tween). After centrifugation (2200 rpm, 2 min) the supernatant was added to equilibrated GBP-Ni-NTA beads (Chromotek) in IP-buffer and rotated for 2 hr at 4°C. After washing three times with Washing-Buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, 0.1% Tween), the GFP-fusion proteins were eluted with 50 mM NaH_2PO_4, 150 mM NaCl, 250 mM Imidazole, 0.05% Tween. The elution buffer was exchanged to 20 mM TrisHCl, pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM DTT for EMSA reactions. The glycosylase activity of the purified Tdg was tested on T/G mismatch containing DNA (data not shown).

Electrophoretic Mobility Shift Assays of Fluorescent DNA Oligonucleotides with GFP-Fusion Proteins
GFP-Tdg and GFP-Dnmt1 at decreasing concentrations (200 nM, 150 nM, 100 nM, 50 nM, 25 nM 12.5 nM and 6.25 nM) were incubated for 30 min on ice with a 1:1 mixture of two distinctly labeled fluorescent 42mers (see Table S4, MWG-Eurofins, 250 nM each) containing a central CG site. The ATTO647N-labeled oligonucleotide contains only canonical bases whereas the ATTO550-labeled DNA bears different cytosine modifications (C, mC, hmC, fC and caC) or an abasic site at the CG position on both strands. Samples were run on a 6% non-denaturating polyacrylamide gel (pre-run 1 hr with 0.5x TBE) at 4°C. Oligonucleotide- and GFP-fluorescence was detected by the Typhoon Scanner (GE Healthcare). Quantifications were done with ImageJ.

DNA Purification and Analysis after NE Incubation
DNA pull-downs were performed as described above, but all amounts were scaled up 3 times. As a control, all baits were also incubated in buffer plus poly-dAdT without nuclear extract for 2 hr at 4°C. The beads were washed 3x using 1 ml of incubation buffer

and 1x using 1M NaCl, 10mM Tris-HCl pH8, 1mM EDTA and 0.05% NP40, to reduce contamination with DNA from the nuclear extracts. Beads were then resuspended in 200 uL incubation buffer and DNA was purified using phenol/chloroform extraction from the beads. The DNA-strands were finally dissolved in milliQ, enzymatically hydrolyzed to nucleosides and analyzed in triplicate (15 pmol each) by MALDI-MS or LC-MS/MS.

GFP Pull-Downs

HeLa wild-type cells and a BAC-GFP transgenic cell line (WDR76) were cultured in SILAC medium for eight cell doublings, after which cells were expanded and nuclear extracts were made. For each pull-down 20 μl of GFP-trap slurry (50% v/v; Chromotek) was washed and incubated for 90 min at 4°C with 1 mg of nuclear extract of (WT L, WT H, GFP L and GFP H) in a total volume of 400 μl incubation buffer (300 mM NaCl, 20 mM HEPES KOH pH 7.9, 20% v/v glycerol, 2 mM MgCl$_2$, 0.2 mM EDTA, 0.1% NP40, complete protease inhibitor w/o EDTA (Roche) and 0.5 mM DTT) in the presence of 2 μl ethidium bromide (10mg/ml, final concentration 50 μg/ml). Beads were then washed two times with this incubation buffer, twice with PBS + 0.5% NP40 and two times with PBS only. During the last wash, beads of light control and heavy GFP pull-down were mixed and vice versa. Bound proteins were then subjected to on-bead trypsin digestion (Hubner and Mann, 2011) and significant proteins were determined as described for the SILAC DNA pull-downs. For the GFP-Tdg pull-down, mESC were cultured in normal mESC medium and a transient transfection with the GFP-Tdg plasmid (15 μg/15cm dish) using PEI (ratio DNA:PEI = 1:3) was performed. GFP-Tdg was purified in a label-free method, thus 3 pull-downs were performed using GFP-trap beads and as a control the same extract was incubated in triplicate with control blocked agarose beads (Chromotek). For each pull-down 20 μl of bead slurry (50% v/v) was washed and incubated for 90 min at 4°C with 1 mg of the nuclear extract in a total volume of 400 μl incubation buffer (150 mM NaCl, 50 mM Tris-HCl pH8, 1mM DTT, 0.25% NP40 and complete protease inhibitor w/o EDTA (Roche)) to mimic the conditions of the DNA pull-downs as close as possible in the presence of 50 μg/ml of ethidium bromide. Beads were then washed two times with 0.5 ml of incubation buffer, twice with PBS + 0.5% NP40 and two times with PBS only, after which bound proteins were on-bead digested. The Tdg-GFP purification was analyzed using a permutation-based t test (FDR = 0.05 & S0 = 3) to determine significant interactors.

Recombinant Protein Expression/DNA Pull-Downs

Klf4(aa396-483), KDM2B(aa606-647), Cxxc5(aa234-293), MBD3(aa1-77) and Rfx5(aa85-173) were cloned into PRP256NB vector, containing a GST with a C-terminal multiple cloning site. Uhrf2 (aa416-626) GST fusion was kindly provided by Dr. Jiemin Wong. hMBD2b-GST was provided by Stefanie Bartels.

Protein expression was performed in *E. coli* BL21-DE3 Codon+ by growing them at 37°C until OD$_{600}$ of 0.5, after which expression was induced using 1mM IPTG and culturing for 3 additional hours at 25°C. Cells were lysed in 50 mM Tris-HCl pH 8.0/ 20% sucrose/ 1 mM EDTA/ 0.5 mM PMSF/ 1 mM DTT/ 1 μg/ml aprotinin using lysozyme and Triton X-100 and repeated freeze-thawing. Bacterial debris was removed by ultracentrifugation.

DNA pull-downs were performed using 2.5 μg DNA coupled to 16.75 μl MyOne beads and 5 μl of bacterial lysate/ nuclear extract in 250 μl total volume (50mM Tris-HCl pH8, 150 mM NaCl, 1mM DTT, 0.25% NP40 and complete protease inhibitors (Roche, EDTA-free)) in the presence of 2.5 μg polydAdT. After 3 times of washing with 0.5 ml of this buffer, beads were boiled in sample buffer. 5% of the input material and 100% of the bound material was loaded on gel for Western blot analyses.

Western Blot

Gels were blotted onto nitrocellulose membranes. Blots were blocked using 5%-skimmed milk in TBST. Used antibodies are: MouseαMBD3 (IBL, 3A3), GoatαMBD2 (Everest Biotech, EB07538), RabbitαRBBP5 (Bethyl, BL766), GoatαJun-C (SantaCruz), RabbitαDNMT1 (Abcam, ab13537), RabbitαCarf (Abcam, ab140519), RabbitαGST (Santa Cruz, SC-138), RabbitαGFP (home made), DonkeyαmouseHRP and DonkeyαRabbitHRP.

NMR-Spectroscopy-Based Interaction Study of Rfx5 and mC DNA

The winged-helix (WH) domain of human Rfx5 (residues 85-173, plus 18 additional residues at the N-terminus) was expressed as a GST-fusion in BL21-DE3 Codon+ bacterial strains at 25°C in M9 minimal medium with ^{15}NH$_4$Cl and/or ^{13}C-glucose. The protein was purified by binding to a Glutathione agarose (GA) column (Sigma) and eluted with 50 mM reduced glutathione (Sigma). After thrombin digestion, Rfx5-WH was purified over a Sephadex-75 (HiLoad 16/60) column in buffer A (50mM KPi pH 7, 100 mM KCl, 5 mM DTT, 0.5 mM PMSF and protease inhibitors). NMR samples used for backbone assignment contained ca. 0.3 mM WH domain in 90/10% H$_2$O/D$_2$O in buffer A. NMR spectra (HNCACB, CBCACONH, HNCA, and HNCO) were recorded at 298K on a 600 or 750 MHz Bruker Avance II spectrometer, processed using the NMRPipe package (Delaglio et al., 1995), and analyzed using CcpNmr Analysis (Vranken et al., 2005). Backbone assignments were obtained for 90 out of 106 residues in the Rfx5-WH construct.

Interaction study with mC DNA was done using an 18bp DNA fragment (see Table S4; (Biolegio)) carrying a single mC on each strand. Annealed DNA oligos were lyophilized and dissolved in buffer A to a stock concentration of 620 μM. The Rfx5-WH domain (103 μM) was titrated with mC DNA, and after each addition (11 points in total) the ^{1}H-^{15}N HSQC spectrum of Rfx5-WH as recorded (298K / 600 MHz Bruker Avance II). Since the DNA sequence used is not palindromic, the two mC may be inequivalent in their capability to bind Rfx5. At high DNA:Rfx ratios, several peaks appear split in two in a roughly 1:1 ratio, suggesting that although the Rfx5-WH domain senses the distinct DNA sequence context of the two mC sites, it recognizes both with similar affinities

(data not shown). Although a few residues showed non-linear titration profiles, most peak displacements were linear. For further analysis, the binding sites were treated as being independent, resulting an apparent dissociation constant for the Rfx5-WH – mC interaction.

Titration data were fitted using using MatLAB scripts (MATLAB version 7.13.0, The MathWorks Inc., 2011) using the fast-exchange assumption for residues with observed chemical shift perturbations between 10 and 30 Hz (fast-exchange regime; 15 residues) in a global fit. The error bars for the observed peak position was set to 1.2 Hz. The overall reduced chi-square for the fit was 2.17. The error in the fitted K_D was estimated using 1000 MonteCarlo simulations resulting in an average of 3.2 ± 0.9 μM. The range of acceptable fits was examined using F-statistics from a grid search, resulting in 95% probability limits of 10 nM < K_D < 16 μM.

A homology model of Rfx5-WH domain was constructed on the basis of the DNA-bound crystal structure of the Rfx1 winged helix domain (PDB-id: 1DP7; 35% sequence identity) using the SwissModel server (Schwede et al., 2003). The model was validated against the predicted backbone dihedral angles from the observed backbone chemical shifts using TALOS+ (Shen et al., 2009). The model of mC bound to the putative binding pocket was constructed in PyMol by superimposing the mC DNA from the UHRF1-mDNA crystal structure (PDB-id 3CLZ) onto the Rfx1-bound DNA, such that the binding pocket and mC are aligned. To achieve a proper fit, the mC base was set to a *syn*-conformation. The side chains orientations of K110 and Y161 were adjusted manually to minimize clashes.

In Silico Analysis of Klf4 ChIP-Seq Profile and Bisulfite Sequencing Data in mESCs Cells and NPCs

Klf4 binding data (ChIP-seq) was taken from (Chen et al., 2008) (GSM288354), and DNA methylation data (whole-genome bisulfite sequencing) was taken from (Stadler et al., 2012) (GSE30202). Annotated Klf4 peak centers (mESC) were extended with 50 bp on both sides to obtain 100-bp Klf4 binding regions. The mean CpG methylation of each 100-bp region was calculated for mESCs and NPCs and plotted as a scatterplot (Suppl. Figure S1B). For each quadrant of this scatterplot, the genomic distribution of the 100-bp Klf4 binding regions was calculated and plotted as a pie chart (Suppl. Figure S1C). Promoters were defined as -/+ 1 kb upstream and downstream from transcription start sites of the RefSeq mm9 annotation. The DNA sequences of the 100-bp Klf4 binding regions were used to search for the GGCGTG motif, and the CpG methylation within these motifs was calculated. The obtained distribution was plotted as a histogram (Suppl. Figure S1D). Analyses were done using Python, Perl and R.

iBAQ

iBAQ was performed essentially as described in (Schwanhäusser et al., 2011). 3.3 μg of UPS2 standard (Sigma) was added to 10 μg of nuclear extract, which was digested using the FASP protocol (Wiśniewski et al., 2009). In addition, 100 μg of NE was digested using FASP after which the peptides were separated into 8 fractions using SAX. Each of these samples was measured during a 4 hr gradient of LC-MSMS. A linear fit was made for the known amounts of the UPS2 standard and the measured iBAQ intensities in the 10 μg sample. Using this curve, iBAQ values of all other identified proteins in the 10 μg sample were converted to amounts. A linear fit was again made using these amounts and the iBAQ values in the eight SAX fractions, which were used to extrapolate absolute protein amounts of all identified proteins in these samples.

Cell Culture and Transfection Experiments

The mammalian GFP-Tet1cd expression vector was generated by PCR amplification of mouse (E14) cDNA encoding the catalytic domain of Tet1 (amino acids 1365 to 2057) and N-terminal GFP fusion. HEK293T cells were grown at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM, *Invitrogen 41966-029*) supplemented with 10% fetal bovine serum and 1% Penicillin-Streptomycin. Cells were passaged at 80% confluency. All transfections were performed using the *jetPRIME* system (*PEQLAB Biotechnologie GmbH*) according to the manufacturer's instructions. HEK293T cells were seeded 24 hr prior to transfection at a density of 3×10^6 cells per 75 cm^2-flask and incubated in 10 ml of medium at 37°C and 5% CO_2 for 24 hr. Cotransfection of GFP-Tet1cd plasmid (6 μg) either with mouse Uhrf2-GFP plasmid DNA (6 μg) (Pichler et al., 2011) or 6 μg of pCMV6-Cdk5Rap1-v2 (*Origene RG216600*) as an unrelated control was carried out in a 75 cm^2 flask containing 10 ml of fresh medium. The transfection solution (500 μl of *jetPRIME* buffer, 12 μg of plasmid DNA and 24 μl of *jetPRIME* reagent) was added to the medium and the cells were incubated at 37°C and 5% CO_2 for 48 hr. After removal of the medium the cells were washed once with PBS and then lysed for DNA extraction according to (Münzel et al., 2010). The DNA was enzymatically digested to the nucleosides and subsequently analyzed by LC-ESI-MS/MS.

LC-MS/MS Analysis of Genomic DNA and Synthetic DNA

The following LC-MS/MS method for the quantification of DNA-nucleosides is based on a further development of our precise and sensitive isotope dilution method ((Pfaffeneder et al., 2011) and manuscript in preparation). In the following we shortly summarize the parameters of the method. Genomic or synthetic DNA was enzymatically digested to the nucleoside level. A specific amount of internal standards with a stable isotope label were spiked to the digestion mixture for precise quantification. The following labeled nucleosides were used as internal standards: [$^{15}N_2$]-dC, [D_3]-mC, [D_2,$^{15}N_2$]-hmC, [$^{15}N_2$]-fC, [$^{15}N_2$]-caC and [D_3]-dT. In case of genomic DNA the dC- or dG-content was determined by LC-UV-Detection.

LC-MS/MS analysis was performed on an Agilent 6490 triple quadrupol mass spectrometer coupled to an Agilent 1290 UHPLC system. The general source-dependent parameters were as follows: Gas Temp 50°C, Gas Flow 15 L/min, Nebulizer 30 psi, Sheath

Gas Heater 300°C, Sheath Gas Flow 11 L/min, Capillary Voltage 2500 V and Nozzle Voltage 500 V. For compound-dependent parameters used for genomic DNA see Table S5, for compound-dependent parameters used for synthetic DNA see Table S6. The transitions of the nucleosides were analyzed in the positive ion selected reaction monitoring mode (SRM) operating MS1 and MS2 under unit mass resolution conditions.

For the analysis a C8 column from Agilent was used (1.8 μm, 2.1 mm x 150 mm). The compounds were separated by a gradient using water and acetonitril with 0.0075% formic acid. The column temperature was maintained at 30°C. The flow rate was 400 μl min^{-1}, and the injection volume amounted to 29 μL. The effluent up to 1.5 min (total run time of 12 min) was diverted to waste by a Valco valve in order to protect the mass spectrometer.

Validation of Quantification Method for Genomic DNA Modifications

In accordance with the FDA guidance for bioanalytical method validation, linearity, precision, and accuracy (i.e., recovery determined from spiked matrix samples) of the established method were investigated. Validation for the established LC-MS/MS quantification method was based on five different series (i.e., calibration functions and quality control samples) accomplished on different days. Calibration standards were analyzed at least in triplicates. Quality control samples to evaluate accuracy, intra- and inter-batch (see intra- and inter-assay) precision were determined using a biological sample with internal standards. Furthermore, each validation experiment was complemented by matrix blanks (analyzed in triplicates) to ensure selectivity and specificity of the method. Additionally, acceptable accuracy (80%–120%) as well as precision (<20% RSD) was required. Linear regression was applied to obtain calibration curves. Therefore, the peak area ratio (y) of the unlabeled nucleoside to the internal standard versus the concentration ratio of the unlabeled nucleoside to the internal standard (x) was plotted. Calibration functions were calculated without weighting. Long-term stability of aqueous solutions of the labeled and unlabeled nucleosides at a storage temperature of −20°C was investigated over two months including several freeze and thaw cycles by analyzing the MS/MS-responses with each batch. Short-term stability at room temperature was studied in overnight experiments. In this process, the results of quantification by LC-ESI-MS/MS directly after preparing the samples were compared with those obtained from samples kept overnight at room temperature.

SUPPLEMENTAL REFERENCES

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475–1489.

Lavery, D.J., and Schibler, U. (1993). Circadian transcription of the cholesterol 7 alpha hydroxylase gene may involve the liver-enriched bZIP protein DBP. Genes Dev. *7*, 1871–1884.

Münzel, M., Globisch, D., Trindler, C., and Carell, T. (2010b). Efficient synthesis of 5-hydroxymethylcytosine containing DNA. Org. Lett. *12*, 5671–5673.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. *31*, 3381–3385.

Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR *44*, 213–223.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2012). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687–696.

Welz, R., and Muller, S. (2002). 5-(benzylmercapto)-1H-tetrazole as activator for 2 '-O-TBDMS phosphoramidite building blocks in RNA synthesis. Tetrahedron Lett. *43*, 795–797.

Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362.
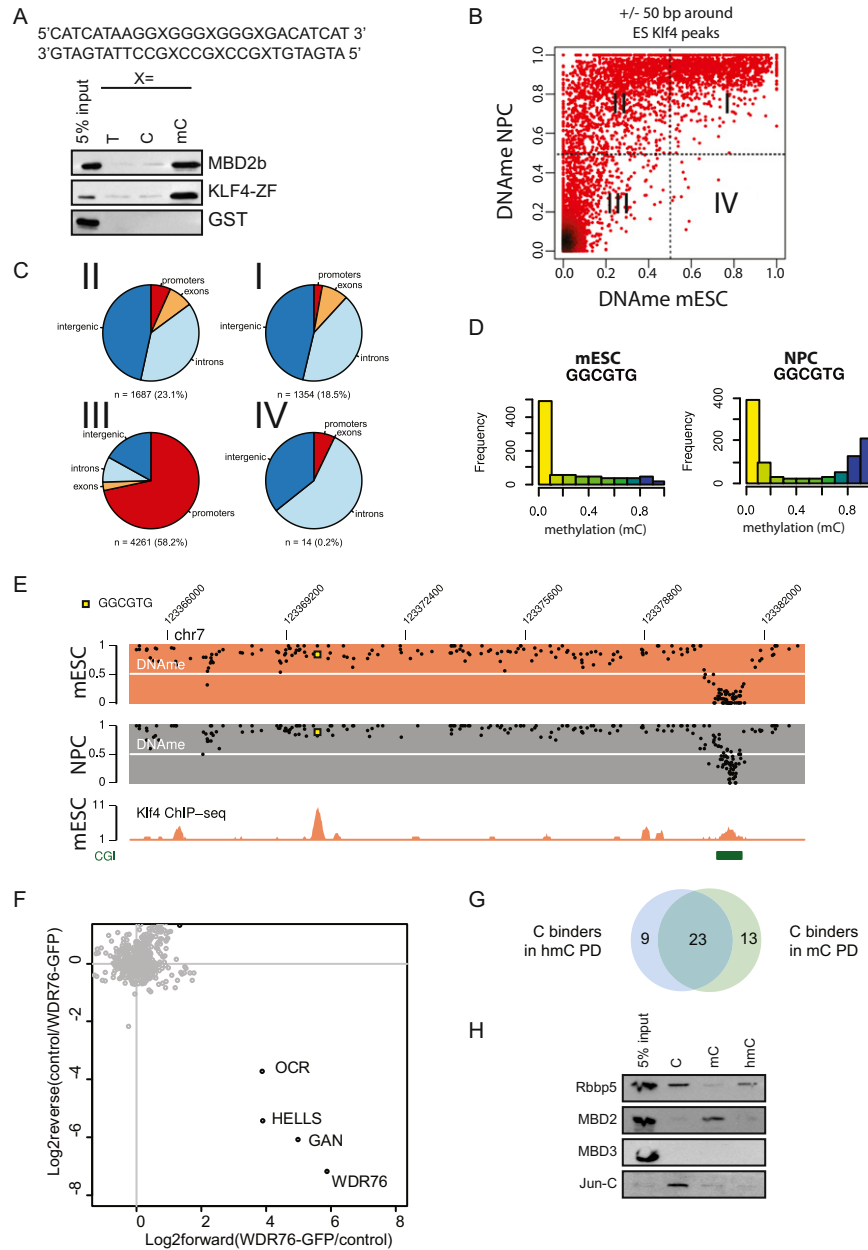
Figure S1. Genome-wide Localization of Klf4 Partially Correlates with DNA Methylation, Related to Figure 1

(A) DNA pull-downs with recombinant GST-fusion proteins of DNA binding domains and Western blotting analysis.

(B) DNA methylation of Klf4 sites in ES cells and NPCs. Whole-genome bisulfite sequencing was used to determine DNA methylation within a window of $+/- 50$ bp around Klf4 peak centers. Darker coloring indicates high density of datapoints.

(C) Pie charts showing the genomic distribution of Klf4 sites as presented in the different quadrants of (B).

(D) Distribution of DNA methylation specifically within the GGCGTG sequence present underneath Klf4 sites.

(E) Example of DNA methylation profiles and Klf4 binding (ChIP-seq), showing binding of Klf4 to both methylated and unmethylated sites. Yellow squares indicate the presence of the GGCGTG sequence underneath Klf4 sites.

(F) SILAC-based GFP-purification from HeLa cells stably expressing WDR76-GFP. Significant interactors are indicated in black (high forward WDR76-GFP/control ratio, low reverse control/WDR76-GFP ratio).

(G) Venn diagram showing the overlap of C-specific readers in the mC and hmC DNA pull-downs from mESC nuclear extracts.

(H) Validation of C and mC specific binders by DNA pull-downs in HeLa nuclear extract and Western blotting for the endogenous proteins.
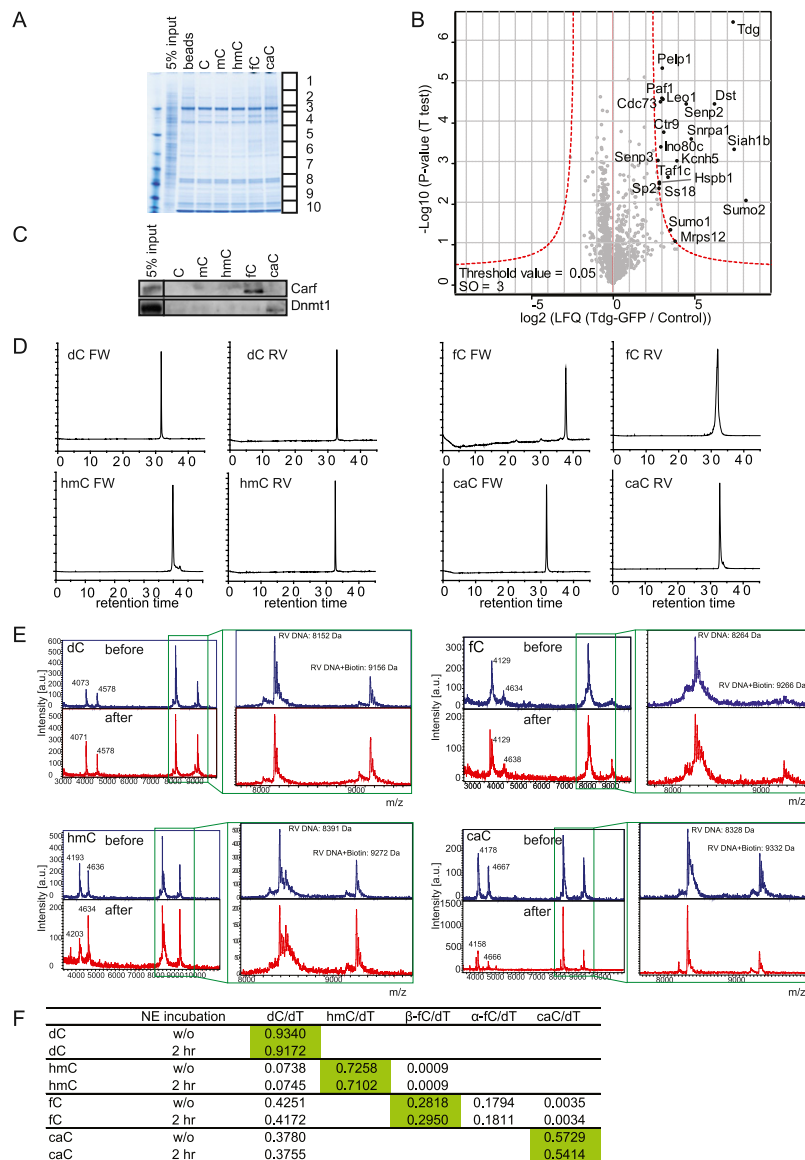
Figure S2. Identification of TDG Interactors, Western Blot Verification of fC and caC Interactors, and Validation of Bait DNA Quality, Related to Figure 2

(A) The indicated immobilized DNA baits were incubated with mouse nuclear extract. Following washes, bound proteins were analyzed by colloidal blue staining. Note that the elution profile of all these baits looks similar, indicating that specific interactors are masked by a large number of high abundant background binders.

(B) Volcano plot of a label-free GFP-Tdg pull-down in mESC nuclear extract. Significant interactors of GFP-Tdg are identified by permutation-based t test (FDR = 0.05 & S0 = 3). The LFQ intensity of the GFP pull-down over the control is plotted against the –Log10 (p-value). The red line indicates the permutation-based FDR. Also see Table SI.

(C) Western blot validation of the fC-specific binding of Carf and caC-specific binding of Dnmt1 in mESC nuclear extract. A single empty lane was removed from the blot.

(D) HPL-Chromatograms of the purified FW and RV DNA obtained from solid phase DNA synthesis showing the purity of the employed strands.

(E) The mass spectra of the DNA before (blue) and after (red) NE incubation as determined by MALDI MS showing the expected m/z before and after NE incubation. Major alterations of the DNA like degradation or strand breaks can be excluded.

(F) Synthetic DNA-strands which were used for DNA pull-downs were compared without (w/o) and with nuclear extract (NE) treatment (2 hr, 4°C) to proof the stability of the indicated modifications. The quantification of the nucleoside content was carried out by LC-MS/MS. For this, the DNA was digested to the nucleoside level and spiked with a specific amount of the following internal standards for precise quantification: $[^{15}N_2]$-dC, $[D_2,^{15}N_2]$-hmC, $[^{15}N_2]$-fC, $[^{15}N_2]$-caC and $[D_3]$-dT. The absolute amount (pmol) of each nucleoside was calculated by calibration curves (not shown). Depicted are ratios of the modified nucleoside (pmol) to deoxy-thymidine (dT; pmol), which were obtained from three independent measurements. The relative standard deviation was between 0.3%–6.2%. No or only marginal loss of the modified nucleosides was observed.

A

mESC

GO biological process		-log(Benj.Hoch.FDR)
C	RNA biosynthetic process	7.61
	cellular biosynthetic process	6.65
	biosynthetic process	6.48
	DNA recombination	5.14
	DNA metabolic process	5.06
	nucleobase-containing compound metabolic process	4.49
	transcription, DNA-dependent	4.48
	nucleic acid metabolic process	4.42
	cellular macromolecule biosynthetic process	4.30
	macromolecule biosynthetic process	4.29
	cellular nitrogen compound metabolic process	4.28
	nitrogen compound metabolic process	4.21
	regulation of macromolecule biosynthetic process	3.85
	regulation of cellular macromolecule biosynthetic process	3.85
	transcription initiation, DNA-dependent	3.82
	regulation of transcription, DNA-dependent	3.80
	regulation of cellular biosynthetic process	3.77
	regulation of biosynthetic process	3.75
	DNA repair	3.71
	regulation of gene expression	3.69
	regulation of nucleobase-containing compound metabolic process	3.53
	regulation of nitrogen compound metabolic process	3.53
	regulation of RNA metabolic process	3.33
	regulation of primary metabolic process	3.32
	regulation of macromolecule metabolic process	3.31
	response to DNA damage stimulus	3.26
	regulation of cellular metabolic process	3.12
	regulation of metabolic process	3.05
	cellular response to stress	2.97
	RNA metabolic process	2.72
	cellular macromolecule metabolic process	2.58
	response to stress	2.57
	macromolecule metabolic process	2.37
	eye morphogenesis	2.05
	cellular metabolic process	2.04
	primary metabolic process	1.86
	metabolic process	1.72
mC	transcription, DNA-dependent	2.84
	RNA biosynthetic process	2.81
	regulation of cellular biosynthetic process	1.76
	macromolecule biosynthetic process	1.76
	one-carbon metabolic process	1.74
	regulation of biosynthetic process	1.73
	regulation of macromolecule biosynthetic process	1.73
	regulation of nitrogen compound metabolic process	1.72
	regulation of RNA metabolic process	1.71
	nucleic acid metabolic process	1.70
	regulation of macromolecule metabolic process	1.70
fC	DNA repair	2.71
	response to DNA damage stimulus	2.60
	DNA metabolic process	2.59
	cellular response to stress	1.79

Cellular Compartment		-log(Benj.Hoch.FDR)
C	transcription factor TFIID complex	2.17
	TFIID complex	7.65
	MLL1 complex	6.06
	histone methyltransferase complex	4.20
	methyltransferase complex	4.37
	transcription factor complex	3.24
	nucleoplasm part	1.73

Domains		-log(Benj.Hoch.FDR)
C	BRLZ	7.40
	bZIP_1	5.57
mC	ZnF_C2H2	2.45
	MBD	2.24

NPC

GO biological process		-log(Benj.Hoch.FDR)
C	regulation of nitrogen compound metabolic process	5.04
	regulation of biosynthetic process	5.03
	regulation of nucleobase-containing compound metabolic process	5.01
	regulation of RNA metabolic process	5.01
	transcription, DNA-dependent	5.00
	regulation of cellular biosynthetic process	4.94
	RNA biosynthetic process	4.93
	regulation of macromolecule biosynthetic process	4.90
	regulation of cellular macromolecule biosynthetic process	4.88
	regulation of transcription, DNA-dependent	4.85
	regulation of gene expression	4.81
	cell differentiation	2.60
	positive regulation of transcription from RNA polymerase II promoter	2.51
	cellular response to metal ion	2.49
	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	2.48
	cellular response to inorganic substance	2.47
	positive regulation of biosynthetic process	2.46
	cellular response to chemical stimulus	2.45
	cellular response to calcium ion	2.45
	regulation of transcription from RNA polymerase II promoter	2.44
	gland development	2.39
	cellular developmental process	2.37
	positive regulation of RNA metabolic process	2.35
	adipose tissue development	2.33
	positive regulation of metabolic process	2.26
	positive regulation of nucleobase-containing compound metabolic process	2.21
	positive regulation of cellular biosynthetic process	2.16
	positive regulation of nitrogen compound metabolic process	2.14
	positive regulation of transcription, DNA-dependent	1.99
	positive regulation of macromolecule metabolic process	1.98
	positive regulation of gene expression	1.92
	response to calcium ion	1.88
	positive regulation of macromolecule biosynthetic process	1.85
	regulation of developmental process	1.83
	response to inorganic substance	1.79
	anatomical structure development	1.78
	regulation of cell differentiation	1.74
	response to metal ion	1.73
mC	positive regulation of macromolecule biosynthetic process	2.25
	positive regulation of cellular biosynthetic process	2.20
	positive regulation of biosynthetic process	2.17
	positive regulation of RNA metabolic process	2.14
	regulation of transcription from RNA polymerase II promoter	2.14
	positive regulation of gene expression	2.11
	positive regulation of cellular metabolic process	2.10
	dorsal spinal cord development	2.07
	positive regulation of metabolic process	2.04
	positive regulation of nucleobase-containing compound metabolic process	2.00
	positive regulation of cellular process	1.95
	positive regulation of nitrogen compound metabolic process	1.94
	positive regulation of biological process	1.86
hmC	DNA metabolic process	4.97
	DNA repair	3.91
	response to DNA damage stimulus	3.61
	base-excision repair	3.32
	cellular response to stress	2.96

Cellular Compartment		-log(Benj.Hoch.FDR)
C	histone methyltransferase complex	2.02
	intracellular membrane-bounded organelle	1.86

Domains		-log(Benj.Hoch.FDR)
C	bZIP_1	14.42
	zf-CXXC	7.91
	MBT	1.97
	Jun	1.77
mC	RFX_DNA_binding	1.98
	MBD	1.81
	Homeobox	1.80

Brain

GO biological process		-log(Benj.Hoch.FDR)
C	transcription, DNA-dependent	5.32
	RNA biosynthetic process	5.31
	regulation of biosynthetic process	5.24
	regulation of transcription, DNA-dependent	5.18
	regulation of macromolecule biosynthetic process	5.02
	regulation of nucleobase-containing compound metabolic process	5.00
	regulation of nitrogen compound metabolic process	4.99
	regulation of cellular macromolecule biosynthetic process	4.99
	regulation of gene expression	4.95
	regulation of cellular biosynthetic process	4.95
	regulation of RNA metabolic process	4.94
	regulation of primary metabolic process	4.70
	regulation of macromolecule metabolic process	4.65
	regulation of metabolic process	4.33
	DNA recombination	4.18
	regulation of cellular metabolic process	4.15
	cellular macromolecule biosynthetic process	4.08
	biosynthetic process	3.59
	response to DNA damage stimulus	2.75
	cellular response to stimulus	2.61
	cellular response to stress	2.58
	regulation of transcription from RNA polymerase II promoter	2.58
	DNA repair	2.47
	anatomical structure morphogenesis	2.46
	cellular response to calcium ion	2.15
	histone H4 acetylation	2.13
	eye morphogenesis	2.06
	response to calcium ion	2.03
	cellular response to metal ion	2.01
	cellular response to inorganic substance	2.00
	response to metal ion	1.71
	DNA metabolic process	1.70
hmC	DNA metabolic process	4.57
	DNA replication	3.54
	base-excision repair	1.78

Cellular Compartment		-log(Benj.Hoch.FDR)
C	histone methyltransferase complex	5.42
hmC	Brd4-Rfc complex	6.11
	DNA replication factor C complex	3.93

Domains		-log(Benj.Hoch.FDR)
C	bZIP_1	8.60
	Jun	1.93
mC	MBD	2.70
hmC	Rep_fac_C	3.61
	AAA	2.11

B

Venn diagrams:

mESC (69 interactors): hmC 6, 3, 1, mC 16, 43; C

NPC (190 interactors): hmC 23, 7, 65, 15, 0, 23, mC 57; C

Brain (107 interactors): hmC 31, 4, 7, 0, 16, 43, mC 6; C

C readers (137 proteins): mESC 12, 8, NPC 64, 13, 12, 11, 17, Brain

mC readers (126 proteins): Brain 21, 3, NPC 83, 1, 4, 9, 5, mESC

hmC readers (85 proteins): NPC 37, 2, Brain 35, 2, 0, 1, 6, 3, mESC

Figure S3. Modification and Cell-Type-Specific GO Term Enrichment Analysis, Related to Figures 1, 2, 3, and 4

(A) Shows GO term enrichment and enriched domains for the different baits (C, mC, hmC, fC and caC) in mESC, NPCs and adult mouse brain.

(B) Venn diagrams showing the overlap between C, mC and hmC readers within each cell type and the overlap between C, mC and hmC readers between mESCs, NPCs and adult mouse brain.

Figure S4. iBAQ Analyses of mESC, NPC, and Adult Mouse Brain Nuclear Extracts, Related to Figure 5

(A) Standard and linear regression curves for the iBAQ of protein abundance in the different nuclear extracts that were used for the DNA pull-downs.

(B) Correlation based clustering of proteins that show at least a 10-fold change in protein levels. Yellow is low abundance, red is high.

(C) GO term enrichment for mESC (indicated in blue in fig S4B), NPC (indicated in red in fig S4B) and adult mouse brain (indicated in green in fig S4) specific proteins.

Table S4. DNA Sequences Used in the Experiments, Related to Figures 1, 2, 3, 4, S1, and S2

Experiment	Forward and Reverse	X=
Mass spec DNA pull-downs, western blot validations	AAG.ATG.ATG.AXG.AXG.AXG.AXG.ATG.ATG TTC.ATC.ATX.GTX.GTX.GTX.GTC.ATC.ATC	C, mC, hmC, fC or caC
Klf4 validation	TTCATCATAAGGXGGGXGGGXGACATCAT ATGATGTXGCCXGCCXGCCTTATGATG	T, C or mC
EMSA	GGATGATGACTCTTCTGGTCXGGATGGTAGTTAAGTGTTGAG CCTACTACTGAGAAGACCAGGXCTACCATCAATTCACAACTC	C, mC, hmC, fC, caC or Abasic
RFX5 NMR	CCTGATGAXGACGTACCG CGGTACGTXGTCATCAGG	mC

Table S5. Compound-Dependent LC-MS/MS Parameters Used in the Analysis of the Nucleoside Contents after Enzymatic Digestion of Genomic DNA, Related to Figure 6

Compound	Precursor Ion (MS1)	Product Ion (MS2)	Dwell time [ms]	CE (V)	Cell Acc (V)
$[^{15}N_2]$-caC	274.08	158.03	90	5	4
caC	272.09	156.04	90	5	4
$[^{15}N_2]$-fC	258.09	142.04	30	5	4
fC	256.09	140.05	30	5	4
$[^{15}N_2,D_2]$-hmC	262.12	146.07	40	27	1
hmC	258.11	142.06	40	27	1
$[D_3]$-mC	245.13	129.09	30	60	1
mC	242.11	126.07	30	60	1
dC	228.1	112.05	1	1	0
dG	268.1	152.06	1	1	0
dT	243.1	127.05	1	3	1
dA	252.11	136.06	1	50	0

Table S6. Compound-Dependent LC-MS/MS Parameters Used in the Analysis of the Nucleoside Contents after Enzymatic Digestion of Synthetic DNA, Related to Figure S2

Compound	Precursor Ion (MS1)	Product Ion (MS2)	Dwell time [ms]	CE (V)	Cell Acc (V)
$[^{15}N_2]$-caC	274.08	158.03	50	2	5
caC	272.09	156.04	50	2	5
$[^{15}N_2]$-fC	258.09	142.04	20	2	5
fC	256.09	140.05	20	2	5
$[^{15}N_2,D_2]$-hmC	262.12	146.07	50	1	1
hmC	258.11	142.06	50	1	1
$[D_3]$-mC	245.13	129.09	50	60	1
mC	242.11	126.07	50	60	1
$[^{15}N_2]$-dC	230.1	114.1	80	2	5
dC	228.1	112.05	80	2	5
$[D_3]$-dT	246.12	130.07	20	3	3
dT	243.1	127.05	20	3	3

8.6 Zusatzmaterialien zu Abschnitt 3.6

Toni Pfaffeneder, Fabio Spada, Mirko Wagner, Caterina Brandmayr, Silvia K. Laube, David Eisen, Matthias Truss, Jessica Steinbacher, Benjamin Hackner, Olga Kotljarova, David Schuermann, Stylianos Michalakis, Olesea Kosmatchev, Stefan Schiesser, Barbara Steigenberger, Nada Raddaoui, Gengo Kashiwazaki, Udo Müller, Cornelia G. Spruijt, Michiel Vermeulen, Heinrich Leonhardt, Primo Schär, Markus Müller, Thomas Carell, *Nat. Chem. Biol.* **2014**, *10, 574–581. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.*

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–$^{15}N_2$)-cytidine ([$^{15}N_2$,D_2] hmC), 5-formyl-2′-deoxycytidine (fC), 5-formyl-2′-deoxy-(N^1,N^3–$^{15}N_2$)-cytidine ([$^{15}N_2$]fC), 5-carboxy-2′-deoxycytidine (caC), 5-carboxy-2′-deoxy-(N^1,N^3–$^{15}N_2$)-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–$^{15}N_2$)-uridine ([$^{15}N_2$]fU) were synthesized according to earlier published work[19,21,42]. All of the synthesized compounds were characterized and purity confirmed by ^{1}H-NMR, ^{13}C-NMR and ESI-MS, and some were additionally validated by ^{15}N-NMR. 8-hydroxy-2′-deoxy-($^{15}N_5$)-guanosine ([$^{15}N_5$]8-oxo-G) (99 atom% ^{15}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*-^{13}C,D_3)-L-Met (99 atom% D and ^{13}C) was purchased from Sigma-Aldrich, and 2′-deoxy-(C^2–^{13}C, N^1,N^3–$^{15}N_2$)-thymidine (99 atom% ^{13}C and ^{15}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-octadinyl-dU (Z) phosphoramidite was synthesized according to literature and characterized by ^{1}H-NMR, ^{13}C-NMR, ^{31}P-NMR and ESI-MS[43]. 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 NHEt$_3$OAc in water) and buffer B (0.1 M NHEt$_3$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 pulldown 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$_3$-S-S-NHS ester ($C_{18}H_{29}N_5O_8S_2$, 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 a 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[29,42,44,45], 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_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 l/min (N_2), 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_2O were digested as follows: An aqueous solution (7.5 µl) of 480 µM $ZnSO_4$, 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]$_2$-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-BL6/N) were dissected at postnatal day 90 and prepared as earlier described[19,21]. 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 3,5-di-*tert*-butyl-4-hydroxytoluene. (BHT, 200 µM) and deferoxamine mesylate salt (desferal, 200 µM) as well as the deaminase inhibitor tetrahydrouridine (THU, 200 µM), according to published methods, to reduce background oxidation or deamination[18]. 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 iPrOH. DNA pellets from cultured cells were washed twice with ice-cold 70% EtOH and resuspended in H_2O 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$_3$, 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_2O 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)[46] were cultured in the presence of serum and LIF as previously described[4]. *Tdg*$^{+/−}$ and *Tdg*$^{−/−}$ mESCs were described previously[47]. *Tdg*$^{−/−}$ mESCs were complemented by random integration of either empty vector (hereafter referred to as *Tdg*$^{−/−}$ mESCs) or a minigene expressing catalytically inactive Tdg (N151A)[47]. Clonal mESC lines with targeted Tdg alleles, R1 cells (strain 129/Sv)[48], J1 cell lines (strain 129S4/SvJae)[49] and a mESC line derived from C57Bl/6/129 mixed background[50] 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 (≥1 ×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 µM $[^{13}C,^{15}N_2]T$. 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 Technology) supplemented as above and with 0.2 mM of either $[methyl-^{13}C,D_3]$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 1×10^5 cells/cm² on gelatin-coated plates in N2B27 medium containing 1,000 U/ml LIF to favor attachment and initial survival[51]. 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 $[methyl-^{13}C,D_3]$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 $[methyl-^{13}C,D_3]$L-Met or natural L-Met. Differentiation of R1 cells in the presence of FGF-2 and ActA was as described[26], 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 2.2×10^5 cells/cm² 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 µg/ml polybrene, shRNA-expressing mESCs were selected with 4 µg/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[53]. 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 µg of Smug1 esiRNAs (34.5 nM) and 20 µl 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 2.5×10^6 cells per 75 cm² flask and incubated in 10 ml of medium. The transfection solution (500 µl of jetPRIME buffer, a specific amount of plasmid DNA (**Supplementary Note 4**) and 20 µl 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 µl of jetPRIME buffer, 5 µg of esiRNA and 20 µl 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 esiRNAEHU079221). **Supplementary Note 4** summarizes the overexpression and knockdown procedures.

Isotope tracing with $[^{13}C,^{15}N_2]$thymidine or $[^{13}C,D_3]$L-Met in HEK-293T cells transfected with Tet1cd. 24 h before transfection, 2.5×10^6 cells were seeded in a 75-cm² flask containing 10 ml either of (for $[^{13}C,^{15}N_2]$thymidine) DMEM medium supplemented with 50 µM $[^{13}C,^{15}N_2]T$ or (for $[^{13}C,D_3]$L-Met) DMEM medium lacking L-Met, L-cystine and pyruvate, which was supplemented with 10% dialyzed FBS, 2 mM $[^{13}C,D_3]$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

system. 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 (Bio-Rad); real-time PCR was performed with SsoFast EvaGreen Supermix (Bio-Rad). 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 µg 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 °C and proteinase K treatment, the oxidized plasmids were purified with GeneJET PCR Purification Kit from Thermo Scientific and eluted in 25 µl water. Samples were then subjected to LC/MS/MS analysis as described[42]. 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 µg (50 µl) of the crude nuclear protein extracts were filled up to 500 µl 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 µM 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 µl 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 7.5×10^7 cells were used. Protein concentration was determined by Bradford assay. For each experiment 250 µg (50 µl) 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 °C and subsequently alkylated with 200 mM iodoacetamide in the dark for 30 min at 25 °C. Protein samples were digested with 0.5 µg trypsin (Promega, Madison, MA, USA) for 16 h at 37 °C. The reaction was stopped using 1 mM phenylmethylsulphonylfluoride. 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 μl were injected for the analysis. The samples were desalted and concentrated on a μ-precolumn cartridge (PepMap100, C18, 5 μM, 100 Å, size 300 μm i.d. x 5 mm) and further processed on a custom-made analytical column (ReproSil-Pur, C18, 3 μM, 120 Å, packed into a 75 μm i.d. x 150 mm and 8 μm 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' 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).

Supplementary Information

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

Toni Pfaffeneder[1,8], Fabio Spada[1,8], Mirko Wagner[1,8], Caterina Brandmayr[1], Silvia Laube[1], David Eisen[1], Matthias Truss[2], Jessica Steinbacher[1], Benjamin Hackner[1], Olga Kotljarova[1], David Schuermann[5], Stylianos Michalakis[3], Olesea Kosmatchev[1], Stefan Schiesser[1], Barbara Steigenberger[1], Nada Raddaoui[1], Gengo Kashiwazaki[1], Udo Müller[4], Cornelia G. Spruijt[6], Michiel Vermeulen[6,7], Heinrich Leonhardt[4], Primo Schär[5], Markus Müller[1]* and Thomas Carell[1]*

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

[2] Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Ziegelstr. 5-9, 10098 Berlin, Germany.

[3] Center for Integrated Protein Science at the Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany.

[4] Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany.

[5] Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland.

[6] Department of Molecular Cancer Research, Cancer Genomics Netherlands, UMC Utrecht, 3584 CG Utrecht, Netherlands.

[7] Present Address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA, Nijmegen, The Netherlands

[8] These authors contributed equally

* Correspondence to: markus.mueller@cup.uni-muenchen.de and thomas.carell@cup.uni-muenchen.de

Supplementary Results
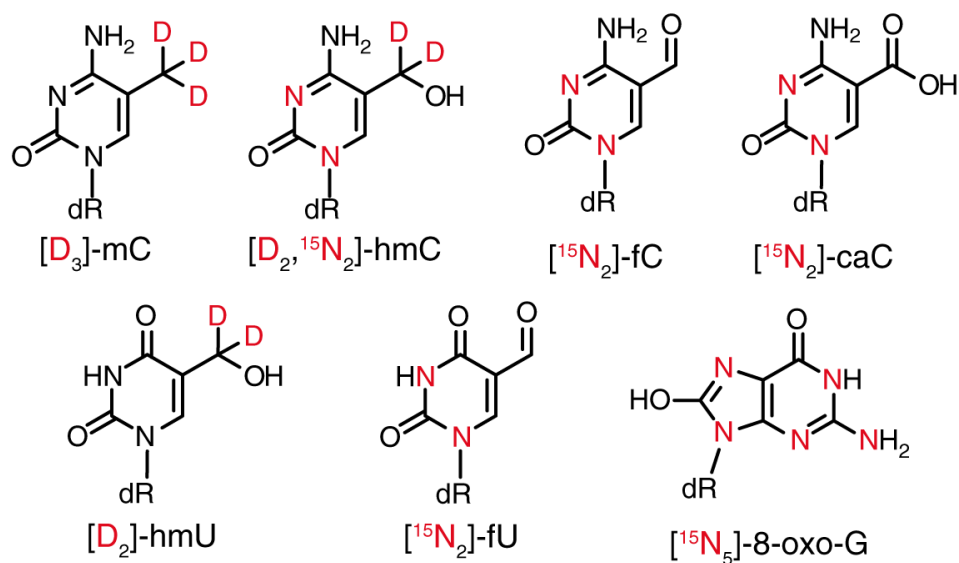
Supplementary Figure 1. Isotopically labeled nucleosides used as internal standards for quantitative LC-MS/MS analysis (dR = -2´-deoxyribose).
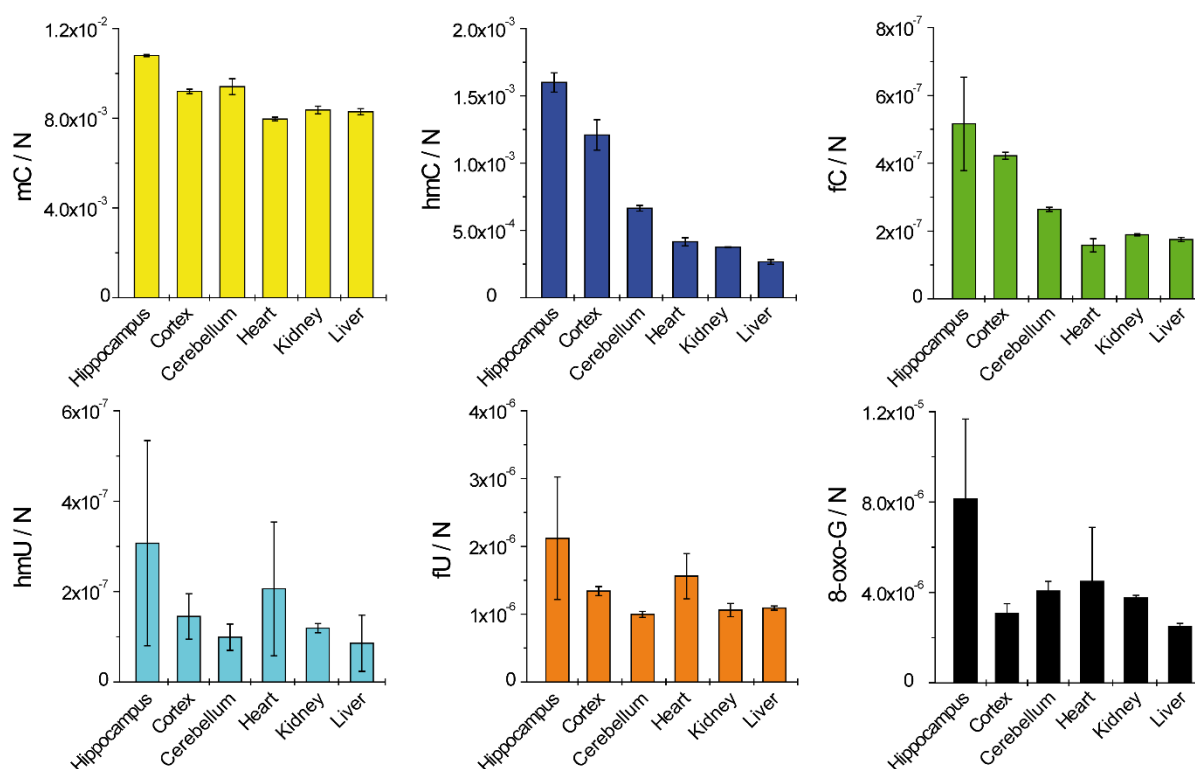
Supplementary Figure 2. DNA modification levels per nucleoside (N) of different mouse tissues from 3 month old individuals ($n = 3$). Depicted are biological mean values ± s.d.
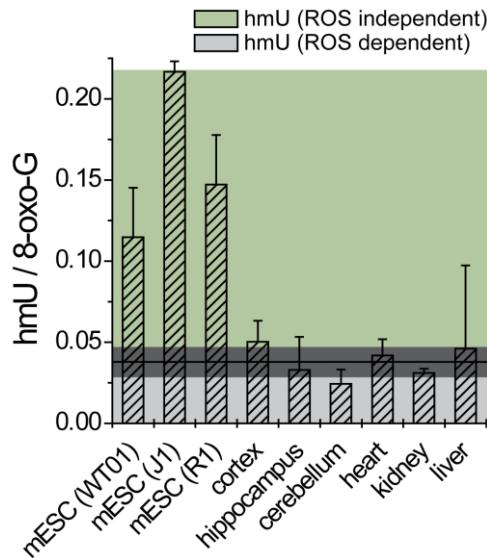
Supplementary Figure 3. hmU levels normalized to the oxidative background marker 8-oxo-G in mESCs ($n_{WT01} = 7$; $n_{J1} = 2$, $n_{R1} = 2$) and mouse tissues ($n = 3$, 3 month old individuals) in order to dissect ROS dependent and ROS independent processes. Normalization was necessary to take deviating background oxidation of DNA sample preparation into account. The light grey area reflect the hmU level fractions, which are generated by ROS dependent processes. The green area reflect the hmU level fractions, which are generated by ROS independent processes. The assignment is based on the assumption, that hmU levels in somatic tissue are exclusively ROS created lesions (derived from the cluster analysis in **Fig. 2b**). The dark grey area reflect the mean value ± s.d. of hmU/8-oxo-G ratios of the tissue data. The difference between the height of the hmU/8-oxo-G ratios of mESCs and the mean of the tissue data give the hmU fraction which is formed by ROS-independent processes. In WT01, J1 and R1 cells about 67%, 83% and 74%, respectively, of the global hmU levels are estimated to be created by ROS independent processes. Bars reflect biological mean values ± s.d. The differences between mESCs (WT01, J1) and mouse tissues are significant ($P = 9.3 \times 10^{-5}$–3.3×10^{-2}; unpaired two-tailed t-test) except for WT01 and liver ($P = 0.15$). The differences between R1 and mouse tissues were mostly not significant on a level < 0.05 ($P = 6.4 \times 10^{-2}$–1.1×10^{-1}) except for R1 and kidney ($P = 3.3 \times 10^{-2}$).

Supplementary Figure 4. Schematic representation of isotope tracing experiments with $[^{13}C,^{15}N_2]$-T (left; blue) and $[methyl$-$^{13}C,D_3]$L-methionine (right; red) and exchange rates of derived genomic isotopologs. Small negligible deviations in the exchange rates are due to differential noise sources. LOD = limit of detection. WT = wild type mESCs; Tdg cm = $Tdg^{-/-}$ mESCs complemented with catalytic inactive Tdg (see **Supplementary Fig. 6**).
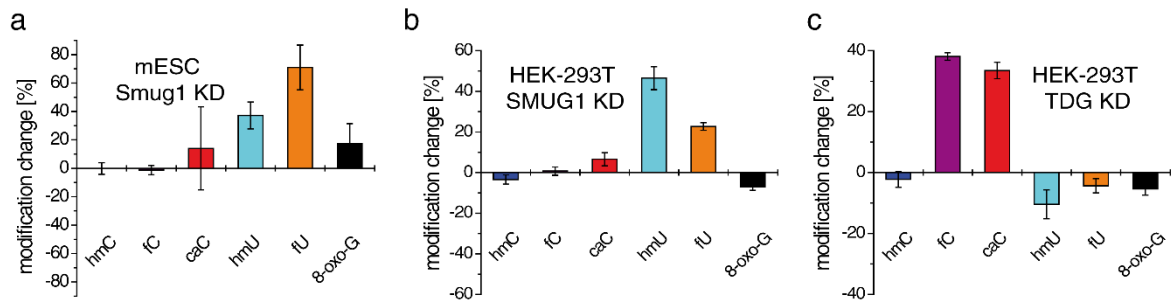
Supplementary Figure 5. (**a**) Effect of Smug1 depletion on modification levels in mESCs (R1). Effect of SMUG1 (**b**) and TDG (**c**) depletion on modification levels in HEK-293T cells overexpressing Tet1cd. Shown is the percent change in modification content per nucleoside in cells co-transfected with esiRNAs targeting TDG or SMUG1 relative to co-transfection with control esiRNA. Depicted are technical mean values ± s.d.
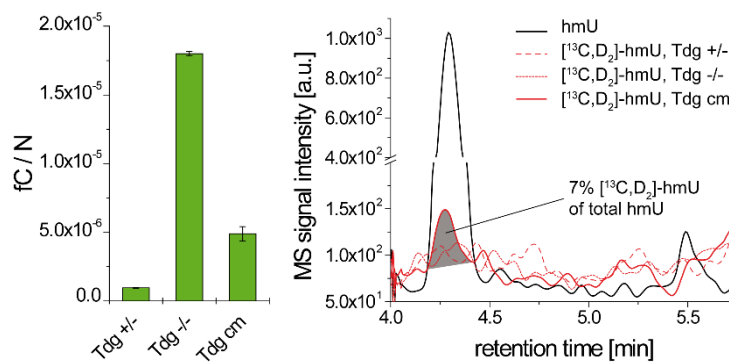
Supplementary Figure 6. Isotope tracing experiments with $Tdg^{+/-}$, $Tdg^{-/-}$ mESCs as well as $Tdg^{-/-}$ mESCs complemented with a catalytic mutant of Tdg (Tdg cm) grown in the presence of $[^{13}C,D_3]$L-Met (200 µM). The catalytic mutant of Tdg is not completely inactive (fC levels are between $Tdg^{+/-}$ and $Tdg^{-/-}$ cells, left). Only in case of $Tdg^{-/-}$ cells complemented with a catalytic inactive Tdg (Tdg cm) labeled hmU was detected, which originated from the deamination of labeled hmC. ~7% $[^{13}C,D_2]$hmU over total hmU was observed. This corresponds to ~0.06% deamination of hmC to hmU under these conditions (2.5×10^{-4} total hmC / N; 2.2×10^{-6} total hmU / N). For comparison, ~2% of hmC was oxidized to fC in these cells and ~6% of hmC was oxidized to fC in $Tdg^{-/-}$ cells. Labeled fU was not observed. Depicted fC-levels represent technical mean values ± s.d.
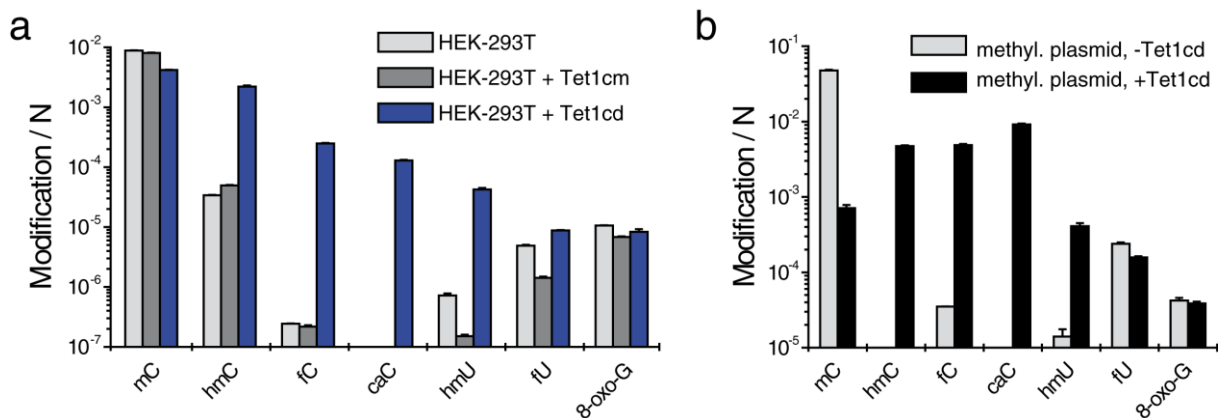
Supplementary Figure 7. Tet1 and Tet2 generate hmU in HEK-293T and in vitro (**a**) Effect of Tet1 overexpression on modified pyrimidines in HEK-293T cells. Modification levels in cells overexpressing WT and catalytic mutant versions of Tet1 catalytic domain (Tet1cd, blue bars and Tet1cm, gray bars, respectively), or a control construct (white bars). Depicted are mean values ± s.d. of technical triplicates on a logarithmic scale. (**b**) Pyrimidine modification levels in methylated plasmid DNA after *in vitro* treatment with Tet1cd. Depicted are mean values ± s.d. of technical duplicates. Note the logarithmic scale.

313
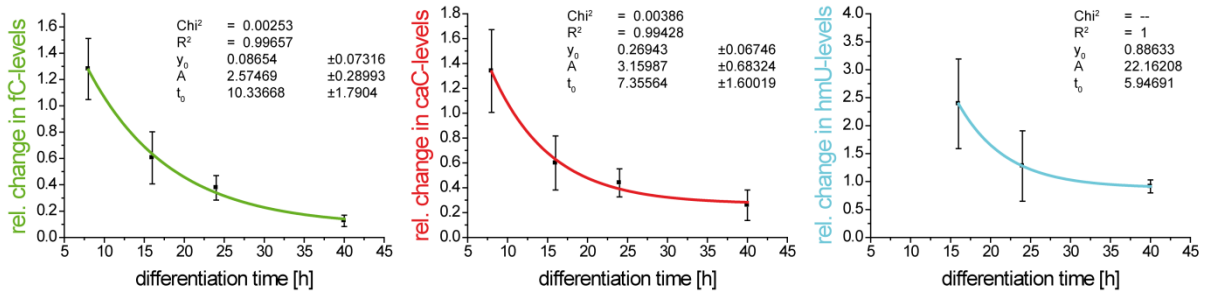
Supplementary Figure 8. Exponential models for fitting the decay curve of fC, caC and hmU in combined data sets from differentiation of R1 and C57Bl6/129-derived mESCs (6 biological independent experiments). In a simplified approach a single exponential decay model ($y = y_0 + A*\exp(-x/t_0)$) was plotted using ORIGIN$^{®}$. The parameters y_0 (offset), t_0 (time constant) and A (amplitude) of each decay function were iteratively optimized until the minimum of the Chi^2 value of the fitting was reached. Half-life times ($t_{1/2} = t_0 * \ln2$) for fC, caC and hmU were 7.2±1.2, 5.1±1.1 and 4.1 h, respectively. Depicted are biological mean values ± s.d.
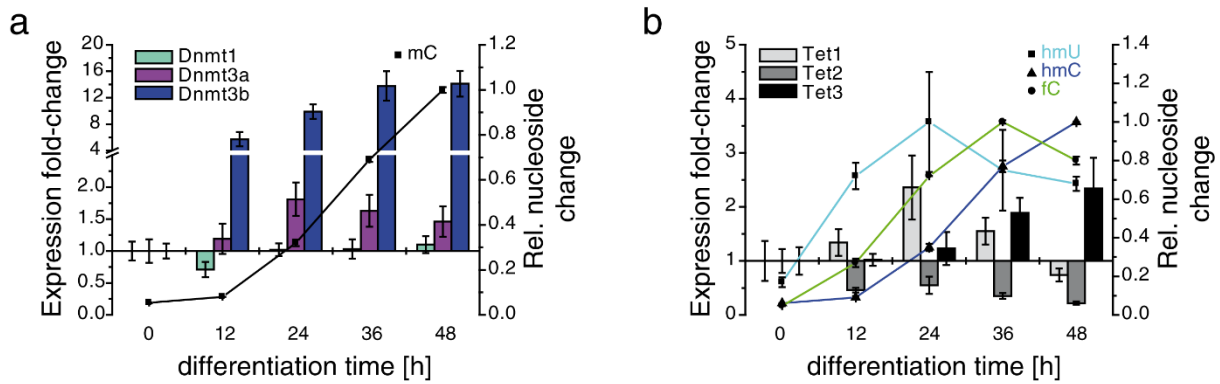
Supplementary Figure 9. Normalized transcript levels of Dnmts (**a**), Tet1–3 (**b**) and normalized modification levels of mC (**a**), hmC, fC and hmU (**b**) during differentiation of naive mESCs in the presence of FGF-2 and ActA.
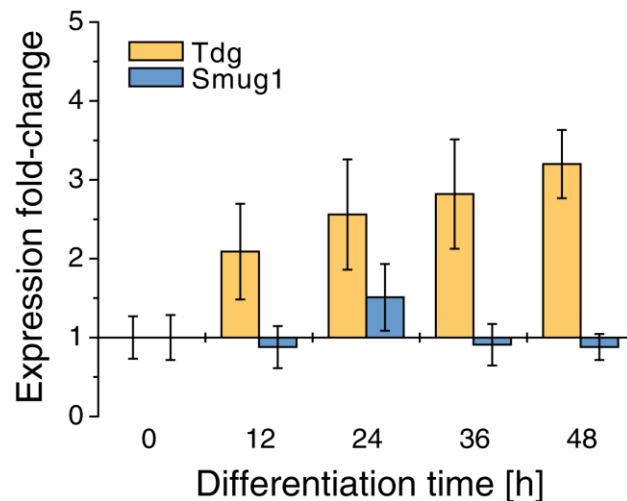
Supplementary Figure 10. Expression level analysis of Tdg and Smug1 during differentiation of mESCs in the presence of FGF-2 and ActA. Expression levels were quantified with respect to the housekeeping gene *Gapdh* and normalized to time point 0 h. Depicted are technical mean values ± s.d.
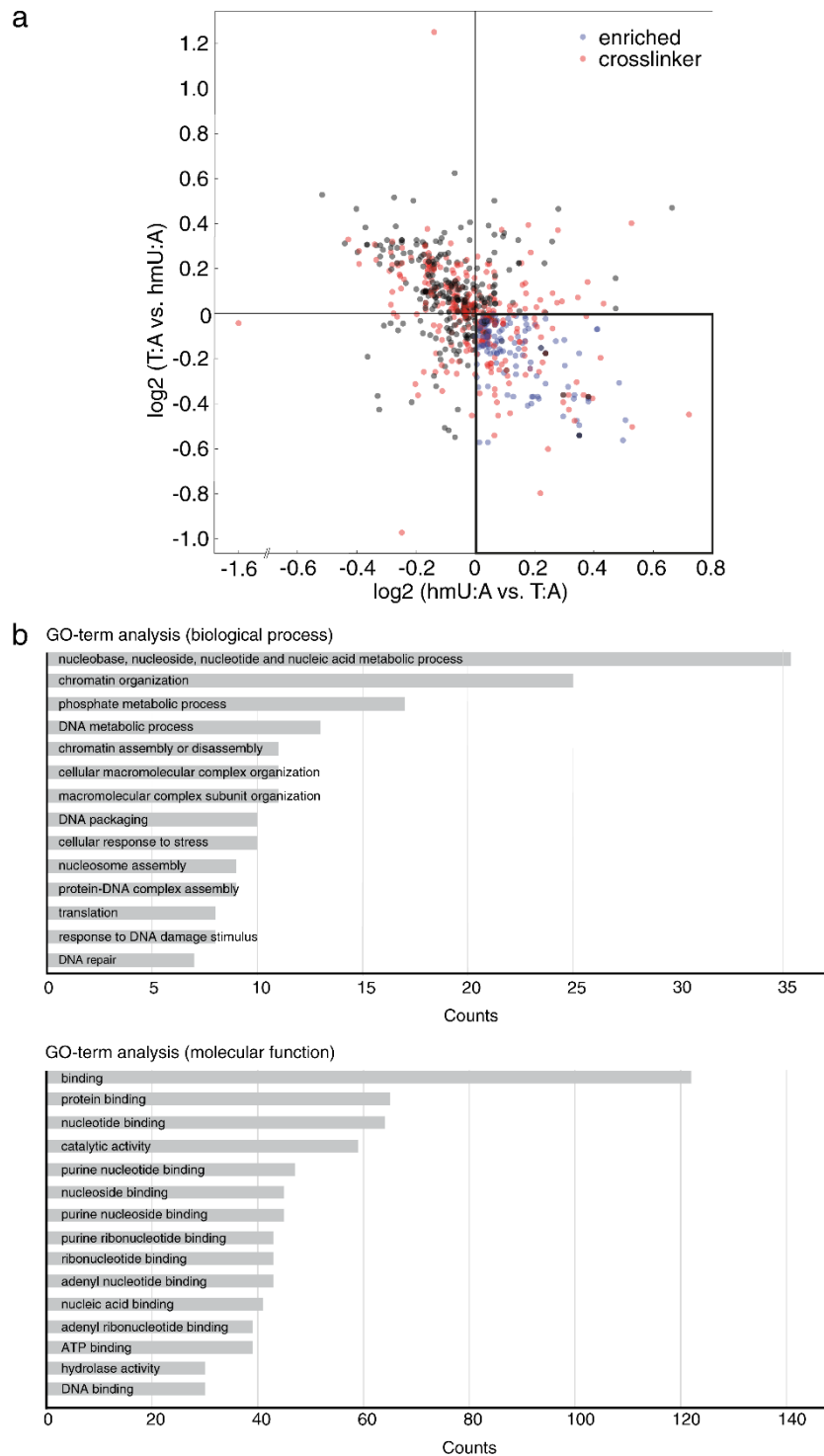
Supplementary Figure 11. (a) Scatterplot of proteins enriched with the hmU:A containing oligomer. Ratios of a forward and a reverse experiment are plotted. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross linker and marked in red. Gray dots are considered unspecific binders. See **Fig. 6** for detailed view. **(b)** Gene Ontology Analysis performed with DAVID Bioinformatics Resources 6.7[4].
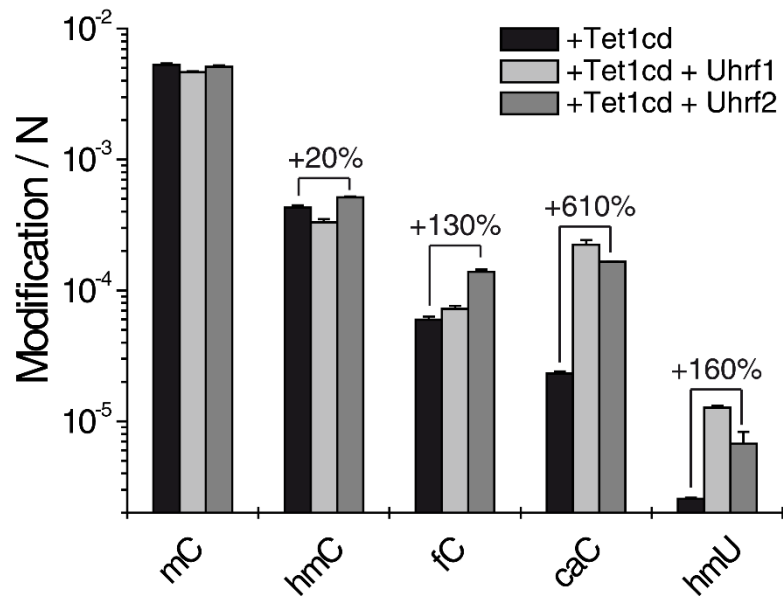
Supplementary Figure 12. Effect on modified pyrimidines in HEK-293T cells upon ectopic co-expression of Tet1cd and Uhrf1 (light gray bars) or Uhrf2 (gray bars). Depicted are mean values ± s.d. of technical triplicates on a logarithmic scale.

Supplementary Table 1. Isotope tracing experiments by supplementing the growth medium of mES cells (LIF), differentiating mESCs (R1, without growth factors) and HEK-293T cells with either $[^{13}C,^{15}N_2]T$ (50 or 100 μM) or $[methyl\text{-}^{13}C,D_3]$L-methionine (0.2 mM). Small deviations in the exchange yields are due to differential noise sources and are negligible. LOD = Limit of detection. In case of $[^{13}C,D]fU$ no difference was observed compared to the natural control.

cell type	growth medium	$[^{13}C,^{15}N_2]T$ / T [%]	$[^{13}C,^{15}N_2]hmU$ / hmU [%]	$[^{13}C,^{15}N_2]fU$ / fU [%]
mESC (2i)	100 μM natural T	0.1	< LOD	< LOD
mESC (2i)	100 μM $[^{13}C,^{15}N_2]T$	76.0	78.2	74.6
HEK + Tet1cd (72 h)	50 μM $[^{13}C,^{15}N_2]T$	73.8	74.2	71.0

cell type	growth medium	$[^{13}C,D_3]mC$ / mC [%]	$[^{13}C,D_2]hmC$ / hmC [%]	$[^{13}C,D_2]hmU$ / hmU [%]	$[^{13}C,D]fU$ / fU [%]
mESC (LIF)	natural methionine	0.1	< LOD	< LOD	3.3
mESC (LIF; 0 h)	$[^{13}C,D_3]$L-methionine	88.9	87.6	< LOD	3.0
diff. mESC (12 h)	$[^{13}C,D_3]$L-methionine	89.3	88.4	< LOD	3.1
diff. mESC (24 h)	$[^{13}C,D_3]$L-methionine	90.1	89.3	< LOD	3.2
diff. mESC (48 h)	$[^{13}C,D_3]$L-methionine	90.4	90.5	< LOD	3.5
mESC $Tdg^{+/-}$	$[^{13}C,D_3]$L-methionine	88.0	87.6	< LOD	3.7
mESC $Tdg^{-/-}$	$[^{13}C,D_3]$L-methionine	87.4	87.2	< LOD	< LOD
mESC $Tdg^{-/-}$ + Tdg cm	$[^{13}C,D_3]$L-methionine	86.9	86.7	7.4	2.2
HEK + Tet1cd (72 h)	$[^{13}C,D_3]$L-methionine	87.4	83.4	< LOD	< LOD

Supplementary Table 2. Assessment of ROS dependent hmU and fU product ratio of T-oxidation in HEK-293T wild type cells where TET activity is lowest (related to **Fig. 3c**). Modified nucleosides / N are given as mean values ± s.d. of three independent technical replicates. When T is oxidized by ROS about 9.8% hmU and 90.2% fU is generated relative to the sum of each other.

n	mC / N		hmC / N		fC / N		caC / N
	techn. mean	s.d.	techn. mean	s.d.	techn. mean	s.d.	techn. mean
1	6.14E-03	1.97E-04	2.92E-05	9.90E-08	2.56E-07	2.59E-09	n.d.
2	6.21E-03	1.32E-05	2.95E-05	2.21E-09	3.10E-07	2.06E-08	n.d.
3	5.76E-03	4.36E-05	3.32E-05	6.55E-07	2.86E-07	2.95E-09	n.d.
4	9.01E-03	2.87E-04	5.23E-05	5.01E-07	3.50E-07	5.00E-09	n.d.
5	8.80E-03	9.62E-05	3.39E-05	5.44E-07	2.43E-07	2.41E-09	n.d.
6	8.55E-03	7.69E-05	3.70E-05	2.17E-07	2.14E-07	1.30E-08	n.d.
biol. mean	7.41E-03		3.59E-05		2.76E-07		
biol. s.d.	1.52E-03		8.56E-06		4.90E-08		

n	hmU / N		fU / N		hmU/(hmU+fU)	fU/(hmU+fU)	8-oxo-G / N	
	techn. mean	s.d.	techn. mean	s.d.	[%]	[%]	techn. mean	s.d.
1	2.66E-07	5.54E-08	5.15E-06	1.59E-07	4.9	95.1	8.49E-06	1.30E-07
2	1.21E-06	1.62E-07	8.02E-06	6.66E-07	13.1	86.9	1.00E-05	1.98E-07
3	3.65E-08	7.35E-09	9.76E-07	8.75E-09	3.6	96.4	3.02E-06	7.02E-08
4	6.31E-07	9.10E-10	3.51E-06	3.28E-08	15.3	84.7	7.43E-06	1.34E-07
5	7.21E-07	5.93E-08	4.89E-06	1.59E-07	12.8	87.2	1.06E-05	1.14E-07
6	3.58E-07	4.95E-08	3.56E-06	9.64E-09	9.1	90.9	7.22E-06	1.15E-08
biol. mean	5.37E-07		4.35E-06		9.8	90.2	7.80E-06	
biol. s.d.	4.13E-07		2.33E-06		4.8	4.8	2.70E-06	

Supplementary Note 1: oligonucleotide sequences for protein pull-down assays

Supplementary Table 3. DNA oligonucleotides used in protein pulldown studies.

ODN	Sequence (5´→ 3´)	Modifications
1	Biotin-GCA-TCC-GGT-CAY-CGT-TCC-TTC-GGA	Y = 5-octadienyl-U
2	Biotin-GCA-TCC-GGT-CAY-CAT-TCC-TTC-GGA	Y = 5-octadienyl-U
3		X= T
4	TCC-GAA-GGA-AXG-ATG-ACC-GGA-TGC	X= hmU
5		X= C
6		X= hmC
7	Biotin-GCT-CAC-GCT-AGY-CGA-CTC-CGT-GCA	Y = 5-octadienyl-U
8	TGC-ACG-GAG-TXG-ACT-AGC-GTG-AGC	X = T
9		Y = hmU

Hybridization scheme:

Pulldown 1: hmU:A vs. T:A : ODN4/2 vs. ODN3/2
Pulldown 2: hmU:G vs. C:G = ODN4/1 vs. ODN5/1
Pulldown 3: hmC:G vs. C:G = ODN6/1 vs. ODN5/1
Pulldown 4 (scrambled sequence): hmU:A vs. T:A = ODN9/7 vs. ODN8/7

Supplementary Note 2: LC-UV-ESI-MS/MS analysis of DNA

Supplementary Table 4. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA. CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ([M+H]$^+$ species) as well as in the negative ([M−H]$^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
			time segment 1.5–4.0 min					
[^{15}N$_2$]caC	274.08	Wide	158.03	Wide	170	5	5	Positive
caC	272.09	Wide	156.04	Wide	170	5	5	Positive
[^{15}N$_2$,D$_2$]hmC	262.12	enhanced	146.07	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
[D$_3$]mC	245.13	enhanced	129.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
C	228.1	enhanced	112.05	enhanced	1	1	0	Positive
			time segment 4.0–6.0 min					
[D$_2$]hmU	259.09	Wide	216.08	Wide	48	7	5	Negative
[D$_2$]hmU	259.09	Wide	126.05	Wide	48	7	5	Negative
hmU	257.08	Wide	214.07	Wide	48	7	5	Negative
hmU	257.08	Wide	124.04	Wide	48	7	5	Negative
[^{15}N$_2$]fU	257.06	Wide	213.05	Wide	48	6	5	Negative
fU	255.06	Wide	212.06	Wide	48	6	5	Negative
			time segment 6.0–9.0 min					
[^{15}N$_5$]8-oxo-G	289.08	Wide	173.04	Wide	120	9	7	Positive
8-oxo-G	284.1	Wide	168.05	Wide	120	9	7	Positive
[^{15}N$_2$]fC	258.09	Wide	142.04	Wide	120	5	5	Positive
fC	256.09	Wide	140.05	Wide	120	5	5	Positive

Supplementary Table 5. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA obtained from cells which were grown in medium supplemented with labeled thymidine ($[^{13}C,^{15}N_2]T$). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$ species) as well as in the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1.5–4.0 min								
$[^{13}C,^{15}N_2]caC$	275.09	wide	159.04	wide	65	5	5	Positive
caC	272.09	wide	156.04	wide	65	5	5	Positive
$[^{13}C,^{15}N_2]hmC$	261.11	enhanced	145.06	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
$[^{13}C,^{15}N_2]mC$	245.13	enhanced	129.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
$[^{13}C,^{15}N_2]C$	231.1	enhanced	115.05	enhanced	40	1	3	Positive
C	228.1	enhanced	112.1	enhanced	40	1	3	Positive
time segment 4.0–6.0 min								
$[^{13}C,^{15}N_2]hmU$	260.08	wide	215.07	wide	50	7	5	Negative
hmU	257.08	wide	214.07	wide	50	7	5	Negative
$[^{13}C,^{15}N_2]fU$	258.06	wide	213.05	wide	50	6	5	Negative
fU	255.06	wide	212.06	wide	50	6	5	Negative
time segment 6.0–9.0 min								
$[^{15}N_5]8\text{-oxo-}G$	289.08	wide	173.04	wide	80	9	7	Positive
8-oxo-G	284.1	wide	168.05	wide	80	9	7	Positive
$[^{13}C,^{15}N_2]fC$	259.09	wide	143.04	wide	80	5	5	Positive
fC	256.09	wide	140.05	wide	80	5	5	Positive
$[^{13}C,^{15}N_2]T$	246.1	enhanced	130.05	enhanced	30	40	3	Positive
T	243.1	enhanced	127.05	enhanced	30	40	3	Positive

Supplementary Table 6. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA obtained from cells which were grown in medium supplemented with labeled [*methyl*-$^{13}C,D_3$]L-methionine. CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$ species) as well as in the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1.5–4.0 min								
$[^{13}C]caC$	273.09	wide	157.04	wide	65	5	5	Positive
caC	272.09	wide	156.04	wide	65	5	5	Positive
$[^{13}C,D_2]hmC$	261.12	enhanced	145.08	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
$[^{13}C,D_3]mC$	246.14	enhanced	130.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
C-dN	228.1	enhanced	112.1	enhanced	40	1	3	Positive
time segment 4.0–6.0 min								
$[^{13}C,D_2]hmU$	260.09	wide	217.09	wide	60	7	5	Negative
hmU	257.08	wide	214.07	wide	60	7	5	Negative
$[^{13}C,D]fU$	257.07	wide	214.07	wide	60	6	5	Negative
fU	255.06	wide	212.06	wide	60	6	5	Negative
time segment 6.0–9.0 min								
$[^{15}N_5]8\text{-oxo-}G$	289.08	wide	173.04	wide	80	9	7	Positive
8-oxo-G	284.1	wide	168.05	wide	80	9	7	Positive
$[^{13}C,D]fC$	258.1	wide	142.06	wide	80	5	5	Positive
fC	256.09	wide	140.05	wide	80	5	5	Positive

Spiking amounts of labeled internal standards for quantitative LC-MS/MS analysis

The quantification of nucleosides of genomic DNA isolated from mESCs or mouse tissue was carried out with the following amounts of internal standards: 51.03 pmol [D_3]mC, 7.655 pmol [$^{15}N_2,D_2$]hmC, 45.6 fmol [$^{15}N_2$]fC, 43.0 fmol [$^{15}N_2$]caC, 108.9 fmol [$^{15}N_5$]8-oxo-G; 160.1 fmol [D_2]hmU and 180.0 fmol [$^{15}N_2$]fU. The quantification of nucleosides of genomic DNA isolated from HEK-293T cells overexpressing Tet was carried out with the following amounts of internal standards: 34.02 pmol [D_3]mC, 5.103 pmol [$^{15}N_2,D_2$]hmC, 303.8 fmol [$^{15}N_2$]fC, 215.1 fmol [$^{15}N_2$]caC, 108.9 fmol [$^{15}N_5$]8-oxo-G; 160.1 fmol [D_2]hmU and 180.0 fmol [$^{15}N_2$]fU. Genomic DNA samples isolated from cells grown in media supplemented with either [$^{13}C,^{15}N_2$]T or [*methyl*-$^{13}C,D_3$]L-methionine were not spiked with internal standards except [$^{15}N_5$]8-oxo-G.

Validation of the LC-UV-MS/MS quantification method:

Method validation, in particular linearity, precision, and accuracy (i.e. determined from matrix samples spiked with isotopically labeled internal standards) of the established method were investigated. Validation for the established LC-UV-ESI-MS/MS quantification method was based on three different series (i.e., calibration functions and quality control samples) accomplished on different days. Each calibration standard (5–8 standard concentrations) was analyzed five times. Each validation experiment was complemented by matrix blanks (analyzed in triplicates) to ensure selectivity and specificity of the method. Linear regression was applied by Origin® 6.0 (Microcal™) to obtain calibration curves. Therefore, the ratio of the area under the curve (A/A*) of the unlabeled nucleoside to the internal standard (*) was plotted against the ratio of the amount of substance (n/n*) of the unlabeled nucleoside to the internal standard (*) (see **Supplementary Fig. 13**). Calibration functions were calculated without weighting. Additionally, acceptable accuracy (80–120%) as well as precision (<20% relative s.d.) was required. Accuracy was proven by computing the amount of substance n from the obtained A/A* ratios of the calibration standards using the respective calibration function. Here, accuracy was defined as the ratio of the used amount of substance to the calculated amount of substance in percent and had to be between 80–120% for each standard concentration. Precision was defined as follows: technical replicates of A/A* ratios for each calibration standard had to have relative standard deviations (r.s.d) smaller than 20%. The lower limit of quantification (LLOQ) was defined as the lowest concentration fulfilling the requirements of accuracy and precision and achieving a response of at least three times the response compared with the blank response. A compilation of absolute and relative LLOQs is shown in **Supplementary Table 7**.

Quality control samples to evaluate intra-batch precision (see below) were investigated using a biological sample spiked with internal standards. Long-term stability of aqueous solutions of the labeled and unlabeled nucleosides at a storage temperature of −20 °C was investigated over two months including several freeze and thaw cycles by analyzing the MS/MS-responses with each batch.

Supplementary Table 7. Compilation of absolute lower limits of quantification [fmol] (LLOQ; see **Supplementary Fig. 13**) and relative LLOQs [per N] depending on the amount of DNA, which is digested. The relative LLOQs were computed by generating ratios of the absolute LLOQ [pmol] to the total amount of nucleosides (N; [pmol]) in the respective amount of DNA [µg]. The total amount of nucleosides was obtained by using the average molar mass of 308.91 g mol^{-1} for the monomeric DNA entity by taking the GC-content (21% C or G) in mouse into account.

	absolute LLOQ [fmol]	relative LLOQ [per N]	relative LLOQ [per N]	relative LLOQ [per N]
DNA amount		**5 µg**	**10 µg**	**25 µg**
mC	104.5	6.5E-06	3.2E-06	1.3E-06
hmC	100.8	6.2E-06	3.1E-06	1.2E-06
fC	2.1	1.3E-07	6.5E-08	2.6E-08
caC	2.0	1.2E-07	6.2E-08	2.5E-08
8-oxo-G	2.0	1.2E-07	6.2E-08	2.5E-08
hmU	6.4	4.0E-07	2.0E-07	7.9E-08
fU	5.7	3.5E-07	1.8E-07	7.0E-08

Intra-batch assay and quantification data processing:

In order to evaluate intra-batch precision (see below) quality samples were investigated using a biological sample spiked with internal standards. The intra-batch-assay was performed for the LC-ESI-MS/MS analysis of the nucleosides G, C, mC, hmC, fC, caC, hmU, fU and 8-oxo-G. For this, a representative mESC DNA sample was analyzed. Technical replicates ($n = 5$; each 4 µg DNA) were independently prepared using the below described digestion protocol. For data processing MassHunter Quantitative Analysis from Agilent was used. The area under the curve (A) was determined by LC-MS/MS for mC, hmC, fC, caC, hmU, fU, 8-oxo-G and for the corresponding labeled internal standards (A*); the area under the curve (A_{UV}) for G and C was determined by LC-UV. The amount of substance (n; pmol) of each nucleoside was computed by using the calibration curves (see **Supplementary Fig. 13**). The total sample volume was 40 µL, the injection volume after sample filtration was 29 µL. Therefore, the obtained values of G and C by LC-UV quantification were corrected by the factor given by the ratio of 40 µL/29 µL. Careful monitoring of the exact pipetting and injection volumes was therefore necessary. The obtained absolute amounts (pmol) of the DNA modifications (X = C, mC, fC, caC, hmU, fU, 8-oxo-G) were then related to the amount of G (pmol) giving ratios of X / G in %. The sum of X / G was defined as 100%. These values were then transferred in X / N values, considering that the G content is 21% in mouse.

The determined A/A* ratios of the DNA nucleosides to the labeled internal standards (see **Supplementary Table 8**) showed a high precision (r.s.d. = 3.9–18%) for each nucleoside. The necessity in using labeled internal standards for quantification is shown by comparing these results with the relative standard deviation (r.s.d. = 7.6–49.5%) of the uncorrected mass signal (A) of the respective DNA modification. Moreover, in order to gain precision between sample batches measured on different days (data not shown), it is even more important to use internal standards. No memory effect was observed during blank experiments performed after several measurements of a sample. The blank analyses were not contaminated by carry-over.
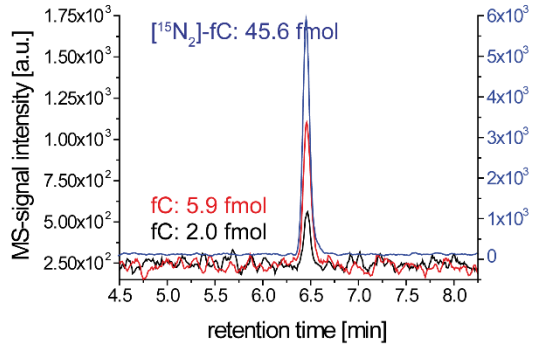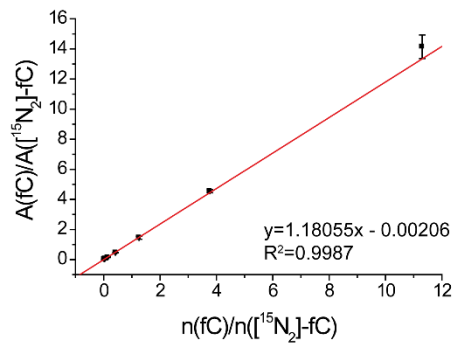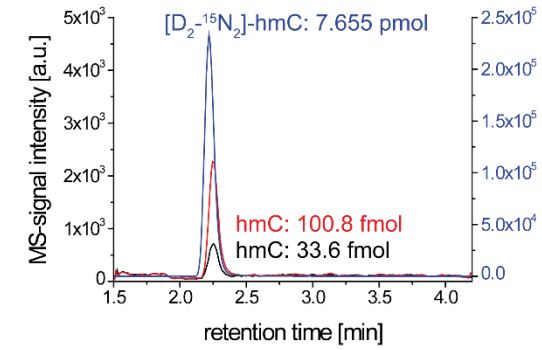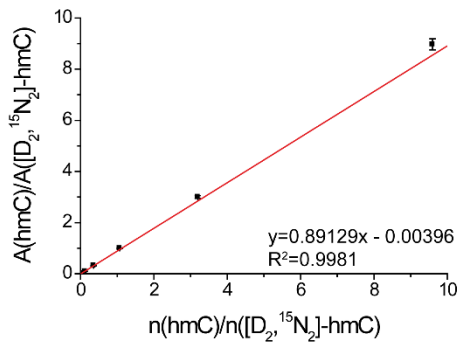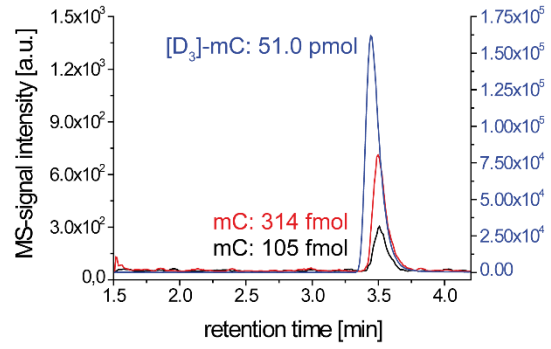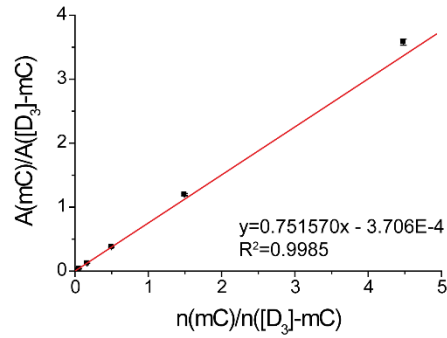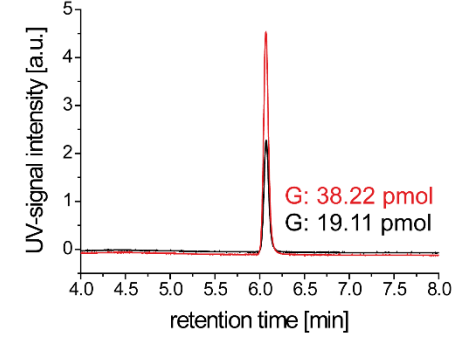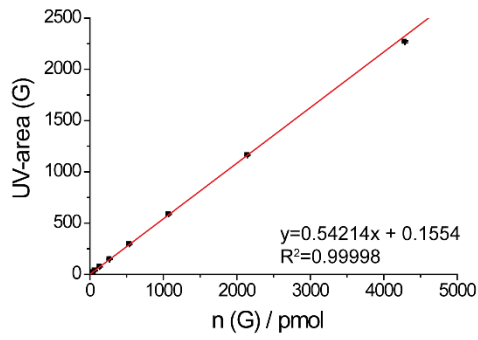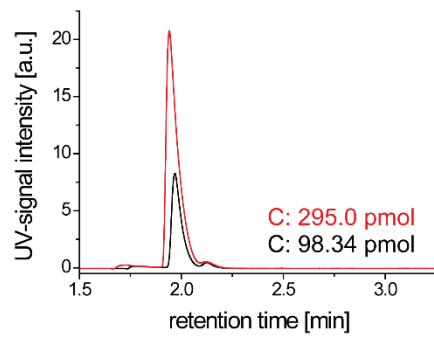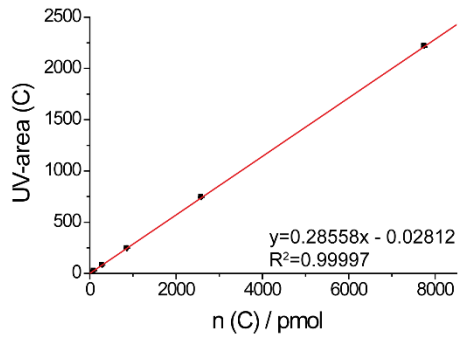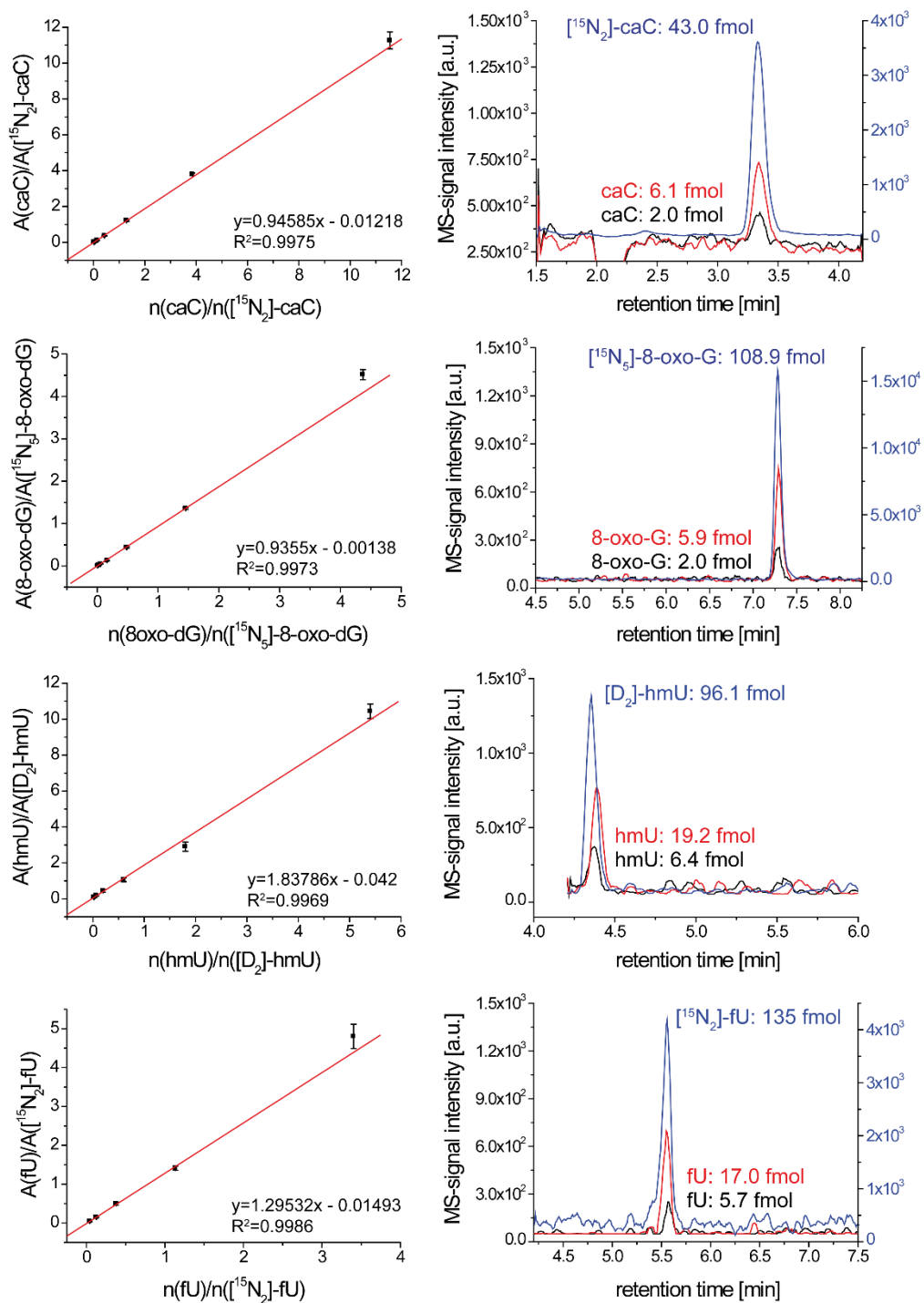
Supplementary Figure 13. UV and LC-MS/MS calibration curves and representative chromatograms of C (UV), G (UV), mC/[D$_3$]mC, hmC/[^{15}N$_2$,D$_2$]hmC, fC/[^{15}N$_2$]fC, caC/[^{15}N$_2$]caC, 8-oxo-G/[^{15}N$_5$]8-oxo-G, hmU/[D$_2$]hmU, fU/[^{15}N$_2$]fU. These were obtained by applying the compound-dependent parameters summarized in Supplementary Table 4. For hmU the MS/MS transition 257→214 was used. Depicted are the means of five technical replicates of one sample batch. Error bars reflect ± s.d. Linearity was given across the following compound amounts in 29 μL injection volume: 10.63–7751 pmol C; 19.11–4892 pmol G; 104.5 fmol–228.6 pmol mC; 100.8 fmol–73.45 pmol hmC; 2.1–515.0 fmol fC; 2.0–496.6 fmol caC; 2.0–475.7 fmol 8-oxo-G; 6.4–519.3 fmol hmU; 5.7–459.6 fmol fU. The amounts of the labeled internal standards in 29 μL injection volume were as follows: 51.03 pmol [D$_3$]mC; 7.655 pmol [^{15}N$_2$,D$_2$]hmC; 45.6 fmol [^{15}N$_2$]fC; 43.0 fmol [^{15}N$_2$]caC; 108.9 fmol [^{15}N$_5$]8-oxo-G; 96.1 fmol [D$_2$]hmU; 135.0 fmol [^{15}N$_2$]fU.

Supplementary Table 8. Intra-batch-assay and quantification data processing.

	$A_{UV}(G)$	n(G) [pmol]	$A_{UV}(C)$	n(C) [pmol]	C / G [%]	C / N
techn. replicate 1	956	2431	485	2341	96.3	2.02E-01
techn. replicate 2	980	2493	495	2391	95.9	2.01E-01
techn. replicate 3	1043	2654	531	2566	96.7	2.03E-01
techn. replicate 4	979	2492	498	2406	96.5	2.03E-01
techn. replicate 5	982	2498	498	2406	96.3	2.02E-01
techn. mean value	988	2513	501	2422	96.4	2.02E-01
s.d.	33	83	18	85	0.3	6.29E-04
r.s.d [%]	3.3	3.3	3.5	3.5	0.3	0.3
	A(mC)	A([D_3]mC)	A(mC)/ A([D_3]mC)	n(mC) [pmol]	mC / G [%]	mC / N
techn. replicate 1	636358	510501	1.247	91.9	3.78	7.94E-03
techn. replicate 2	678284	512180	1.324	97.7	3.92	8.23E-03
techn. replicate 3	565889	447103	1.266	93.4	3.52	7.39E-03
techn. replicate 4	664238	546597	1.215	89.6	3.60	7.56E-03
techn. replicate 5	719777	570445	1.262	93.1	3.73	7.82E-03
techn. mean value	652909	517365	1.263	93.1	3.71	7.79E-03
s.d.	57206	46591	0.040	2.9	0.16	3.30E-04
r.s.d [%]	8.8	9.0	3.1	3.1	4.2	4.2
	A(hmC)	A([D_2,$^{15}N_2$]hmC)	A(hmC)/ A([D_2,$^{15}N_2$]hmC)	n(hmC) [pmol]	hmC / G [%]	hmC / N
techn. replicate 1	80754	84338	0.958	8.51	0.350	7.35E-04
techn. replicate 2	115869	108774	1.065	9.46	0.380	7.97E-04
techn. replicate 3	264594	269496	0.982	8.72	0.329	6.90E-04
techn. replicate 4	139093	133830	1.039	9.23	0.371	7.78E-04
techn. replicate 5	116163	106074	1.095	9.73	0.389	8.18E-04
techn. mean value	143295	140502	1.028	9.13	0.364	7.64E-04
s.d.	70941	74213	0.057	0.51	0.024	5.12E-05
r.s.d [%]	49.5	52.8	5.6	5.6	6.7	6.7
	A(fC)	A([$^{15}N_2$]fC)	A(fC)/ A([$^{15}N_2$]fC)	n(fC) [pmol]	fC / G [%]	fC / N
techn. replicate 1	163427	31747	5.148	0.199	8.18E-03	1.72E-05
techn. replicate 2	178366	32585	5.474	0.211	8.48E-03	1.78E-05
techn. replicate 3	196827	35504	5.544	0.214	8.07E-03	1.69E-05
techn. replicate 4	193392	33755	5.729	0.221	8.88E-03	1.86E-05
techn. replicate 5	193493	35959	5.381	0.208	8.32E-03	1.75E-05
techn. mean value	185101	33910	5.455	0.211	8.38E-03	1.76E-05
s.d.	14070	1816	0.214	0.008	3.17E-04	6.66E-07
r.s.d [%]	7.6	5.4	3.9	3.9	3.8	3.8
	A(caC)	A([$^{15}N_2$]caC)	A(caC)/ A([$^{15}N_2$]caC)	n(caC) [pmol]	caC / G [%]	caC / N
techn. replicate 1	1444	6666	0.217	0.0104	4.28E-04	8.99E-07
techn. replicate 2	1735	8205	0.211	0.0102	4.08E-04	8.57E-07
techn. replicate 3	2111	9709	0.217	0.0104	3.93E-04	8.26E-07
techn. replicate 4	1985	8301	0.239	0.0114	4.59E-04	9.63E-07
techn. replicate 5	1927	8581	0.225	0.0108	4.31E-04	9.05E-07
techn. mean value	1840	8292	0.222	0.0106	4.24E-04	8.90E-07
s.d.	260	1089	0.011	0.0005	2.48E-05	5.20E-08
r.s.d [%]	14.1	13.1	4.8	4.6	5.8	5.8

	A(hmU)	A([D$_2$]hmU)	A(hmU)/ A([D$_2$]hmU)	n(hmU) [pmol]	hmU / G [%]	hmU / N
techn. replicate 1	894	5799	0.154	0.0090	3.69E-04	7.74E-07
techn. replicate 2	1278	6368	0.201	0.0130	5.23E-04	1.10E-06
techn. replicate 3	1561	7679	0.203	0.0133	5.00E-04	1.05E-06
techn. replicate 4	1695	8148	0.208	0.0137	5.49E-04	1.15E-06
techn. replicate 5	1236	9010	0.137	0.0075	3.00E-04	6.29E-07
techn. mean value	1333	7401	0.181	0.0113	4.48E-04	9.41E-07
s.d.	312	1309	0.033	0.0029	1.08E-04	2.27E-07
r.s.d [%]	23.4	17.7	18.1	25.3	24.2	24.2
	A(fU)	A([^{15}N$_2$]fU)	A(fU)/ A([^{15}N$_2$]fU)	n(fU) [pmol]	fU / G [%]	fU / N
techn. replicate 1	7527	19661	0.383	0.0553	2.27E-03	4.78E-06
techn. replicate 2	8672	22769	0.381	0.0550	2.21E-03	4.63E-06
techn. replicate 3	11884	23930	0.497	0.0711	2.68E-03	5.63E-06
techn. replicate 4	10143	27784	0.365	0.0528	2.12E-03	4.45E-06
techn. replicate 5	14349	30152	0.476	0.0682	2.73E-03	5.73E-06
techn. mean value	10515	24859	0.420	0.0605	2.40E-03	5.04E-06
s.d.	2695	4148	0.061	0.0085	2.82E-04	5.93E-07
r.s.d [%]	25.6	16.7	14.5	14.0	11.7	11.7
	A(8-oxo-G)	A([^{15}N$_5$]8-oxo-G)	A(8-oxo-G)/ A([^{15}N$_5$]8-oxo-G)	n(8-oxo-G) [pmol]	8oxo-G / G [%]	8oxoG / N
techn. replicate 1	237655	245919	0.966	0.113	4.63E-03	9.73E-06
techn. replicate 2	256991	266182	0.965	0.113	4.51E-03	9.48E-06
techn. replicate 3	310924	294412	1.056	0.123	4.64E-03	9.74E-06
techn. replicate 4	282245	299479	0.942	0.110	4.41E-03	9.26E-06
techn. replicate 5	327930	305105	1.075	0.125	5.01E-03	1.05E-05
techn. mean value	283149	282219	1.001	0.117	4.64E-03	9.75E-06
s.d.	37187	25225	0.060	0.007	2.28E-04	4.79E-07
r.s.d [%]	13.1	8.9	6.0	6.0	4.9	4.9

Supplementary Note 3: LC-MS/MS quantification results of genomic DNA

Supplementary Table 9. LC-MS/MS quantification results of mESC (WT01, J1 and R1) wild type, knockdown (KD) and knockout cells (related to **Fig. 1c, 2a and 4**). The results of independent biological replicates are shown, the biological mean values / N and the biological standard deviation (± s.d.). n.d. = not detected.

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (WT01) 1	2.01E-01	8.08E-03	6.53E-04	9.35E-06	6.01E-07	6.94E-07	3.51E-06	6.75E-06
mESC (WT01) 2	2.02E-01	7.71E-03	6.41E-04	9.81E-06	5.90E-07	7.04E-07	2.52E-06	5.95E-06
mESC (WT01) 3	2.02E-01	7.78E-03	6.82E-04	9.26E-06	5.24E-07	4.81E-07	1.60E-06	4.02E-06
mESC (WT01) 4	2.01E-01	8.17E-03	6.92E-04	1.03E-05	6.44E-07	4.73E-07	2.17E-06	4.90E-06
mESC (WT01) 5	2.02E-01	6.72E-03	8.63E-04	1.36E-05	7.25E-07	3.83E-07	4.34E-06	5.76E-06
mESC (WT01) 6	2.03E-01	6.43E-03	7.70E-04	1.32E-05	3.81E-07	5.14E-07	1.30E-06	3.17E-06
mESC (WT01) 7	2.03E-01	6.10E-03	8.18E-04	1.08E-05	6.77E-07	4.78E-07	1.03E-06	3.47E-06
biol. mean value	2.02E-01	7.28E-03	7.31E-04	1.09E-05	5.92E-07	5.32E-07	2.35E-06	4.86E-06
± s.d.	7.75E-04	8.48E-04	8.63E-05	1.81E-06	1.13E-07	1.21E-07	1.21E-06	1.36E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
Tet1 KD (WT01) 1	2.01E-01	8.99E-03	2.76E-04	3.31E-06	2.21E-07	1.42E-07	7.98E-07	3.21E-06
Tet1 KD (WT01) 2	2.02E-01	7.77E-03	1.39E-04	2.53E-06	2.20E-07	6.44E-08	2.00E-06	3.73E-06
Tet1 KD (WT01) 3	2.03E-01	6.37E-03	2.36E-04	6.51E-06	2.07E-07	3.07E-07	4.13E-06	7.08E-06
biol. mean value	2.02E-01	7.71E-03	2.17E-04	4.11E-06	2.16E-07	1.71E-07	2.31E-06	4.67E-06
± s.d.	1.33E-03	1.31E-03	7.03E-05	2.11E-06	7.87E-09	1.24E-07	1.69E-06	2.10E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
Tet2 KD (WT01) 1	2.00E-01	9.41E-03	4.45E-04	6.12E-06	2.38E-07	9.51E-08	5.18E-07	2.82E-06
Tet2 KD (WT01) 2	2.02E-01	7.25E-03	3.38E-04	4.52E-06	9.02E-08	3.23E-07	6.01E-07	1.78E-06
Tet2 KD (WT01) 3	2.02E-01	7.36E-03	3.98E-04	6.25E-06	3.36E-07	1.21E-07	2.50E-06	5.21E-06
Tet2 KD (WT01) 4	2.02E-01	7.69E-03	3.90E-04	6.78E-06	3.37E-07	n.d.	1.16E-06	6.52E-06
biol. mean value	2.02E-01	7.93E-03	3.93E-04	5.92E-06	2.50E-07	1.80E-07	1.20E-06	4.08E-06
± s.d.	1.04E-03	1.01E-03	4.36E-05	9.73E-07	1.16E-07	1.25E-07	9.16E-07	2.17E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (J1) 1	2.04E-01	5.53E-03	3.61E-04	2.18E-06	7.95E-07	1.48E-06	3.98E-06	6.68E-06
mESC (J1) 2	2.04E-01	5.63E-03	3.32E-04	2.17E-06	9.30E-07	1.41E-06	4.19E-06	6.64E-06
biol. mean value	2.04E-01	5.58E-03	3.47E-04	2.17E-06	8.62E-07	1.44E-06	4.08E-06	6.66E-06
± s.d.	4.96E-05	6.97E-05	2.06E-05	8.94E-09	9.55E-08	4.91E-08	1.51E-07	2.93E-08

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
DNMT1 -/- (J1) 1	2.07E-01	2.72E-03	1.80E-04	1.36E-06	3.10E-07	1.68E-06	4.56E-06	7.22E-06
DNMT1 -/- (J1) 2	2.07E-01	2.71E-03	1.77E-04	1.55E-06	2.94E-07	1.56E-06	3.90E-06	6.73E-06
biol. mean value	2.07E-01	2.72E-03	1.79E-04	1.46E-06	3.02E-07	1.62E-06	4.23E-06	6.97E-06
± s.d.	5.80E-06	3.35E-06	2.28E-06	1.31E-07	1.11E-08	8.47E-08	4.73E-07	3.45E-07

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
DNMT3ab -/- (J1) 1	2.10E-01	4.12E-04	3.01E-05	2.46E-07	n.d.	8.02E-07	6.25E-07	2.68E-06
DNMT3ab -/- (J1) 2	2.10E-01	3.76E-04	3.54E-05	5.67E-07	n.d.	2.60E-06	7.86E-07	5.73E-06
DNMT3ab -/- (J1) 3	2.10E-01	3.81E-04	3.03E-05	3.71E-07	n.d.	2.27E-06	3.94E-07	2.32E-06
DNMT3ab -/- (J1) 4	2.10E-01	2.67E-04	2.67E-05	4.52E-07	n.d.	1.71E-06	4.50E-07	1.79E-06
DNMT3ab -/- (J1) 5	2.10E-01	2.50E-04	1.07E-05	3.78E-07	n.d.	7.41E-07	4.51E-06	7.68E-06
DNMT3ab -/- (J1) 6	2.10E-01	2.32E-04	1.60E-05	2.64E-07	n.d.	5.28E-07	2.87E-06	5.50E-06
biol. mean value	2.10E-01	3.20E-04	2.48E-05	3.80E-07		1.44E-06	1.61E-06	4.28E-06
± s.d.	9.60E-05	7.83E-05	9.51E-06	1.19E-07		8.76E-07	1.70E-06	2.35E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (R1) 1	2.01E-01	8.37E-03	2.00E-04	1.04E-06	n.d.	1.52E-06	2.56E-06	1.21E-05
mESC (R1) 2	2.03E-01	6.76E-03	3.30E-04	2.00E-06	3.03E-07	9.23E-07	1.20E-06	n.d.
mESC (R1) 3	2.01E-01	8.68E-03	2.70E-04	2.12E-06	n.d.	1.77E-06	4.67E-06	1.05E-05
biol. mean value	2.02E-01	7.94E-03	2.67E-04	1.72E-06	3.03E-07	1.40E-06	2.81E-06	8.42E-06
± s.d.	9.92E-04	1.03E-03	6.52E-05	5.91E-07		4.35E-07	1.75E-06	5.00E-06

Supplementary Table 10. LC-MS/MS quantification results of different mouse organs of 3 months old wild type individuals (*n*) (related to **Fig. 1d**, **2** and **Supplementary Fig. 2**). Compiled are mean values / N obtained from three independent technical replicates and the standard deviation (± s.d.).

n	organ	DNA isolation	C / N		mC / N		hmC / N		fC / N	
			techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.
1	cerebellum		2.01E-01	2.30E-04	8.06E-03	2.40E-04	6.59E-04	1.08E-05	3.45E-07	3.14E-08
1	cerebellum	BHT, + Desf.,	2.01E-01	2.67E-04	8.27E-03	2.59E-04	7.37E-04	1.76E-05	2.86E-07	1.30E-08
1	cerebellum	BHT, + Desf., +THU	2.01E-01	3.66E-04	8.52E-03	3.45E-04	7.29E-04	2.14E-05	2.97E-07	2.03E-08
2	cerebellum	BHT, + Desf., +THU	2.00E-01	3.92E-04	9.81E-03	2.80E-04	6.37E-04	4.00E-05	2.73E-07	1.16E-08
3	cerebellum	BHT, + Desf., +THU	2.00E-01	3.95E-04	8.95E-03	3.45E-04	6.90E-04	2.91E-05	2.61E-07	1.57E-08
4	cerebellum	BHT, + Desf., +THU	2.00E-01	4.55E-04	9.46E-03	3.60E-04	6.69E-04	2.45E-05	2.59E-07	3.89E-09
1	kidney		2.01E-01	1.72E-04	8.13E-03	1.59E-04	4.37E-04	2.79E-05	2.27E-07	1.10E-08
1	kidney	BHT, + Desf.	2.02E-01	3.38E-04	7.79E-03	3.29E-04	4.33E-04	8.84E-06	2.25E-07	2.75E-08
1	kidney	BHT, + Desf., +THU	2.02E-01	1.70E-04	7.71E-03	1.76E-04	4.27E-04	2.15E-05	2.11E-07	1.19E-08
2	kidney	BHT, + Desf., +THU	2.01E-01	4.44E-04	8.59E-03	3.75E-04	3.78E-04	1.35E-05	1.86E-07	1.43E-08
3	kidney	BHT, + Desf., +THU	2.01E-01	1.46E-04	8.20E-03	1.02E-04	3.73E-04	1.67E-05	1.88E-07	8.82E-09
4	kidney	BHT, + Desf., +THU	2.01E-01	3.17E-04	8.30E-03	2.45E-04	3.79E-04	2.06E-05	1.94E-07	2.45E-08
1	cortex	BHT, + Desf., +THU	2.00E-01	4.48E-04	9.06E-03	3.37E-04	1.12E-03	7.21E-05	4.09E-07	4.17E-08
2	cortex	BHT, + Desf., +THU	1.99E-01	2.05E-04	9.29E-03	1.48E-04	1.37E-03	2.05E-05	4.34E-07	1.14E-08
3	cortex	BHT, + Desf., +THU	2.00E-01	4.65E-04	9.23E-03	3.73E-04	1.14E-03	7.48E-06	4.24E-07	1.92E-08
1	hippocampus	BHT, + Desf., +THU	1.98E-01	1.88E-04	1.08E-02	2.11E-04	1.56E-03	6.83E-05	4.49E-07	2.82E-08
2	hippocampus	BHT, + Desf., +THU	1.98E-01	2.13E-04	1.08E-02	1.73E-04	1.55E-03	3.08E-05	7.08E-07	4.67E-08
3	hippocampus	BHT, + Desf., +THU	1.97E-01	1.78E-04	1.09E-02	1.44E-04	1.71E-03	2.66E-05	3.89E-07	2.66E-08
1	heart	BHT, + Desf., +THU	2.02E-01	3.25E-04	7.96E-03	2.87E-04	4.08E-04	3.32E-05	1.85E-07	1.66E-08
2	heart	BHT, + Desf., +THU	2.02E-01	1.68E-04	7.87E-03	1.31E-04	3.84E-04	2.30E-05	1.48E-07	2.64E-09
3	heart	BHT, + Desf., +THU	2.01E-01	4.25E-04	8.07E-03	3.42E-04	4.56E-04	4.39E-05	1.41E-07	1.35E-08
1	liver	BHT, + Desf., +THU	2.01E-01	2.32E-04	8.35E-03	1.92E-04	2.44E-04	3.24E-06	1.66E-07	2.23E-08
2	liver	BHT, + Desf., +THU	2.02E-01	3.16E-04	8.10E-03	2.61E-04	2.73E-04	8.57E-06	1.78E-07	2.26E-08
3	liver	BHT, + Desf., +THU	2.01E-01	3.04E-04	8.42E-03	2.39E-04	2.82E-04	3.44E-05	1.81E-07	2.91E-08

n	organ	DNA isolation	hmU / N		fU / N		8-oxo-G / N	
			techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.
1	cerebellum		7.27E-08		2.03E-06	2.29E-07	5.09E-06	1.92E-07
1	cerebellum	BHT, + Desf.,	n.d.		6.89E-07	2.65E-08	2.96E-06	6.04E-08
1	cerebellum	BHT, + Desf., +THU	n.d.		8.05E-07	7.34E-08	3.30E-06	3.79E-08
2	cerebellum	BHT, + Desf., +THU	1.30E-07	7.94E-08	1.02E-06	4.83E-08	3.86E-06	2.70E-07
3	cerebellum	BHT, + Desf., +THU	6.01E-08	3.67E-08	9.33E-07	9.52E-08	3.66E-06	2.26E-07
4	cerebellum	BHT, + Desf., +THU	1.07E-07	7.00E-08	1.03E-06	8.63E-08	4.67E-06	5.69E-07
1	kidney		n.d.		7.58E-07	7.38E-08	3.66E-06	2.63E-07
1	kidney	BHT, + Desf.	n.d.		7.00E-07	5.30E-08	3.24E-06	3.73E-07
1	kidney	BHT, + Desf., +THU	n.d.		7.16E-07	1.73E-08	3.48E-06	2.63E-07
2	kidney	BHT, + Desf., +THU	1.08E-07	7.89E-08	1.08E-06	1.45E-07	3.73E-06	1.42E-07
3	kidney	BHT, + Desf., +THU	1.29E-07	9.06E-08	1.17E-06	1.67E-07	3.91E-06	5.99E-08
4	kidney	BHT, + Desf., +THU	n.d.		9.34E-07	6.36E-08	3.65E-06	1.57E-07
1	cortex	BHT, + Desf., +THU	1.46E-07	6.19E-09	1.28E-06	1.22E-07	2.51E-06	3.74E-08
2	cortex	BHT, + Desf., +THU	8.24E-08	2.81E-08	1.30E-06	7.35E-08	3.13E-06	1.16E-07
3	cortex	BHT, + Desf., +THU	2.05E-07	5.83E-08	1.43E-06	3.72E-08	3.58E-06	2.83E-07
1	hippocampus	BHT, + Desf., +THU	2.74E-07	6.06E-08	2.18E-06	1.29E-07	6.30E-06	2.17E-07
2	hippocampus	BHT, + Desf., +THU	6.00E-07	5.50E-08	3.20E-06	3.46E-07	1.31E-05	7.69E-07

328

3	hippocampus	BHT, + Desf., +THU	4.68E-08	1.85E-08	9.85E-07	2.26E-07	5.03E-06	8.96E-08
1	heart	BHT, + Desf., +THU	4.14E-07	7.28E-08	1.97E-06	3.46E-07	7.85E-06	1.69E-06
2	heart	BHT, + Desf., +THU	9.81E-08	4.23E-09	1.15E-06	7.63E-08	2.48E-06	1.18E-07
3	heart	BHT, + Desf., +THU	1.04E-07	3.36E-08	1.55E-06	7.55E-08	3.14E-06	2.27E-07
1	liver	BHT, + Desf., +THU	n.d.		1.13E-06	9.97E-08	2.69E-06	3.93E-07
2	liver	BHT, + Desf., +THU	1.48E-07	8.67E-08	1.09E-06	1.96E-07	2.43E-06	9.41E-08
3	liver	BHT, + Desf., +THU	2.34E-08		1.05E-06	1.18E-07	2.39E-06	4.69E-08

Supplementary Table 11. LC-MS/MS quantification results of $n = 6$ independent mESC differentiation experiments without growth factors (related to **Fig. 5a**,**b**). Compiled are mean values / N obtained from three independent technical measurements and their standard deviation (s.d.). K = C57Bl6/129 derived mES cell line.

n	Sample	C / N		mC / N		hmC / N		fC / N		caC / N	
	mESC diff. time	techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.
1	K, t= 0h	2.06E-01	1.47E-04	3.79E-03	1.52E-04	4.25E-04	8.84E-06	1.68E-05	9.20E-07	1.21E-06	5.80E-08
1	K, t= 8h	2.06E-01	9.23E-05	3.43E-03	9.96E-05	5.33E-04	1.27E-05	1.71E-05	7.46E-07	1.09E-06	2.97E-08
1	K, t= 16h	2.06E-01	1.07E-04	3.88E-03	1.15E-04	4.93E-04	9.98E-06	6.69E-06	2.33E-07	3.92E-07	4.81E-08
2	R1, t= 0h	2.08E-01	1.38E-04	1.87E-03	1.05E-04	3.89E-04	3.33E-05	2.17E-05	5.00E-07	1.52E-06	6.10E-08
2	R1, t= 8h	2.08E-01	2.55E-05	1.60E-03	1.73E-05	5.50E-04	1.14E-05	2.91E-05	1.12E-06	1.91E-06	3.30E-08
2	R1, t= 16h	2.08E-01	7.54E-05	1.74E-03	7.53E-05	5.87E-04	1.28E-05	1.89E-05	4.48E-07	1.04E-06	2.38E-08
3	K, t= 0h	2.05E-01	6.83E-05	4.24E-03	6.36E-05	3.84E-04	6.03E-06	9.95E-06	1.46E-07	1.74E-07	1.70E-08
3	K, t= 8h	2.05E-01	7.08E-05	4.08E-03	7.14E-05	4.85E-04	5.17E-06	1.47E-05	7.35E-07	2.60E-07	1.84E-08
3	K, t= 24h	2.05E-01	5.00E-05	4.65E-03	5.74E-05	4.48E-04	7.86E-06	4.00E-06	1.67E-07	8.74E-08	2.02E-08
4	R1, t= 0h	2.07E-01	6.37E-05	2.49E-03	5.86E-05	4.61E-04	1.43E-05	1.71E-05	4.28E-07	3.25E-07	4.80E-08
4	R1, t= 8h	2.07E-01	5.45E-05	2.27E-03	4.25E-05	5.53E-04	1.69E-05	2.63E-05	8.05E-07	5.71E-07	2.45E-08
4	R1, t= 24h	2.06E-01	5.12E-05	3.00E-03	5.28E-05	5.93E-04	7.13E-06	8.39E-06	2.73E-07	1.74E-07	2.04E-08
5	K, t=0 h	2.04E-01	1.02E-04	5.38E-03	1.04E-04	3.89E-04	4.27E-06	7.66E-06	2.13E-07	2.29E-07	1.73E-08
5	K, t=8 h	2.04E-01	1.74E-04	5.10E-03	1.72E-04	4.65E-04	4.64E-06	1.01E-05	5.36E-07	2.37E-07	8.38E-10
5	K, t=16 h	2.04E-01	2.04E-04	5.88E-03	2.09E-04	5.22E-04	9.13E-06	4.40E-06	2.13E-07	1.92E-07	1.34E-08
5	K, t=24 h	2.04E-01	1.84E-04	5.40E-03	1.79E-04	4.01E-04	6.47E-06	2.06E-06	1.87E-07	1.00E-07	1.16E-08
5	K, t=40 h	2.04E-01	1.71E-04	5.72E-03	1.71E-04	3.00E-04	1.20E-06	1.20E-06	5.32E-08	7.93E-08	
6	R1, t=0 h	2.06E-01	2.16E-05	3.56E-03	3.01E-05	5.85E-04	1.80E-05	2.08E-05	4.72E-07	5.49E-07	5.15E-08
6	R1, t=8 h	2.05E-01	9.98E-05	3.83E-03	8.48E-05	7.63E-04	1.51E-05	2.05E-05	5.32E-07	8.74E-07	9.94E-09
6	R1, t=16 h	2.06E-01	4.53E-05	3.44E-03	3.95E-05	6.89E-04	7.42E-06	1.20E-05	5.65E-07	3.03E-07	8.59E-09
6	R1, t=24 h	2.06E-01	1.19E-04	3.57E-03	8.65E-05	6.16E-04	3.52E-05	7.18E-06	2.03E-07	1.55E-07	2.87E-08
6	R1, t=40 h	2.05E-01	6.38E-06	4.95E-03	1.56E-05	4.42E-04	1.11E-05	1.97E-06	4.48E-08	9.46E-08	1.84E-08

n	Sample	hmU / N		fU / N		8-oxo-G / N	
	mESC diff. time	techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.
1	K, t= 0h	3.75E-07	1.67E-09	3.89E-06	2.73E-07	6.66E-06	2.74E-07
1	K, t= 8h	1.08E-06	1.57E-07	4.83E-06	3.45E-07	7.79E-06	3.05E-07
1	K, t= 16h	1.01E-06	1.76E-07	5.11E-06	5.69E-07	8.07E-06	5.57E-07
2	R1, t= 0h	1.14E-06	3.65E-07	3.82E-06	4.10E-07	6.70E-06	1.72E-07
2	R1, t= 8h	2.97E-06	4.13E-07	5.23E-06	9.97E-08	9.08E-06	5.89E-07
2	R1, t= 16h	3.75E-06	6.55E-07	5.15E-06	5.72E-07	8.81E-06	5.37E-07
3	K, t= 0h	8.76E-07	2.02E-08	8.86E-06	9.44E-07	1.35E-05	9.84E-07
3	K, t= 8h	1.29E-06	1.76E-07	6.90E-06	9.81E-07	1.00E-05	6.53E-07
3	K, t= 24h	6.55E-07	5.89E-08	6.72E-06	7.12E-07	9.73E-06	5.55E-07

4	R1, t= 0h	1.22E-06	1.58E-07	8.62E-06	7.83E-07	1.26E-05	8.90E-08
4	R1, t= 8h	1.91E-06	2.69E-07	6.02E-06	4.56E-07	8.05E-06	2.35E-07
4	R1, t= 24h	1.82E-06	3.67E-07	6.60E-06	6.25E-07	1.06E-05	6.27E-07
5	K, t=0 h	9.98E-07	2.95E-07	7.04E-06	1.35E-06	9.98E-06	9.08E-07
5	K, t=8 h	2.97E-06	2.62E-07	6.89E-06	5.67E-07	9.64E-06	5.85E-07
5	K, t=16 h	1.43E-06	3.46E-07	2.71E-06	1.92E-07	3.82E-06	3.04E-07
5	K, t=24 h	8.01E-07	2.16E-07	6.60E-06	5.93E-07	9.72E-06	1.33E-07
5	K, t=40 h	8.30E-07	3.44E-08	7.78E-06	1.31E-07	1.13E-05	3.07E-07
6	R1, t=0 h	1.85E-06	3.00E-07	8.62E-06	1.94E-07	1.12E-05	7.13E-07
6	R1, t=8 h	3.77E-06	6.22E-07	4.01E-06	7.44E-07	6.02E-06	1.41E-06
6	R1, t=16 h	3.94E-06	8.01E-08	5.93E-06	3.24E-07	7.61E-06	2.01E-07
6	R1, t=24 h	3.84E-06	3.90E-07	8.92E-06	1.03E-06	1.23E-05	2.83E-07
6	R1, t=40 h	1.84E-06	9.73E-08	8.44E-06	2.94E-07	1.26E-05	1.02E-06

Supplementary Table 12. Relative modification levels of combined data sets from differentiation (0–40 h) of R1 and C57Bl6/129-derived mESCs without growth factors (related to **Fig. 5a** and **Supplementary Fig. 8**). In order to obtain these, the absolute modification levels of t = 0 h time points compiled in **Supplementary Table 11** were set as 1 and the modification levels of later time points respectively related to these. Summarized are the biological mean values at each differentiation time point and the standard deviation (± s.d.).

	relative C / N		relative mC / N		relative hmC / N	
time	biol. mean	± s.d.	biol. mean	± s.d.	biol. mean	± s.d.
t = 0 h	1.00000E+00		1.000E+00		1.000E+00	
t = 8 h	1.00024E+00	1.2480E-03	9.435E-01	7.579E-02	1.270E+00	8.082E-02
t = 16 h	9.98952E-01	1.4308E-03	1.004E+00	7.157E-02	1.296E+00	1.629E-01
t = 24 h	9.98556E-01	1.4742E-03	1.077E+00	9.513E-02	1.133E+00	1.176E-01
t = 40 h	9.96347E-01	3.4067E-03	1.228E+00	2.314E-01	7.634E-01	1.034E-02

	relative fC / N		relative caC / N	
time	biol. mean	± s.d.	biol. mean	± s.d.
t = 0 h	1.000E+00		1.000E+00	
t = 8 h	1.280E+00	2.318E-01	1.339E+00	3.329E-01
t = 16 h	6.051E-01	1.980E-01	5.991E-01	2.175E-01
t = 24 h	3.769E-01	9.373E-02	4.393E-01	1.122E-01
t = 40 h	1.255E-01	4.352E-02	2.589E-01	1.225E-01

	relative hmU / N		relative fU / N		relative 8-oxo-G / N	
time	biol. mean	± s.d.	biol. mean	± s.d.	biol. mean	± s.d.
t = 0 h	1.000E+00		1.000E+00		1.000E+00	
t = 8 h	2.258E+00	6.597E-01	9.221E-01	3.417E-01	9.024E-01	3.178E-01
t = 16 h	2.390E+00	8.012E-01	9.337E-01	4.755E-01	8.978E-01	4.410E-01
t = 24 h	1.278E+00	6.283E-01	8.742E-01	1.357E-01	9.086E-01	1.623E-01
t = 40 h	9.129E-01	1.155E-01	1.043E+00	8.856E-02	1.129E+00	3.665E-04

Supplementary Table 13. LC-MS/MS quantification results of mESC differentiation with the growth factors FGF-2 and ActA (related to **Supplementary Fig. 9**). Modified nucleosides / N are given as mean values ± s.d. of three independent technical replicates.

Sample	C / N		mC / N		hmC / N		fC / N	
EpiLC diff. time	techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.
t= 0h	2.096E-01	1.40E-05	3.92E-04	1.35E-05	4.57E-05	6.48E-07	2.88E-07	4.52E-09
t= 12h	2.093E-01	1.01E-05	5.80E-04	4.94E-06	6.85E-05	5.47E-06	1.65E-06	1.34E-07
t= 24h	2.074E-01	1.17E-04	2.29E-03	1.02E-04	2.60E-04	1.41E-05	4.45E-06	6.30E-09
t= 36h	2.044E-01	9.71E-05	4.96E-03	7.38E-05	5.81E-04	2.30E-05	6.16E-06	5.27E-08
t= 48h	2.020E-01	8.85E-05	7.23E-03	8.83E-05	7.53E-04	8.90E-07	4.92E-06	1.31E-07

Sample	hmU / N		fU / N		8-oxo-G / N	
EpiLC diff. time	techn. mean	± s.d.	techn. mean	techn. mean	± s.d.	techn. mean
t= 0h	4.76E-07	7.05E-08	1.87E-06	9.10E-08	4.00E-06	1.10E-07
t= 12h	2.01E-06	1.90E-07	3.91E-06	7.84E-07	9.80E-06	3.07E-07
t= 24h	2.78E-06	7.20E-07	2.42E-06	3.65E-07	5.19E-06	1.45E-07
t= 36h	2.10E-06	5.82E-07	2.20E-06	1.51E-07	5.42E-06	4.12E-07
t= 48h	1.90E-06	1.03E-07	3.74E-06	3.74E-07	6.69E-06	3.08E-07

Supplementary Table 14. HEK-293T wild type vs. HEK + Tet1cm vs. HEK + Tet1cd (related to **Supplementary Fig. 7a**). Modified nucleosides / N are given as mean values of three independent technical replicates.

Nucleosides	HEK-293T wild type		HEK + Tet1cm		HEK + Tet1cd	
	techn. mean	± s.d.	techn. mean	± s.d.	techn. mean	± s.d.
C / N	2.01E-01	9.67E-05	2.02E-01	1.09E-04	2.03E-01	3.91E-04
mC / N	8.80E-03	9.62E-05	8.04E-03	1.09E-04	4.18E-03	3.93E-05
hmC / N	3.39E-05	5.44E-07	4.95E-05	1.04E-06	2.21E-03	9.42E-05
fC / N	2.43E-07	2.41E-09	2.18E-07	1.17E-08	2.48E-04	5.96E-06
caC / N	n.d.		n.d.		1.29E-04	3.35E-06
hmU / N	7.21E-07	5.93E-08	1.51E-07	9.07E-09	4.24E-05	2.51E-06
fU / N	4.89E-06	1.59E-07	1.42E-06	6.66E-08	8.79E-06	1.07E-07
8-oxo-G / N	1.06E-05	1.14E-07	6.83E-06	1.47E-07	8.31E-06	9.03E-07

Supplementary Table 15. DNA modification levels of Tet1 *in vitro* assay (related to **Supplementary Fig. 7b**). Plasmid DNA with full CpG methylation was treated with commercially available Tet1.

plasmid	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
untreated	4.8E-02	n.d.	n.d.	n.d.	n.d.	4.4E-06	8.6E-06
untreated	4.6E-02	n.d.	n.d.	n.d.	n.d.	4.0E-06	9.4E-06
+Tet1	7.6E-04	4.8E-03	5.0E-03	9.0E-03	3.8E-04	1.5E-04	4.0E-05
+Tet1	6.5E-04	4.7E-03	4.8E-03	9.3E-03	4.4E-04	1.6E-04	3.7E-05
−Tet1	4.8E-02	n.d.	3.5E-05	n.d.	1.6E-05	2.5E-04	4.0E-05
−Tet1	4.7E-02	n.d.	3.5E-05	n.d.	1.2E-05	2.3E-04	4.5E-05

Supplementary Table 16. Effect of Smug1 depletion on modification levels in mESCs (R1) and effect of TDG and SMUG1 depletion on modification levels in HEK-293T cells overexpressing Tet1cd (related to **Supplementary Fig. 5**). Percent change values for modified nucleosides of cells treated with esiRNA (targeting Smug1/SMUG1 or TDG) with respect to unrelated control esiRNA. The absolute modification content of HEK-293T cells was normalized based on Tet1cd expression levels (determined by TECAN reading). The percent change is given as a mean value of three independent technical replicates.

	mESC Smug1 KD		HEK-293T + Tet1cd / TDG KD		HEK-293T + Tet1cd / SMUG1 KD	
Nucleosides	**Percent change**	**± s.d.**	**Percent change**	**± s.d.**	**Percent change**	**± s.d.**
hmC	-0.15	4.07	-2.22	2.61	-3.43	2.24
fC	-1.26	3.29	38.15	1.25	0.77	2.05
caC	14.04	29.24	33.51	2.72	6.59	3.24
hmU	37.14	9.47	-10.41	4.74	46.46	5.65
fU	71.00	15.73	-4.36	2.34	22.66	1.91
8-oxo-G	17.27	14.26	-5.36	2.04	-6.97	1.77

Supplementary Table 17. HEK-293T wild type, HEK with Tet1cd-overexpression, Tet1cd/Uhrf1 co-overexpression or Tet1cd/Uhrf2 co-overexpression (related to **Supplementary Fig. 12**). Modified nucleosides / N are given as mean values of three independent technical replicates ± s.d.

	HEK-293T WT		HEK + Tet1cd		HEK + Tet1cd + Uhrf1		HEK + Tet1cd + Uhrf2	
Nucleosides	**techn. mean**	**± s.d.**	**techn. mean**	**± s.d.**	**techn. mean**	**± s.d.**	**techn. mean**	**± s.d.**
C / N	2.04E-01	2.97E-05	2.04E-01	1.24E-04	2.05E-01	8.32E-05	2.04E-01	1.15E-04
mC / N	6.01E-03	2.74E-05	5.29E-03	1.22E-04	4.63E-03	8.93E-05	5.11E-03	1.15E-04
hmC / N	2.86E-05	1.57E-06	4.29E-04	1.85E-05	3.31E-04	1.93E-05	5.13E-04	9.85E-06
fC / N	4.51E-07	5.21E-08	5.97E-05	3.43E-06	7.21E-05	4.24E-06	1.39E-04	5.61E-06
caC / N	1.77E-07	2.67E-09	2.32E-05	8.50E-07	2.23E-04	1.84E-05	1.66E-04	1.21E-06
hmU / N	8.18E-07	8.89E-08	2.55E-06	6.26E-08	1.27E-05	4.30E-07	6.74E-06	1.59E-06
fU / N	7.60E-06	3.39E-07	5.40E-06	5.26E-07	5.64E-06	8.42E-07	1.22E-05	2.10E-06
8-oxo-G / N	1.27E-05	7.08E-07	9.23E-06	3.22E-07	6.93E-06	1.09E-06	1.62E-05	8.43E-07

Supplementary Note 4: materials in cell culture

Supplementary Table 18. Overexpression plasmids and esiRNAs used in HEK-293T cell experiments.

Figure	Experiment	Sample	Plasmid DNA	esiRNA
S7a	HEK-293T +/- Tet1xx	HEK + Tet1cd	GFP-Tet1cd (7.5 µg)	x
		HEK + Tet1cm	mCh-Tet1cm (7.5 µg)	x
		Wild type	pCMV6-Cdk5Rap1-v2 (7.5 µg)	x
S5b,S5c	Tet1cd with TDG or SMUG1 KD	HEK + Tet1cd	GFP-Tet1cd (10 µg)	CDK5RAP1 esiRNA (5 µg)
		HEK + Tet1cd with TDG KD	GFP-Tet1cd (10 µg)	TDG esiRNA (5 µg)
		HEK + Tet1cd with SMUG1 KD	GFP-Tet1cd (10 µg)	SMUG1 esiRNA (5 µg)
S12	HEK-293T	Tet1cd	GFP-Tet1cd (6 µg)	x
	HEK-293T	Tet1cd + Uhrf1	GFP-Tet1cd, GFP-Uhrf1[1] (each 6 µg)	x
	HEK-293T	Tet1cd + Uhrf2	GFP-Tet1cd, GFP-Uhrf2[1] (each 6 µg)	x

Supplementary Table 19. Knockdown (KD) efficiencies by Tet relative to SCR shRNAs.

	FWD	REV	Reference
Tet1	GAGCCTGTTCCTCGATGTGG	CAAACCCACCTGAGGCTGTT	*Ito et. al.*[2]
Tet2	TGTTGTTGTCAGGGTGAGAATC	TCTTGCTTCTGGCAAACTTACA	*Ito et. al.*[2]
actin	AAGGCCAACCGTGAAAAGAT	GTGGTACGACCAGAGGCATAC	This work

Supplementary Table 20. Primers for qPCR analysis of Tet, Dnmt, Tdg and Smug1 of EpiLC differentiation and Smug1 knockdown samples.

	FWD	REV	Reference
Gapdh	CATGGCCTTCCGTGTTCCTA	CTTCACCACCTTCTTGATGTCATC	*Szwagierzcak et al.*[3]
Tet1	CCAGGAAGAGGCGACTACGTT	TTAGTGTTGTGTGAACCTGATTTATTGT	*Szwagierzcak et al.*[3]
Tet2	ACTTCTCTGCTCATTCCCACAGA	TTAGCTCCGACTTCTCGATTGTC	*Szwagierzcak et al.*[3]
Tet3	GAGCACGCCAGAGAAGATCAA	CAGGCTTTGCTGGGACAATC	*Szwagierzcak et al.*[3]
Dnmt1	CCTAGTTCCGTGGCTACGAGGAG	TCTCTCTCCTCTGCAGCCGACTC	This work
Dnmt3a	GCTTTCTTCTCAGCCTCCCT	CCATGCCAAGACTCACCTTC	This work
Dnmt3b	CTGGCACCCTCTTCTTCATT	ATCCATAGTGCCTTGGGACC	This work
Tdg	GTCTGTTCATGTCGGGGCTGAGTGAG	CTGCAGTTTCTGCACCAGGATGCGC	This work
Smug1	CACTGGGGCCTACCCATGA	CTCCCAAGCATAATCCACCG	This work

Supplementary Note 5: correlation analysis results of modification levels

Supplementary Table 21. Correlation analysis of DNA modification levels comparing mouse tissues from three months old individuals (cortex, hippocampus, cerebellum, heart, liver and kidney). Pearson coefficients (p) and significance values (s) are summarized. n=24 independent DNA samples (see **Supplementary Table 10**). Highlighted in gray are strong to very strong correlations ($|p| > 0.7$) with significance levels (s) lower than 0.001 (marked with *). Additionally, moderate correlations ($0.7 > |p| > 0.6$) with significance level lower than 0.001 are highlighted in light gray.

		C	mC	hmC	fC	8-oxo-G	hmU	fU
C	p	1.000	-0.986*	-0.935*	-0.832*	-0.534	-0.466	-0.479
	s		0.000	0.000	0.000	0.007	0.022	0.018
mC	p	-0.986*	1.000	0.863*	0.769*	0.534	0.464	0.460
	s	0.000		0.000	0.000	0.007	0.022	0.024
hmC	p	-0.935*	0.863*	1.000	0.887*	0.481	0.420	0.472
	s	0.000	0.000		0.000	0.017	0.041	0.020
fC	p	-0.832*	0.769*	0.887*	1.000	0.663*	0.586	0.649*
	s	0.000	0.000	0.000		0.000	0.003	0.001
8-oxo-G	p	-0.534	0.534	0.481	0.663*	1.000	0.837*	0.835*
	s	0.007	0.007	0.017	0.000		0.000	0.000
hmU	p	-0.466	0.464	0.420	0.586	0.837*	1.000	0.871*
	s	0.022	0.022	0.041	0.003	0.000		0.000
fU	p	-0.479	0.460	0.472	0.649*	0.835*	0.871*	1.000
	s	0.018	0.024	0.020	0.001	0.000	0.000	

Supplementary Table 22. Correlation analysis of DNA modification levels during early mESC differentiation (0-40 h). Pearson coefficients (p) and significance values (s) are summarized. n=22 independent DNA samples (see Supplementary Table 11). Highlighted in gray are strong to very strong correlations ($|p| > 0.7$) with significance levels (s) lower than 0.001 (marked with *). Additionally, weak correlations of hmC/hmU with mC/C are highlighted in pale pink.

		C	mC	hmC	fC	caC	8-oxo-G	hmU	fU
C	p	1	-0.997*	0.324	0.806*	0.699*	-0.032	0.299	-0.145
	s		0.000	0.141	0.000	0.000	0.889	0.176	0.518
mC	p	-0.997*	1	-0.399	-0.815*	-0.693*	0.057	-0.356	0.160
	s	0.000		0.066	0.000	0.000	0.801	0.104	0.476
hmC	p	0.324	-0.399	1	0.404	0.182	-0.312	0.783*	-0.227
	s	0.141	0.066		0.062	0.417	0.157	0.000	0.309
fC	p	0.806*	-0.815*	0.404	1	0.797*	-0.266	0.303	-0.299
	s	0.000	0.000	0.062		0.000	0.232	0.170	0.177
caC	p	0.699*	-0.693*	0.182	0.797*	1	-0.435	0.165	-0.567
	s	0.000	0.000	0.417	0.000		0.043	0.463	0.006
8-oxo-G	p	-0.032	0.057	-0.312	-0.266	-0.435	1	-0.080	0.959*
	s	0.889	0.801	0.157	0.232	0.043		0.723	0.000
hmU	p	0.299	-0.356	0.783*	0.303	0.165	-0.080	1	-0.022
	s	0.176	0.104	0.000	0.170	0.463	0.723		0.921
fU	p	-0.145	0.160	-0.227	-0.299	-0.567	0.959*	-0.022	1
	s	0.518	0.476	0.309	0.177	0.006	0.000	0.921	

Supplementary References

1. Pichler, G. et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* **112**, 2585-93 (2011).

2. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-3 (2011).

3. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **38**, e181 (2010).

4. Huang, D.W., Sherman, B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.***4**(1):44-57 (2009).