

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie & Pharmazie
der Ludwig-Maximilians-Universität München

Synthese und Untersuchung von epigenetisch relevanten DNA-Modifikationen

René Johannes Rahimoff

aus Moosburg a. d. Isar (Deutschland)

2018

Erklärung

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

Eidesstattliche Erklärung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 16.08.2018

(René Rahimoff)

Dissertation eingereicht am: 24.08.2018.....

1. Berichtstatter: Prof. Dr. Thomas Carell

2. Berichtstatter: Dr. Stefanie Kellner

Datum der mündlichen Prüfung: 26.10.2018.....

„There is a crack in everything, that's how the light gets in.“

Leonard Cohen, Anthem (1992)

Meiner Familie und meinem lieben Freund Florian.

Publikationsliste

Teile dieser Arbeit wurden bereits publiziert oder auf Konferenzen vorgestellt.

Publikationen

- „5-Formylcytosine to cytosine conversion by C-C bond cleavage in vivo“ Katharina Iwan[#], **René Rahimoff**[#], Angie Kirchner[#], Fabio Spada[#], Arne S. Schröder, Olesea Kosmatchev, Shqiponja Ferizaj, Jessica Steinbacher, Edris Parsa, Markus Müller, Thomas Carell, *Nature Chem. Bio.* **2018**, *14*, 72 – 78.
- „5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells“ **René Rahimoff**[#], Olesea Kosmatchev[#], Angie Kirchner[#], Toni Pfaffeneder, Fabio Spada, Victor Brantl, Markus Müller, Thomas Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359 – 10364.
- „Quantitative LC-MS Provides No Evidence for m⁶dA or m⁴dC in the Genome of Mouse Embryonic Stem Cells and Tissues“ Sarah Schiffers, Charlotte Ebert, **René Rahimoff**, Olesea Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, Fabio Spada, Stylianos Michalakis, Jörg Nickelsen, Markus Müller, Thomas Carell, *Angew. Chem. Int. Ed.* **2017**, *56*, 11268 – 11271.
- „Epigenetic Modifications and the Regulation of Gene Expression“ Iacovos N. Michaelides, **René Rahimoff**, Thomas Carell, *GIT Labor Fachzeitschrift* **2016**.

[#]: Diese Autoren haben zu gleichen Teilen beigetragen.

Konferenzbeiträge

- „A Novel Probe for the Detection and Quantification of Abasic Sites in Genomic DNA“, XVIIth Symposium on Nucleic Acid Components (Cesky Krumlov, Tschechische Republik, Juni 2017), mit dem *Nucleic Acids Research Poster Prize* für die beste mündliche Präsentation ausgezeichnet.
- „A Novel Probe for the Detection and Quantification of Abasic Sites in Genomic DNA“, PSL Chemical Biology Symposium (Paris, Frankreich, Dezember 2016).

Weitere Publikationen

- „*Stage Specific Effects of Soluble Copper and Copper Oxide Nanoparticles During Sea Urchin Embryo Development and Their Relation to Intracellular Copper Uptake*“ Christina Torres-Duarte, Karla M. Ramos-Torres, **René Rahimoff**, Gary N. Cherr, *Aquat. Toxicol.* **2017**, *189*, 134 – 141.
- „*Practical Large-Scale Regioselective Zincation of Chromone Using TMPZnCl·LiCl Triggered by the Presence or Absence of MgCl₂*“ Lydia Klier, Dorothee S. Ziegler, **René Rahimoff**, Marc Mosrin, Paul Knochel, *Org. Process Res. Dev.* **2017**, *21*, 660 – 663.

Danksagungen

Meinem Doktorvater *Professor Dr. Thomas Carell* möchte ich hier an erster Stelle danken. In den letzten drei Jahren hatte ich die Möglichkeit, unter deiner Anleitung an einer herausfordernden Themenstellung zu arbeiten und genöß dabei sämtliche Freiheiten und dein ständiges Vertrauen in meine Arbeit und Verantwortung. Außerdem bin ich dir für deine Unterstützung außerhalb wissenschaftlicher Sachverhalte überaus dankbar, ich kann mit Stolz und einem sehr guten Gefühl auf das zusammen Erreichte zurückblicken und freue mich auf alles, was in Zukunft auf mich zukommt.

Dr. Stefanie Kellner danke ich vielmals für die Übernahme des Zweitgutachtens.

Den Mitgliedern der Prüfungskommission, *Professor Dr. Oliver Trapp*, *Prof. Dr. Anja Hoffmann-Röder*, *Prof. Dr. Konstantin Karaghiosoff* und *Dr. Dorian Didier* sei für das Beisitzen der mündlichen Prüfung gedankt.

Markus Müller, dir möchte ich für die vielen angeregten wissenschaftlichen und nicht-wissenschaftlichen Diskussionen, die hilfreichen und kreativen Ratschläge und deine Hilfe in allen möglichen Belangen recht herzlich danken.

Frau *Slava Gärtner* sei an dieser Stelle für die Durchführung aller bürokratischen Aufgaben gedankt.

Kristof Hufnagel danke ich für seine Hilfe und seinen unermüdlichen Einsatz im Arbeitskreis Carell, du warst mir sehr oft eine große Hilfe bei der HPLC-Aufreinigung. *Kerstin Kurz* danke ich ebenfalls für ihren Einsatz im Arbeitskreis, die außerterminlichen Bestellungen von dringend benötigten Chemikalien und den Urwald an Pflanzen, den ich dank dir mittlerweile bei mir zu Hause stehen habe.

Für das gewissenhafte und schnelle Korrekturlesen großer Teile dieser Arbeit bedanke ich mich bei *Angie Kirchner*, *Katharina Iwan*, *Markus Müller* und *Alexander Schön*.

Der Analytikabteilung, im speziellen Herrn *Werner Spahl* (bestechlich mit Schokoriegeln), danke ich für die gewissenhafte und zeitnahe Messung meiner Verbindungen.

Meinen Praktikanten *Johannes Harenberg*, *Julian Schackmann*, *Timo Denk* und *Manuel Loos* danke ich für ihren Enthusiasmus und ihren großen Einsatz während ihrer Bachelorarbeiten oder Forschungspraktika, ihr wart mir eine unglaubliche Hilfe bei der Anfertigung dieser Dissertation. Es ist schön zu sehen,

wo die Reise für euch hingehet oder mittlerweile hinging und ich bin froh, dass ich ein Teil davon sein durfte.

Meinen Masterstudenten *Alexander Schön* und *Simon Veth* sei für ihr Vertrauen, ihre Geduld und ihren unermüdlichen Arbeitswillen während der Anfertigung ihrer Masterarbeiten gedankt. Ich hoffe, ihr habt von mir genauso viel gelernt, wie ich von euch. Es hat mir unglaublich viel Spaß gemacht, euch beide zu betreuen und euren wissenschaftlichen und persönlichen Fortschritt mitzuerleben. Außerdem freut es mich, dass ihr euch beide für eine Promotion im Arbeitskreis Carell entschieden habt. Ich wünsche euch viel Erfolg und alles Gute für eure Zukunft!

Katharina Iwan und *Angie Kirchner*, mit euch beiden war es eine intensive, aber allen voran spannende, aufschlussreiche und spaßige Zeit während der Durchführung unserer Projekte. Euch beide kann man sich als Arbeitskollegen nur wünschen, ich werds' vermissen, danke euch!

Olesea Kosmatchev, *Fabio Spada* und *Sarah Schiffers*, euch dreien sei auch für die Kooperation während verschiedenster Phasen dieser Dissertation vielmals gedankt.

Mit nahezu allen Mitgliedern des Arbeitskreis Carells habe ich während meiner Doktorarbeit eine tolle Zeit erlebt. Ich danke euch für den offenen wissenschaftlichen Austausch, die Unterstützung in sämtlichen Dingen, die Biere und die Riojas (sprich: Rioutscha). Besonders hervorzuheben sind: *Sidney Becker*, *Martin „Razorblade“ Rossa*, *Korbinian Brunner*, *Simon Geiger*, *Edris Parsa*, *Hidenori Okamura*, *Nobuhiro Tago*, *Iacovos Michaelides*, *Stefan Schiesser*, *Florian Kink*, *Matthias Kurz*, *Mirko „Dirkos“ Wagner*, *Michael Stadlmeier*, *Bastien Viverge*, *Samuele Stazzoni*, *Stefano Groce*, *Franziska Traube* und *Arne Schröder*.

Meine Freundinnen und Freunden außerhalb des Arbeitskreises möchte ich an dieser Stelle auch gerne erwähnen. Was ich erlebe, denke und mich als Person ausmacht ist ein Produkt aus allen Erfahrungen, die ich mit euch allen teile und geteilt habe, dafür bin ich euch unendlich dankbar: *Eyke Liegmann*, *Niklas Colling*, *William Moody*, *Finian Vogel*, *Amrei Aufderheide*, *Diana Pereira Moreira*, *Tobias* und *Markus Klaas*, *Andreas Huber*, *Florian Pichlmeier*, *Martina Schernhammer*, *Yanna Löbig*, *David Hammond*, *Kilian Knoll*, *Florian Rott*, *Max Ganiek*, *Marcus Krone* und *Rebecca Sturm*.

Besondere Erwähnung haben meine zwei Telefonjoker und Kletterfreunde *Christoph „Giveher“ Körner* und *Nico „Rainer“ Becker* verdient. Mit euch ist es immer eine Freude in den Bergen oder der

Kletterhalle, danach kann es einem nur gut gehen. Danke für die Ernennung zum vollwertigen El Chorro-Reisemitglied, eine Ehre!

Das Beste kommt natürlich zum Schluss: ich möchte mich bei meiner Familie bedanken, insbesondere bei meiner Oma *Gertraud Pletschacher* für die mehr als wichtigen Lebensweisheiten und Ansichten, die so manchem Großstädter vermutlich schon zu progressiv wären. Den allergrößten Dank verdienen jedoch meine Mama *Clarissa Rahimoff* und mein Papa *Dieter Rahimoff*. Ohne eure Unterstützung, euer Vertrauen und euer Verständnis wäre das Alles für mich irgendwann unmöglich gewesen. Ihr habt das alles möglich gemacht vor knapp 30 Jahren, dazu herzlichen Glückwunsch, ich bin euch unendlich dankbar dafür!

Inhaltsverzeichnis

1	Einleitung	1
1.1	Historisches zu Isotopenverfolgungsexperimenten	1
1.2	DNA und die Vererbarkeit von genetischer Information	4
1.3	Epigenetik	5
1.3.1	Histonmodifikationen zur Kontrolle der Genexpression	6
1.3.2	Genexpressionskontrolle durch mdC als zentralen Schalter	9
1.3.3	Verteilung und Funktion von 5-Hydroxymethyl-2'-desoxycytidin	11
1.3.4	5-Formyl-2'-desoxycytidin und 5-Carboxy-2'-desoxycytidin als Intermediate im epigenetischen Metabolismus	15
1.4	Aktive Demethylierung von mdC	16
1.4.1	Basen-Exzisions-Reparatur gekoppelte aktive Demethylierung	17
1.4.2	Demethylierung über einen C-C-Bindungsbruch	20
2	Aufgabenstellung	24
3	Veröffentlichte Arbeiten	28
3.1	Deformylierung von 5-Formyl-desoxycytidin zu Desoxycytidin in genomischer DNA als alternativer aktiver Demethylierungsweg über einen C-C-Bindungsbruch	28
3.2	Anwendung eines neuartigen Reagenzes zur Quantifizierung von abasischen und β - eliminierten abasischen Stellen in genomischer DNA embryonaler Stammzellen im epige- netischen Kontext	70
3.3	Entwicklung der Synthese eines Nukleosidstandards zur Untersuchung des m ⁶ dA und m ⁴ dC Gehalts in embryonalen Stammzellen von Mäusen	100
4	Unveröffentlichte Arbeiten	124
4.1	Entwicklung und Synthese C5- und C6-deuterierter Cytidine und deren epigenetischer Metabolismus in embryonalen Stammzellen	124
4.1.1	Zellbiologischen Daten der verfütterten Testmoleküle	137
4.1.2	Zusammenfassung und Ausblick	142
4.1.3	Projektbeiträge	143
4.1.4	Synthesevorschriften und Analytik	144
4.2	Synthese eines photoinduzierbaren Hydroxylamin-Testmoleküls zur <i>in cellulo</i> -Markierung von abasischen Stellen	154
4.2.1	Projektbeiträge	162

Inhaltsverzeichnis

4.2.2	Synthesevorschriften, Analytik und experimentelle Details	162
5	Material und Methoden	183
6	Abkürzungsverzeichnis	186
	Literatur	189

Zusammenfassung

Durch epigenetische Kontrolle der Genexpression ist es multizellulären Organismen möglich, hochspezialisierte Zelltypen aus einer gemeinsamen Vorläuferzelle, der embryonalen Stammzelle, zu entwickeln. Im menschlichen Körper unterscheiden sich diese ausdifferenzierten Zellen stark in ihrer Morphologie und Funktion, obwohl sie alle die gleiche DNA als Erbsubstanz besitzen. Daher gibt es eine zweite Informationsebene, welche die gezielte Expression von bestimmten Genen zu bestimmten Zeitpunkten steuert und so eine große Variabilität erzeugen kann. In einem dieser Kontrollszenarien wird beispielsweise die Genexpression durch chemische Modifikation an der C5-Position des 2'-Desoxycytidins (dC) kontrolliert. In diesem Zusammenhang ist schon lange bekannt, dass durch enzymatische Methylierung von dC zu 5-Methyl-2'-desoxycytidin (mdC) in CpG-reichen Regionen im Genom die Aktivität eines Gens abgeschaltet werden kann. Wird diese Modifikation wieder entfernt, so kann das Gen wieder abgelesen und exprimiert werden. Jedoch ist letzterer Vorgang noch nicht eingehend geklärt. Vorallem ist die direkte Entfernung einer Methylgruppe an einem aromatischen Csp²-Zentrum thermodynamisch sehr aufwändig und praktisch schwer vorstellbar. 2009 wurde mit der Entdeckung der durch die Tet-Enzyme vermittelten Erzeugung von 5-Hydroxymethyl-2'-desoxycytidin (hmdC) in embryonalen Stammzellen der genetische Code um die sechste Base erweitert und gleichzeitig wurde deutlich, dass diese DNA-Base eine Funktion bei der aktiven Demethylierung haben könnte. 2011 gelang schließlich die Entdeckung der höher oxidierten Derivate 5-Formyl-2'-desoxycytidin (fdC) und 5-Carboxy-2'-desoxycytidin (cadC) im genetischen Material embryonaler Stammzellen. Auch hier wurde schnell eine potentielle Funktion während der aktiven Demethylierung von mdC zu dC vorgeschlagen.

Der Schwerpunkt der hier vorgelegten Dissertation lag in der Untersuchung der Entfernung der DNA-Modifikationen fdC und cadC im Rahmen der aktiven Demethylierung von mdC zu dC. Im Moment werden für diesen Reaktionsschritt größtenteils zwei verschiedene Mechanismen diskutiert. Zum einen wurde kurz nach der Entdeckung von fdC und cadC *in vitro* gezeigt, dass diese Nukleotide Substrate der *Thymin DNA Glykosylase* (Tdg) sind, welche diese Basen erkennen kann und sie an der glykosidischen Bindung der 2'-Desoxyribose abtrennt. Dadurch entsteht eine abasische Stelle in der DNA, welche im Rahmen der *Base Excision Repair* (BER) prozessiert und aufgefüllt wird. Das Intermediat, welches während der aktiven Demethylierung von mdC zu dC auftritt, sollte in einem ersten Teil dieser Arbeit genauer untersucht werden. Durch Entwicklung und Anwendung einer auf UHPLC-MS/MS basierenden Analyse-methode, bei der mit Hilfe eines Hydroxylamins die abasischen Stellen der genomischen DNA von embryonalen Stammzellen der Maus (mESC) derivatisiert werden, sollten diese kurzlebigen Intermediate zunächst quantifiziert werden. Dabei wurde mit Hilfe stabil-isotopenmarkierter interner Standards eine Isotopenverdünnungsmethode angewandt. Dadurch konnten exakte quantitative Daten zu abasischen-

und β -eliminierten abasischen Stellen erhalten werden, was bis dato mit der etablierten Technik nicht möglich war. Mit dieser Technologie konnte gezeigt werden, dass von dC-stammende, abasische Stellen in embryonalen Stammzellen nicht akkumulieren und, falls eine aktive Demethylierung über die BER abläuft, dieser Vorgang vermutlich in einem geschlossenen Komplex stattfindet, welcher die Exposition der abasischen Stelle nicht zulässt.

Im zweiten Teil dieser Arbeit wurde mit Hilfe von Isotopenverfolgungsexperimenten und der Verfütterung von fluorierten DNA-Nukleosiden erstmals eine direkte Demodifizierung von fdC zu dC in mESCs und anderen somatischen Zelllinien beobachtet. Dazu wurden entweder 2'-(*R*)-Fluor-modifizierte dC Derivate (F-hmdC und F-fdC) oder mit stabilen Isotopen markiertes fdC an Zellen verfüttert. Nach Isolation und Verdau der genomischen DNA auf Nukleosidebene und anschließender UHPLC-MS/MS-Analyse wurde eine direkte Demodifizierung von fdC zu dC festgestellt, welche universell in allen getesteten Zelltypen stattfand und Tet-unabhängig war. Weiterhin wurde festgestellt, dass demodifiziertes fdC im Laufe des Metabolismus wieder zu mdC methyliert wird, womit eine regulative Funktion der Demodifizierung von fdC zu dC während der Genexpressionskontrolle suggeriert wird. Außerdem konnten durch die Entwicklung und Verfütterung verschiedener an C5-Position deuterierter dC-Derivate weitere Hinweise auf die Beteiligung eines direkten C-C-Bindungsbruchs während der aktiven Demethylierung in mESCs gefunden werden. Diese Experimente könnten der Grundstein für die Entdeckung der dafür verantwortlichen biologischen Entität darstellen, welche diese Reaktionen im Rahmen der Genexpression durchführt und somit eine Demethylierung ohne die Erzeugung von potentiell mutagenen, abasischen Stellen ermöglicht.

Neben den an C5-modifizierten dC-Derivaten wurden in den letzten Jahren mehrere Studien veröffentlicht, welche von der Entdeckung vom N6-methylierten 2'-Desoxyadenosin (m^6dA) im genetischen Material von Säugetieren berichteten. Durch Synthese interner Standards beider Modifikationen konnte mit Hilfe der im Arbeitskreis etablierten UHPLC-MS/MS-Methodik gezeigt werden, dass diese Modifikationen, entgegen anderer Berichte, welche sich größtenteils auf immunologische Nachweismethoden stützten, nicht im genetischen Material von mESCs vorhanden sind. Darüber hinaus erhärtet sich eher der Verdacht, dass es sich bei den zuvor berichteten Entdeckungen von m^6dA und m^4dC um Verunreinigungen aus dem genomischen Material von Bakterien handeln könnte, in denen m^6dA und m^4dC feste Bestandteile sind.

Abstract

Epigenetic control of gene expression enables multicellular organisms to develop highly specialized cells that derive from a single progenitor, the embryonic stem cell. Although all of these highly specialized and differentiated cells contain the same genetic material, they differ vastly both regarding their morphology and function. Consequently, a second layer of information must exist that controls the well orchestrated expression of genes at specific time points to generate such a huge variety of cell functions. Gene expression can, for example, be controlled through chemical modification of the C5-position of 2'-deoxycytidine (dC). Methylation of this position to 5-methyl-2'-deoxycytidine (mdC) in CpG-rich sequences is well known to deactivate transcriptional activity. Eradication of this so called epigenetic mark goes in hand with reactivation of gene expression. However, the mechanisms of active removal of the methyl group is still not fully understood, especially since the replacement of an aromatic Csp²-bound methyl group is thermodynamically quite challenging. In 2009, 5-hydroxymethyl-2'-deoxycytidine (hmdC) was found to be a Tet-mediated oxidation product of mdC that is actively generated within the genetic material of mouse embryonic stem cells (mESCs) and neurons. This sixth base of the genetic material was instantly regarded as a possible intermediate of the enigmatic active demethylation pathway. Finally, in 2011, the discovery of the higher oxidized derivatives 5-formyl-2'-deoxycytidine (fdC) and 5-carboxy-2'-deoxycytidine (cadC) in mESCs fueled additional discussions about their potential role as actively generated intermediates during the demethylation of mdC to dC.

The focus of this dissertation was to investigate the underlying mechanisms of active demethylation of mdC to dC. In general, two mechanisms for their removal are discussed of which one is processing of fdC and cadC in the framework of the so called *base excision repair* (BER) pathway. It was shown *in vitro* that *thymine DNA glycosylase* (Tdg) can cleave both fdC and cadC bases at the glycosidic bond of the 2'-deoxysugar, which generates an abasic site that is further processed through BER. Investigation of the abasic site intermediate was the main subject in the first part of this thesis. To this end, the development and application of an UHPLC-MS/MS based method, which is based on a hydroxylamine probe for the derivatization of abasic sites, was performed to analyze the composition of these intermediates in mESCs. Utilization of the isotope-dilution method and addition of isotopically labeled internal standards enabled exact quantification of both abasic and, for the first time also, of β -eliminated abasic sites, the latter are generated through bifunctional glycosylases. Until today, this was not possible with the established derivatization techniques. Finally, it was shown that abasic sites that derive from dC do not accumulate during BER based active reprogramming of the cells and that this process might take place in a closed complex, in which the abasic site intermediate is immediately removed.

In the second part of this thesis, utilizing isotope-tracing experiments together with the feeding of fluorine-labelled nucleosides, a direct *C*5-demodification of fdC to dC was shown to occur in both mESCs and other somatic cells. To enable this, 2'-(*R*)-fluorinated dC-derivatives (F-hmdC and F-fdC) and isotopically labeled fdC were fed to cells and after isolation of the genomic DNA and enzymatic hydrolysis, the constituents of the mixture were analyzed via UHPLC-MS/MS. It was shown that fdC can be converted to dC *in vivo* and that this Tet-independent process takes places in a variety of different cells. Moreover, it was shown that demodified fdC is subsequently remethylated to mdC, suggesting a pathway that acts in a regulative fashion during gene expression. To substantiate this process, a series of *C*5-deuterated dC-probes were synthesized to gain further insight into a possible active demethylation pathway in mESCs, that involves a direct *C-C*-bond cleavage at the *C*5-position of dC. These experiments could represent the foundation to finding the biological entity that is responsible for the active demethylation of mdC to dC, avoiding the generation of potentially mutagenic abasic site intermediates.

Besides *C*5-modified dC derivatives, in the last couple of years scientists reported the discovery of amino-methylated 2'-deoxyadenosin (dA) m⁶dA in mESCs. Synthesis of internal standards of these modifications and utilization of the established UHPLC-MS/MS-methodology showed, that these modifications, in contrast to published reports that mostly rely on immunological detection methods, are no constituents of the genetic material of mESCs. Moreover, it was suggested that m⁶dA and m⁴dC might be bacteria-derived impurities that were introduced into the genome of mESCs artificially and therefore misleadingly interpreted as novel DNA modifications.

1 Einleitung

1.1 Historisches zu Isotopenverfolgungsexperimenten

“Hydrogen, we have Deuterium, you have nothing, carbon, we have ^{13}C , you have ^{11}C with a half-life of 20 minutes, nitrogen, we have ^{15}N , you have nothing and oxygen, we have ^{18}O and you have nothing.“

Mit diesen provokanten Worten wandte sich im Frühling 1938 *Harold C. Urey*, vier Jahre nachdem er mit dem Nobelpreis für Chemie aufgrund seiner Entdeckung des Deuteriums ausgezeichnet wurde, an keinen geringeren als *Ernest Orlando Lawrence of Berkeley*.^[1] Der Geschichte nach soll *Lawrence of Berkeley*, seines Zeichens der Erfinder des Zyklotrons und daher 1939 ebenfalls mit dem Nobelpreis für Physik ausgezeichnet, nach diesem Zusammentreffen seine Forschungsgruppe angewiesen haben, ein radioaktives Isotop mit einer Halbwertszeit zu finden, welche lang genug ist, damit es für biologische Untersuchungen verwendet werden kann. Kurz darauf gelang es *S. Ruben* und *M. D. Kamen* ^{14}C darzustellen, dessen Halbwertszeit sie auf 10^3 - 10^5 Jahre schätzten.^[2]

Aufgrund seiner Anstrengungen und Erfolgen zur Anreicherung der natürlichen stabilen Isotope wie zum Beispiel dem ^{13}C , ^{15}N und ^{18}O kann *Harold C. Urey* zu Recht als einer der Väter von Isotopenverfolgungsexperimenten genannt werden. Eine von *Urey's* Doktorandinnen, *Mildred Cohn*, erinnerte sich an seinen Pioniergeist. So wollte er schon lange, bevor überhaupt irgendjemand an Isotopenverdünnungsmethoden gedacht hatte, das Körpervolumen des Menschen mit H_2^{18}O bestimmen. Mäuse sterben, sobald man in ihrer Ernährung 25% des natürlichen Wassers mit D_2O substituiert. Die Toxizität ist dabei auf einen H/D-Austausch bei den lebenswichtigen Enzymen und den damit einhergehenden, deaktivierenden Isotopeneffekten zurückzuführen. *Urey* war sich hingegen sicher, dass die Verwendung von H_2^{18}O theoretisch in keinsten Art und Weise letal sein sollte und schlug *Mildred Cohn* vor, sie solle Gegenstand des Experiments sein und somit der erste Mensch, dessen Körpervolumen exakt bestimmt werden würde. *Mildred Cohn* lehnte dankend ab, wurde 96 Jahre alt und eine der wichtigsten Pionierinnen auf dem Gebiet der Kernresonanzspektroskopie.

Urey's Labor versorgte unter anderem *Rudolph Schoenheimer*, welcher in den 30er Jahren ebenfalls Professor an der Columbia Universität in New York war, mit verschiedenen stabilen Isotopen. Dadurch war es *Schoenheimer* und seinem Mitarbeiter *David Rittenberg* möglich, bahnbrechende Fortschritte in der Erforschung verschiedenster metabolischer Reaktionswege in Tieren zu untersuchen. 1937 veröffentlichten sie die Ergebnisse einer Studie, bei der sie Mäuse mit einer spezifischen Diät für mehrere Wochen fütterten, die unter anderem die Beimischung von 1.5% D_2O in deren Wasser beinhaltete. Sie fanden dabei heraus, dass die langkettigen Fettsäuren Palmitin und Stearin über einen gewissen Zeitraum Deuterium in ihre Struktur einbauten. Da sie in ihrer Publikation jedoch nachwiesen, dass diese Fettsäuren *in vitro* selbst unter sehr harschen Bedingungen keine austauschbaren Wasserstoffatome besitzen, schlussfolgerten *Rittenberg* und *Schoenheimer*, dass diese Deuterium-Atome durch einen

1 Einleitung

unbekannten biosynthetischen Vorgang in den Zielmolekülen landeten, welcher vermutlich Fettsäuren aus kleineren Bestandteilen zusammensetzt. Außerdem konnten sie beobachten, dass aus den Mäusen isoliertes Cholesterin ebenfalls zu einem kleinen Prozentsatz Deuterium eingebaut hatte. Auch hier konnten sie einen passiven H/D-Austausch, welcher durch das im Medium vorhandene D₂O hervorgerufen werden könnte, durch *in vitro* Experimente ausschließen. Letzten Endes kamen sie zu der Erkenntnis, dass der Cholesterin-Biosyntheseweg daher mit dem Fettsäuremetabolismus zusammenhängen könnte.^[3] Heutzutage weiß man, dass bei der Cholesterinbiosynthese Acetyl-Coenzym A (Acetyl-CoA) benötigt wird, welches aus der β -Oxidation beim Fettsäureabbau von zum Beispiel Palmitinsäure entsteht.^[4,5] Das heißt, dass eine enolisierbare Carbonylverbindung entsteht, welche daraufhin in D₂O einen H/D-Austausch vollziehen kann. Auch diese Zusammenhänge wurden schon 1937 von *Rittenberg* und *Schoenheimer* postuliert.^[3] Zu genaueren Untersuchungen des biochemischen Metabolismus wurde jedoch frühzeitig bemerkt, dass es noch einiger anderer stabiler Isotope bedarf, mit denen die Ausgangsmoleküle markiert werden können. *Urey* und *Greiff* zeigten anhand theoretischer Rechnungen zunächst, dass es zwischen den leichten und schweren Isotopen eines Elements auch zu kleinen, physikalischen Unterschieden kommen sollte.^[6] *Lewis et al.* konnten 1933 nachweisen, dass H₂¹⁸O von H₂¹⁶O durch fraktionierte Destillation getrennt werden kann.^[7] Dies wurde auf einen unterschiedlichen Dampfdruck beider Spezies zurückgeführt, welcher schließlich von *Wahl* und *Urey* gemessen werden konnte. Dieser physikalische Unterschied wurde letztendlich von *Huffman* und *Urey* ausgenutzt, um damit den Anteil von H₂¹⁸O in einer Wasserprobe um das 4.5-fache anzureichern.^[8-10] Aufgrund dieser kleinen physikalischen Unterschiede, die eine Anreicherung ermöglichten, wurde von *Harold C. Urey* schnell das Potential chemischer Austauschreaktionen erkannt und somit war er bahnbrechend an der Entwicklung neuer Methoden beteiligt, mit denen stabile Isotopen letztendlich stärker angereichert werden konnten. 1937 untersuchte er Austauschreaktionen zwischen gasförmigen NH₃ und NH₄NO₃ oder (NH₄)₂SO₄, die er für die Anreicherung von ¹⁵N über eine fraktionierte Destillation ausnutzte.^[11] Dabei wurde in einem iterativen Verfahren Ammoniak evaporiert, kondensiert und als Salz in wässriger Lösung gebunden, welches dann erneut basisch behandelt wurde, sodass gasförmiger Ammoniak evaporieren konnte. Somit konnte ein Anreicherungsfaktor um das 6.5-fache gegenüber dem natürlichen Vorkommen von ¹⁵N erzielt werden. 1938 gelang es *Thode et al.* erstmals eine Methode zu beschreiben, bei der sie innerhalb zweier Wochen durch NH₄NO₃/NH₃-Austausch einen 14.8%-igen ¹⁵N-Stickstoffgehalt erreichen konnten.^[12] Damit wurde dieses stabile Isotop im Vergleich zum natürlichen Vorkommen um das 41-fache angereichert. Die Gerätschaft wurde weiterhin von den Autoren als „günstiger Apparat (500-600 Dollar)“ beworben, mit dem Wissenschaftler „in kurzer Zeit (zwei oder drei Monate)“ größere Mengen an ¹⁵N (150 mg/24 Stunden) erhalten können. Ein weiterer Meilenstein in der Anreicherung von stabilen Isotopen gelang 1939 *Roberts et al.*, indem sie die Anreicherung von ¹³C beschrieben.^[13]

1 Einleitung

Dabei wurden erneut chemische Reaktionen ausgenutzt, in diesem Fall wurde durch Zutropfen von konzentrierter H_2SO_4 zu einer wässrigen NaCN -Lösung gasförmiges HCN erzeugt, welches über eine lange, mit Glasspiralen gefüllte Säule strömte. Am Ende der Säule wurde das HCN wässrig gebunden, mit Hilfe einer Vakuumpumpe destilliert und durch massenspektrometrische Analyse wurde eine Anreicherung von ^{13}C in der Gasphase ermittelt. Durch weitere Feinjustierung des Experiments und dem Aufbau der Apparatur gelang es *Hutchinson et al.* letztlich, HCN mit einer ^{13}C -Anreicherung von insgesamt 25% zu erzeugen, außerdem konnten pro Tag 150 mg schwerer Blausäure dargestellt werden. Mit der damit vorhandenen Verfügbarkeit der stabilen Isotope ^{15}N , ^{13}C und ^2H wurden nun eine Reihe von bahnbrechenden Studien veröffentlicht. 1938 wurden erste Verfütterungsversuche an Ratten durchgeführt, denen an der α -Aminoposition ^{15}N -markiertes DL-Tyrosin verabreicht wurde.^[14] Eine ca. 2%-ige ^{15}N -Markierung war dabei schon ausreichend, um zu sehen, dass nur ein geringer Teil von ^{15}N -markierten Stickstoff im Urin der Ratten gefunden werden konnte und ein Großteil im Gewebe der Ratte verblieb. Vorallem in den isolierten Proteinen konnte eine ^{15}N -Markierung nachgewiesen werden. Diese tauchte unter anderem in den Aminosäuren Histidin und Arginin auf und in den „Monoamindicarbonsäuren“, wobei Glutamin- und Asparaginsäure gemeint waren. Außerdem konnte im Falle von Arginin gezeigt werden, dass neben der α -Aminogruppe der Stickstoff der Guanidinium-Gruppe markiert war. In einer weiterführenden Studie von *Schoenheimer et al.* wurden Ratten mit ^{15}N -Ammoniumcitrat gefüttert, das als interne, markierte Ammoniakquelle fungierte.^[15] Aus dem Gewebe der Ratten konnten sie dann erfolgreich ^{15}N -markiertes Histidin isolieren, das in einem *in vitro* Experiment seine ^{15}N -Markierung nach saurer Desaminierung zu Imidazolactat verlor. Dabei erkannten sie ein wichtiges Intermediat der Histidin-Biosynthese und schlussfolgerten aus ihren Beobachtungen, dass Histidin selbst in adulten Ratten ständigen Desaminierungs-Aminierungs-Prozessen an der α -Aminogruppe ausgesetzt sein muss. Im selben Jahr konnte das Team erfolgreich L-Leucin, bei dem das komplette Kohlenstoffgerüst ^{13}C - und die α -Aminogruppe ^{15}N -markiert war, an Ratten verfüttern und das Schicksal dieser Aminosäure untersuchen. Erneut kamen sie zu der Erkenntnis, dass die nach Isolierung des Gewebes gefundenen Aminosäuren an der α -Aminoposition ^{15}N -markiert waren und auch das reisolierte, deuterierte Leucin zum Teil seine Markierung an der Aminogruppe verlor. Dies deuteten die Forscher erneut mit fortlaufenden Aminierungs-Deaminierungs-Prozessen, die auf biosynthetischer Ebene ablaufen müssen. Außerdem postulierten sie, dass sich die Peptidkette in den Proteinen ständig öffnet und schließt, da auch Stickstoffe der Amidbindungen markiert gefunden werden konnten.^[16] An diesem Punkt sei erwähnt, dass hier nur ein kleiner Auszug der Arbeit von *Rittenberg* und seinen Kollegen vorgestellt wurde. Alleine im Jahr 1939 konnte er mit Hilfe der von *Urey* bereitgestellten stabilen Isotope in 14 verschiedenen Publikationen zeigen, wie man mit Hilfe von ^{15}N - und ^2H -Markierungen Licht in komplexe, biologisch-relevante Prozesse bringen kann.

Heutzutage ist die Verwendung von Isotopenverfolgungsexperimenten immer noch weit verbreitet. Die fortlaufenden Entwicklungen, zum Beispiel in der NMR-Spektroskopie, erlauben die Untersuchung des Metaboloms in verschiedenen Geweben oder Biofluiden wie dem Urin oder dem Blutserum, und das in kürzester Zeit.^[17] Fortschritte in der Massenspektrometrie und der davor gekoppelten Hochleistungs-Flüssigkeitschromatographie erlauben ebenfalls eine genaue, schnelle und breit anwendbare Analyse verschiedenster metabolischer Reaktionswege in komplexen Systemen. Sie werden in Zukunft noch weitere Entdeckungen ermöglichen, anhand derer Krankheiten besser verstanden und letztendlich behandelt werden können.^[18–20] Auch im Rahmen dieser Arbeit konnten neue Wege des epigenetischen Metabolismus von DNA-Derivaten aufgeklärt und näher untersucht werden, was ohne die Koppelung von Massenspektrometrie mit Isotopenverfolgungsexperimenten durch Verfütterung markierter Metaboliten unmöglich gewesen wäre.

1.2 DNA und die Vererbarkeit von genetischer Information

Mittlerweile ist allgemein bekannt, dass DNA als genetischer Informationsspeicher dient und somit den Konstruktionsplan von Lebewesen liefert. *Gregor Mendel* konnte 1866 als einer der ersten mit seinen Versuchen zur Kreuzung von Pflanzen die Vererbarkeit genetischer Informationen mit statistischen Methoden untermauern, besser bekannt als die *Mendelschen Regeln*.^[21] *Friedrich Miescher* beschrieb kurz darauf die Isolation und Entdeckung einer neuen Substanzklasse an organischen Molekülen aus Eiterzellen, welche er „Nuclein“ nannte, die DNA.^[22,23] *Sutton* und *Boveri* entdeckten 1902 beziehungsweise 1904, dass sowohl Spermium als auch Eizelle die gleiche Anzahl an Chromosomen liefern und die Entwicklung des Embryos von den einzelnen Chromosomen abhängig sind. Damit gelang es ihnen, die *Mendel'schen Regeln* mit Hilfe von biochemischen Entitäten zu erklären. *Boveri* und *Sutton* waren außerdem die ersten Wissenschaftler, die das Konzept von Genen beschrieben, welche sie auf den Chromosomen lokalisiert fanden.^[24–27] Knapp 40 Jahre später zeigten *Avery et al.*, dass es sich bei den auf den von *Sutton* und *Boveri* beschriebenen Genen um DNA handelt.^[28] *Chargaff* beschrieb in seiner 1953 erschienenen Publikation, dass dieses Biopolymer aus vier verschiedenen Basen besteht, welche in einem bestimmten Verhältnis zueinander auftreten: Die Menge an Adenosin in DNA entspricht der Menge an Thymin und die Menge an Guanosin entspricht der Menge an Cytosin, daraus leiteten sich die sogenannten *Chargaff Regeln* ab.^[29] Woher diese ihre Gesetzmäßigkeit erhalten, wurde weiterhin durch die Aufklärung der molekularen Struktur von DNA sichtbar, die mit Hilfe der *Chargaff Regeln* richtig interpretiert werden konnte. Aus der Kristallstruktur geht hervor, dass sich 2'-Desoxyadenosin (dA) mit Thyminidin (dT) und 2'-Desoxycytidin (dC) mit 2'-Desoxyguanidin (dG) in einer Doppelhelix paart.^[30–33] *Watson* und *Crick* postulierten damals schon, dass diese spezifische Basenpaarung einen potentiellen Mechanismus zur Speicherung von genetischen Informationen darstellen könnte.^[34] 1958

gelang es schließlich *Meselson* und *Stahl*, die semi-konservative Replikation und damit die Vererbung von genetischen Material auf biochemisch-molekularer Ebene in *E. coli* Bakterien zu beweisen.^[35] *Nirenberg* und *Matthaei* zeigten 1961 außerdem, dass das Basentriplet UUU für den Einbau von Phenylalanin verantwortlich ist.^[36] Diese Experimente schufen die Grundlage für die Entschlüsselung des genetischen Codes, genauer, welche Basenabfolge in der DNA für die jeweilige Aminosäure-Sequenz im Peptid eines Proteins kodiert. Die dafür jeweils verantwortlichen Basentriplets wurden letztlich durch *Crick et. al* drei Jahre später beschrieben.^[37] Diese Entdeckung und das zentrale Dogma der Molekularbiologie („DNA macht RNA und RNA macht Protein“), wie es ebenfalls zunächst von *Crick* formuliert^[38] und später von ihm konkretisiert wurde,^[39] setzt im wesentlichen die Wichtigkeit der DNA-Sequenz voraus. Jedoch sind mittlerweile mehrere Stufen der Genexpressionskontrolle bekannt, welche unter anderem auf Protein- oder direkt auf DNA-Ebene stattfinden. Auf diese epigenetischen Phänomene wird in den nächsten Abschnitten näher eingegangen.

1.3 Epigenetik

Wie zuvor ausgeführt, liegt der Bauplan eines jeden Lebewesens in der DNA gespeichert vor. Die einzelnen Zellen eines komplexen Organismus wie dem des Menschen besitzen die gleichen genetischen Informationen. Jedoch unterscheiden sich Zellen nicht nur in ihrer Morphologie, sondern logischerweise auch in ihrer Funktion. Diese spezialisierten, ausdifferenzierten Zellen (wie zum Beispiel Neuronen oder auch ein einfaches Haar), entspringen alle gemeinsamen Vorläufern, den embryonalen Stammzellen. Diese unterwerfen sich einer bestimmten Hierarchie, deren Anfang in einer Zygote begründet liegt. Diese entsteht nach der Befruchtung einer Eizelle mit dem Spermium, beides selbst hochspezialisierte und ausdifferenzierte Zellen. Aus dieser sogenannten totipotenten Vorläuferzelle proliferiert und differenziert im Laufe der Entwicklung jede weitere Zelle eines Embryos, Fötus oder adulten Organismus.^[40,41] Nachdem sich die Zygote über die Morula zum pluripotenten Blastozyst teilt, entwickelt sich eine embryonale Stammzelle, die sich schließlich in eines der drei Keimblätter (ektoderm, mesoderm oder endoderm) oder die Keimbahn (multipotente Stammzellen) spezialisiert (siehe **Abbildung 1.1**).^[42–45] Die Entwicklung embryonaler Stammzellen unterliegt damit in einem hohen Grad einer dynamischen Kontrolle zur Ausdifferenzierung und Umprogrammierung, welche nicht nur in eine Richtung ablaufen kann. Somatische, spezialisierte und damit ausdifferenzierte Zellen können zu embryonalen Stammzellen umprogrammiert werden. 1962 konnte dies von *Gurdon* gezeigt werden, in dem er den Zellkern einer somatischen Zelle in eine entkernte, befruchtete Eizelle eines Frosches implantierte. Daraus entwickelte sich ein völlig normaler Froschembryo. So konnte gezeigt werden, dass sich eine multipotente Zelle wieder in eine pluripotente umwandeln lässt.^[46] Über die Jahre hinweg konnte dieser Vorgang weiterentwickelt werden, so dass mittlerweile die Umprogrammierung von somatischen Zellen selbst zu

induzierten pluripotenten Stammzellen *in vitro* durch die Verwendung eines Cocktails aus verschiedenen Transkriptionsfaktoren möglich ist.^[47–49]

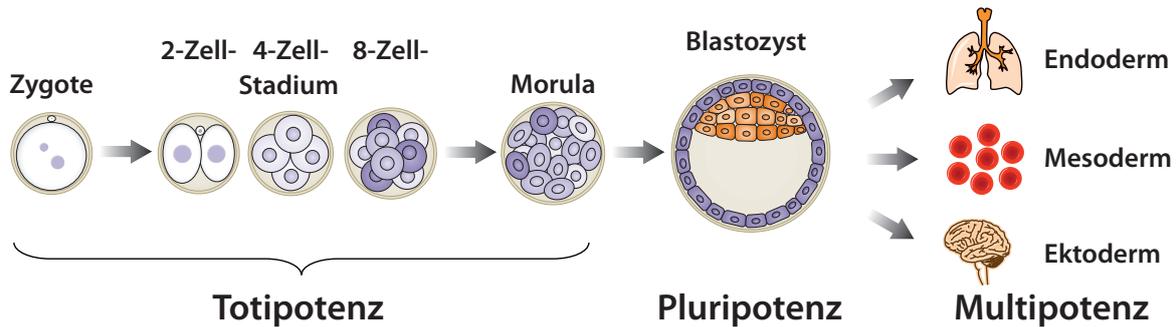


Abbildung 1.1: Übersicht der Entwicklung einer Zygote hin zu einer pluripotenten Stammzelle, welche sich in die unterschiedlichen Keimbälter entwickeln kann. Abbildung modifiziert nach Wu et al.^[50]

Für den geordneten Ablauf dieser beschriebenen Prozesse bedarf es hochorchestrierter Mechanismen, welche die Genexpressionskontrolle sowohl zeitlich als auch räumlich an distinkten Stellen im Genom steuern können. Conrad Waddington war vermutlich einer der ersten Wissenschaftler, die den Namen der Epigenetik einführten. Er beschrieb das Phänomen, wie ein Genotyp einen bestimmten Phänotyp während der Entwicklung erzeugt, als Epigenese.^[51] Robin Holliday fasste die Epigenetik auf einer molekular-biologischen Ebene präziser zusammen.^[52] Über die Jahre hinweg wurde die Definition der Epigenetik mit fortschreitendem Wissensstand weiter konkretisiert, so kann man sie heute als Veränderungen in der Genexpression deuten, welche nicht auf die DNA Sequenz zurückzuführen sind.^[53] Dieser Beschreibung kann man entnehmen, dass die genomische DNA auf einer weiteren „Metaebene“ verändert werden muss. Diese kann sich auf die Ebene der Nukleotide, oder auf der mit der DNA assoziierten Proteine, befinden. Im letzteren Fall handelt es sich um kovalente Modifikationen an Histonen, die die DNA zum Chromatin kondensieren. Auf die epigenetische Kontrolle, die über die Chromatinstruktur erreicht wird, wird im nächsten Kapitel kurz näher eingegangen.

1.3.1 Histonmodifikationen zur Kontrolle der Genexpression

Damit die DNA in die Zelle passt, muss sie in einer extrem kompakten Art und Weise organisiert sein. Pro diploider Zelle hat das Genom in etwa sechs Milliarden Basenpaare, wobei jedes Basenpaar ungefähr 0.34 nm lang ist. Das bedeutet, dass sich in jeder Zelle ca. zwei Meter DNA befinden. Der menschliche Organismus enthält 50×10^{12} Zellen, somit würde eine Aneinanderreihung der Gesamt-DNA eine Strecke bilden, die 300 Mal von der Erde zur Sonne und wieder zurück reicht. Roger Kornberg beschrieb mit dem Nucleosome Core Particle (NCP) als erster die Chromatinstruktur in seiner einfachsten Organisation, die sich aus einem Komplex von acht Histonen zusammensetzt.^[54] Die konkrete Zusammensetzung des NCP ist aus der Analyse seiner Kristallstruktur ersichtlich: Das Histonoktamer bildet sich aus jeweils

1 Einleitung

zwei Tetrameren, wobei ein Tetramer aus zwei Kopien der Histone H3 und H4, das andere aus je zwei Kopien von H2A und H2B besteht.^[55] In der von *Luger et al.* gezeigten Struktur windet sich eine DNA von 147 Basenpaaren um diesen Komplex, der ungefähr einen Durchmesser von 10 nm besitzt. Vier aufeinander folgende NCPs bilden die nächste Organisationsstufe, die sogenannte 30 nm Faser, die über die Linker-Histone H1 miteinander verbunden sind (**Abb. 1.2 A**). Zunächst wurden die Linker-Histone H1 bei der Rekonstitution der 30 nm-Faser nicht berücksichtigt, dadurch entstand ein vereinfachtes Bild dieser DNA-Verpackung.^[56]

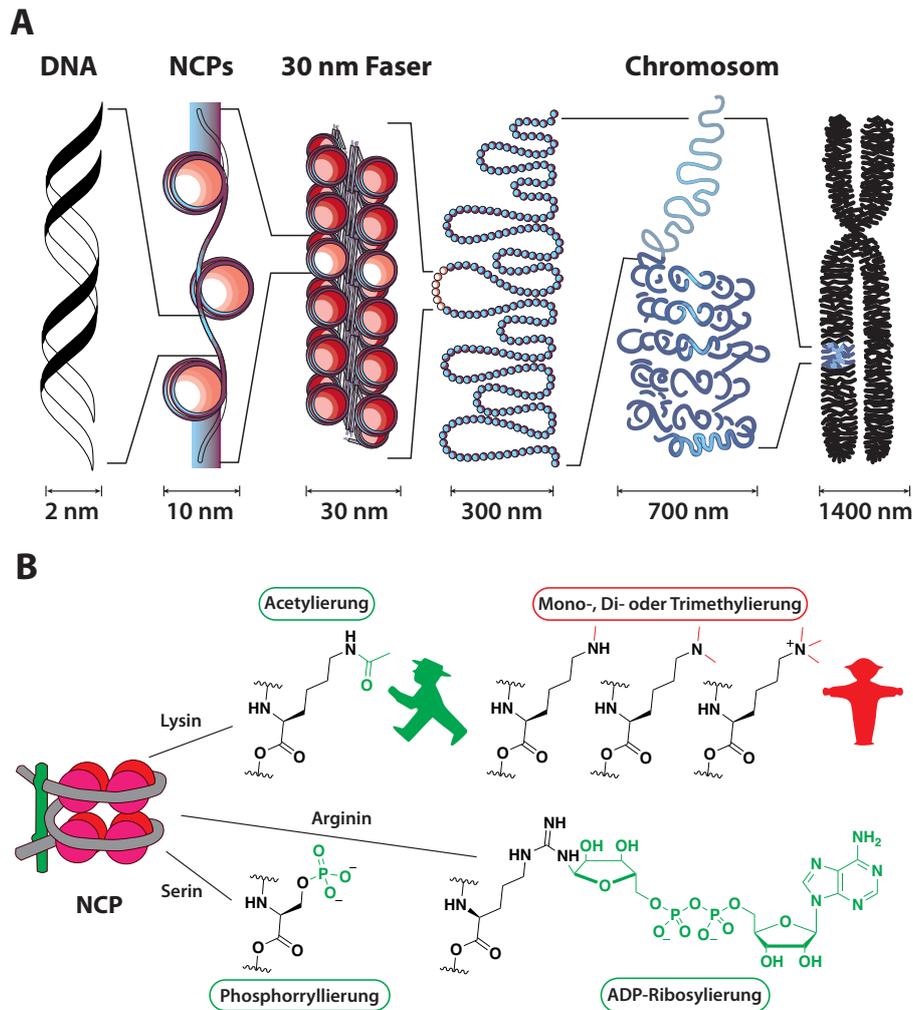


Abbildung 1.2: **A** Darstellung der einzelnen Organisationsformen der Chromatinstruktur (frei bearbeitet nach *Felsenfeld et al.*^[57]) **B** Beispiele einiger post-translationaler Modifikationen an den Histon-Schwänzen von Nucleosomen. Diese können eine Transkription durch Bildung von Euchromatin fördern oder die DNA für Transkriptionsfaktoren durch Ausbildung der Heterochromatinstruktur unzugänglich machen.

Mittlerweile konnte diese Struktur jedoch durch Cryo-Elektronenmikroskopie-Analyse konkretisiert werden: Anstelle einer Aneinanderreihung der NCPs kommt es bei der 30 nm Faser eher zu einer der DNA-ähnlichen Doppelhelixstruktur.^[58] Die 30 nm-Faser organisiert sich dann letztendlich zum Chromosom. Über eine weitere Kontrollebene kann durch posttranslationale Modifikationen der Histonchwänze die

1 Einleitung

Zugänglichkeit der DNA moduliert werden. Man unterscheidet zwischen Hetero- und Euchromatin, wobei erstere Form kondensiert ist und somit Faktoren wie die RNA-Polymerasen nicht an die DNA binden und diese transkribieren können. Im Euchromatin hingegen sind die Histone so moduliert, dass die DNA in „geöffneter“ Form für die Transkription der Gene zugänglich ist.^[59] Ob das Nukleosom nun in Form von Hetero- oder Euchromatin vorliegt, wird im einfachsten Fall durch rein elektrostatische Phänomene ermöglicht. Dazu wird durch eine Vielzahl von kovalenten Modifikationen der *N*-terminalen Histonden die Kompaktheit des Chromatins moduliert (**Abb. 1.2 B**). Lysine werden unter anderem acetyliert, methyliert, ubiquitiniert, SUMOyliert, crotonyliert, butyryliert oder propionyliert. Arginine können ebenfalls methyliert, ADP-ribosyliert oder citrulliniert werden, Threonine und Serine findet man unter anderem glykosyliert oder phosphoryliert.^[60,61,62,63,64,65] Die Aminogruppe der Lysin-Seitenkette wird zum Beispiel durch Trimethylierung quaternisiert und damit permanent positiv geladen. Dadurch kann eine positive Wechselwirkung mit dem negativ geladenen Phosphatrückgrat der DNA stattfinden, somit wird der DNA-Histon-Komplex kompakter und die DNA damit unzugänglicher. Die Lysine K9 und K27 im Histon H3 treten in diesem Kontext häufig auf, eine Trimethylierung geht hier mit einer Inaktivierung der Genexpressionsaktivität einher. Durch Acetylierung der Aminoseitenkette der zuvor genannten Lysine kann genau das Gegenteil bewirkt werden. Hier wird das freie Elektronenpaar quasi maskiert, es kann nicht zur Ausbildung einer positiven Ladung durch Protonierung kommen. Die Histone können dadurch delokalisiert und die DNA damit zugänglich gemacht werden. Phosphorylierte Serine findet man ebenfalls in Euchromatin. Durch die negative Ladung der Phosphatgruppe kommt es hier zu einer Abstoßung mit dem negativ geladenen Phosphatrückgrat der DNA.^[59] Die Veränderung der Chromatinstruktur durch elektrostatische Kräfte ist sehr gut verstanden, jedoch wird dadurch nur ein stark vereinfachtes Bild der Expressionskontrolle erzeugt. Die post-translationalen Modifikationen an den Histonschwänzen können eine Vielzahl an Reaktionen innerhalb der Zelle auslösen, wie zum Beispiel die Rekrutierung verschiedenster *Reader*-Proteine, die das Nukleosom verschieben oder im Falle der Erzeugung von Heterochromatin verdichten können. Es wurde unter anderem gezeigt, dass mono-ADP-ribosyliertes Arginin im Histon H3 durch dauerhafte Aktivierung des Transkriptionsfaktors P300 zur Proliferation eines kolorektalen Karzinoms führen kann.^[66]

Wie beim DNA-Basentriplett und der Genexpressionskontrolle durch die Methylierung der *C5*-Position von dC (*vide infra*), postuliert man für die Histonmodifikationen einen sogenannten Histon-*Code*, wenn dieser auch sehr kompliziert und noch uneindeutig ist.^[67] Alleine die Tatsache, dass mono-methyliertes Lysin K4 im Histon H3 mit einer Aktivierung oder eine K12-Acetylierung im Histon H4 mit einer Deaktivierung der Genexpression einhergeht, soll zeigen, dass mehrere sich gegenseitig beeinflussende Faktoren offensichtlich für die Ausbildung von Hetero- oder Euchromatin verantwortlich sind.^[68] Im ersten Fall wird durch die Methylierung von K4 eine aktive Rekrutierung von Histonacetyltransferasen

in Gang gesetzt, welche wiederum die Acetylierung von K9 und K14 und damit eine transkriptionelle Aktivierung bewirken.^[64]

Zur Stilllegung von Genen bedienen sich Zellen jedoch nicht nur der Veränderung der Chromatinstruktur. Durch kovalente Modifikationen an dC kann eine Genexpressionskontrolle auch direkt auf der DNA-Ebene geschehen. In den nächsten Abschnitten soll daher erläutert werden, wie ausgehend von 5-Methyl-2'-desoxycytidin (mdC) Gene dynamisch an-, beziehungsweise ausgeschaltet werden können.

1.3.2 Genexpressionskontrolle durch mdC als zentralen Schalter

Die am häufigsten in der DNA vorkommende, aktiv erzeugte Modifikation ist mdC. Ca. 4-6% aller dCs in genomischer DNA liegen an der C5-Position methyliert vor,^[69] wobei diese Modifikation häufig in CpG-Dinukleotiden auf gegenüberliegenden Stellen in genomischer DNA zu finden sind.^[70] Von diesen CpG-*islands* im Genom sind ca. 70-80% methyliert, außerdem kann in ca. 70% der Promotorregionen von Genen mdC gefunden werden.^[71,72] Schon früh konnte eine regulatorische Funktion dieser zuerst von *Hotchkiss* aus Kalbsthymus-DNA isolierten Verbindung gezeigt werden.^[73-75] Die allgemein akzeptierte Meinung, dass eine Methylierung von CpG-Sequenzen generell die Expression von Genen blockiert, ist jedoch eine sehr starke Vereinfachung eines relativ komplizierten Sachverhalts, bei dem viele verschiedene Faktoren zusammenspielen. So ist die Lokalisation der mdCs innerhalb eines Gens entscheidend. Liegt ein Promotor demethyliert vor, so wird die Expression des Gens stimuliert. *Downstream* des Gens können jedoch trotzdem CpG-Inseln methyliert vorliegen, dabei wird der Elongationsprozess, also die Verlängerung der transkribierten RNA, aber nicht gestoppt.^[76,77] Wenn CpGs in Promotorregionen methyliert sind, kann es zu einer Bindung von sogenannten *Methylated CpG Binding Proteins* (MBDs) kommen, welche dann weitere Repressoren wie zum Beispiel HDACs binden, wobei die Chromatinstruktur von Eu- zu Heterochromatin verändert werden kann (*vide supra*).^[78,79] Die Methylgruppe von mdC ist weiterhin in der großen Furche lokalisiert und ermöglicht dadurch eine Interaktion mit den verschiedenen regulativ wirkenden Proteinen und Proteinkomplexen (siehe **Abbildung 1.3**). Wie die zuvor diskutierten Modifikationen der Histonschwänze wird auch die Methylgruppe an mdC postsynthetisch eingeführt. Dies geschieht durch die Familie der DNA Methyltransferasen (Dnmts), welche den Kofaktor S-Adenosylmethionin (SAM) als Methyl donor verwenden. Als katalytische Unterformen kommen in Eukaryoten Dnmt1, Dnmt3a und Dnmt3b vor. Außerdem ist mit Dnmt3l eine weitere Form bekannt, welche zum jetzigen Wissensstand lediglich regulative Funktionen übernimmt und selbst nicht in der Lage ist, dC zu methylieren.^[80-82] Bei Dnmt1 spricht man von einer sogenannten *maintenance* Methyltransferase, welche das Methylierungsmuster während der Replikation aufrecht erhält und direkt an der Replikationsgabel lokalisiert ist.^[83,84,85]

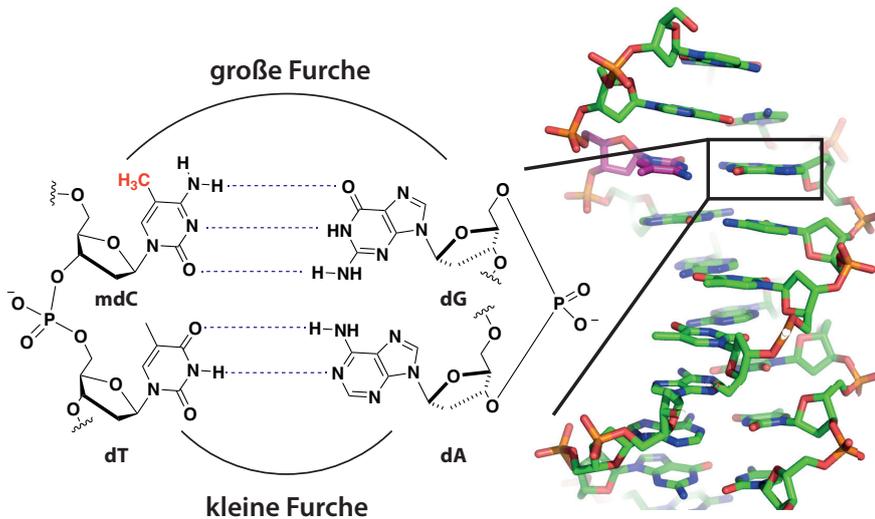


Abbildung 1.3: Darstellung der räumlichen Orientierung von mdC in einem DNA-Strang. DNA-Struktur erstellt in PyMol von *Markus Müller*.

Die Modifikation wird im Rahmen der Vervielfältigung der DNA quasi vererbt. Bei Dnmt3a und Dnmt3b handelt es sich um *de novo* Methyltransferasen.^[86,87] Diese methylieren dC replikationsunabhängig und können während der Differenzierung Methylierungsmuster in die DNA einführen. **Abbildung 1.4** zeigt mechanistisch, wie die Methylierung von dC zu mdC auf molekularer Ebene abläuft.

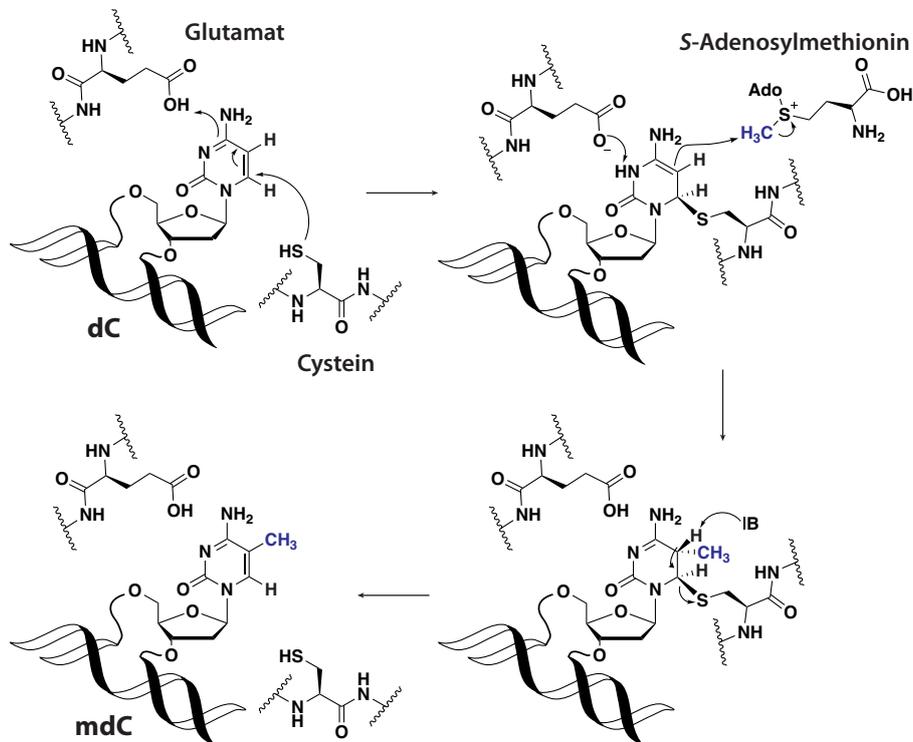


Abbildung 1.4: Darstellung des Mechanismus der Dnmt-Protein-katalysierten Methylierung von dC zu mdC.

1 Einleitung

Im aktiven Zentrum des Enzyms greift zunächst ein Cystein an der C6-Position von Cytosin an, wobei ein Glutamat als Proton-*Shuttle* den Stickstoff im Ring protoniert. Die dabei generierte nukleophile Spezies greift nun nach erneuter Deprotonierung die durch SAM bereitgestellte Methylgruppe an. Dabei entsteht eine C5-C6-gesättigte Spezies. Nach *syn*-Eliminierung des C5-Protons wird das Cystein schließlich wieder eliminiert und als Produkt entsteht mdC.

Der Mechanismus der Deaktivierung von Genen durch mdC ist zwar schon lange bekannt, jedoch sind noch lange nicht alle Mechanismen geklärt, wie diese Modifikation kurz nach der Befruchtung wieder verschwindet. Bei der Umprogrammierung des Genoms während der Embryogenese kommt es zu einem drastischen Rückgang der mdC Menge im väterlichen und mütterlichen Genom (**Abb. 1.5**). Dieser ist, gemessen an der Teilungsrate der Zellen, so massiv, dass er nicht durch passive, replikationsbedingte Verdünnung von mdC erklärt werden kann und es somit aktive Demethylierungsmechanismen zur Entfernung von mdC geben muss.

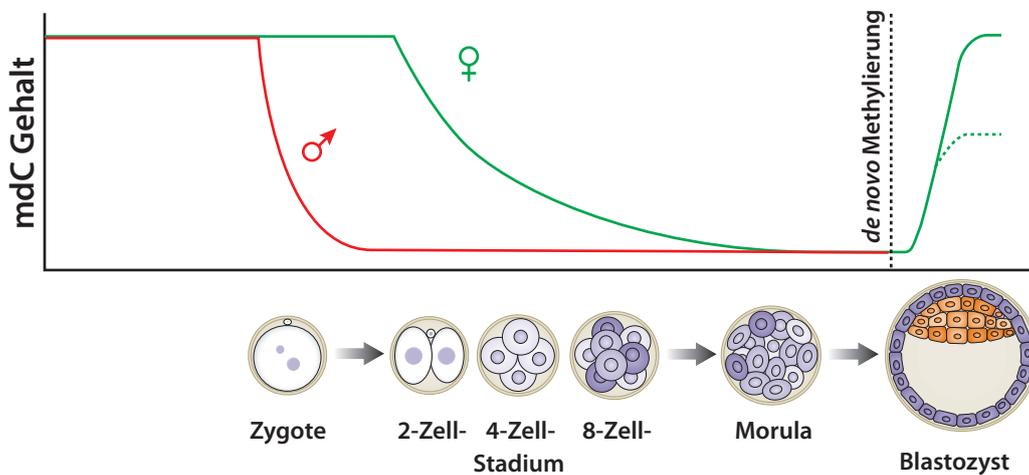


Abbildung 1.5: Illustration der Entwicklung des globalen mdC Gehalts nach der Befruchtung der Eizelle. Abbildung frei bearbeitet nach *Wu et al.*^[50]

In Pflanzen ist die direkte Entfernung von mdC über die Ros/Demeter-Enzyme gut beschrieben, jedoch ist in Säugetieren ein solcher Mechanismus nicht bekannt.^[50] 2009 wurde eine neue, enzymatisch erzeugte DNA Modifikation in embryonalen Stammzellen von Mäusen (mESC) entdeckt, von der man ausging, dass sie ein Intermediat bei der aktiven Demethylierung sein könnte. Bei dieser Modifikation handelt es sich um 5-Hydroxymethyl-2'-desoxycytidin (hmdC), welches ausgehend von mdC durch eine Tet-(*Ten eleven translocation*) vermittelte Oxidation gebildet wird.^[88,89] Auf diese Spezies und die dabei implizierten Funktionen wird im nächsten Abschnitt näher eingegangen.

1.3.3 Verteilung und Funktion von 5-Hydroxymethyl-2'-desoxycytidin

Die hydroxymethylierte Base hmdC wurde schon relativ früh als fester Bestandteil des genetischen Systems in Phagen beschrieben.^[90,91] Dort wird die 5-Hydroxymethylgruppe durch Glykosylierung mit

1 Einleitung

D-Glucose weiter modifiziert, sodass die DNA vom enzymatischen Abbau durch Restriktionsenzyme geschützt wird. Gleichzeitig produziert der Virus seine eigenen Endonukleasen, die die Wirts-DNA degradieren können.^[92] *Yura et al.* konnten hmdC auch im Gewebe von Ratten finden, jedoch waren hier mit 15% Hydroxymethylierung aller mdCs die beschriebenen Mengen enorm groß.^[93] Weiterhin wurde hmdC später erneut in Säugetier DNA entdeckt, jedoch waren diesmal die Mengen so gering, dass man bei hmdC von einer durch ROS (*reactive oxygen species*)-induzierten Schädigung von mdC ausging.^[94,95] 2009 gelang es jedoch zwei Forschergruppen unabhängig voneinander das Vorkommen von hmdC in genomischer DNA von Säugetieren nachzuweisen.^[88,89] *Tahiliani et al.* konnten außerdem das für die Oxidation von mdC zu hmdC verantwortliche Tet1-Enzym identifizieren. Dieses ist homolog zu den eukaryotischen α -Ketoglutarat abhängigen Eisen(II)-dioxygenasen *J Binding Protein* (JBP) 1 und 2.^[96,97] Diese oxidieren Thymin zu 5-Hydroxymethyl-2'-desoxyuracil (hmdU), welches dann ebenfalls mit Glukose zur sogenannten Base J glykosyliert werden kann. Die Parasiten, welche die Base J in ihrem Genom enthalten (zum Beispiel Leishmanien und Trypanosomen) sind unter anderem für die afrikanische Schlafkrankheit verantwortlich.^[98] Der postulierte Mechanismus der Hydroxymethylierung mit Hilfe einer α -Ketoglutarat abhängigen Eisen(II)-dioxygenase ist in **Abbildung 1.6** gezeigt.^[99,100,101,102]

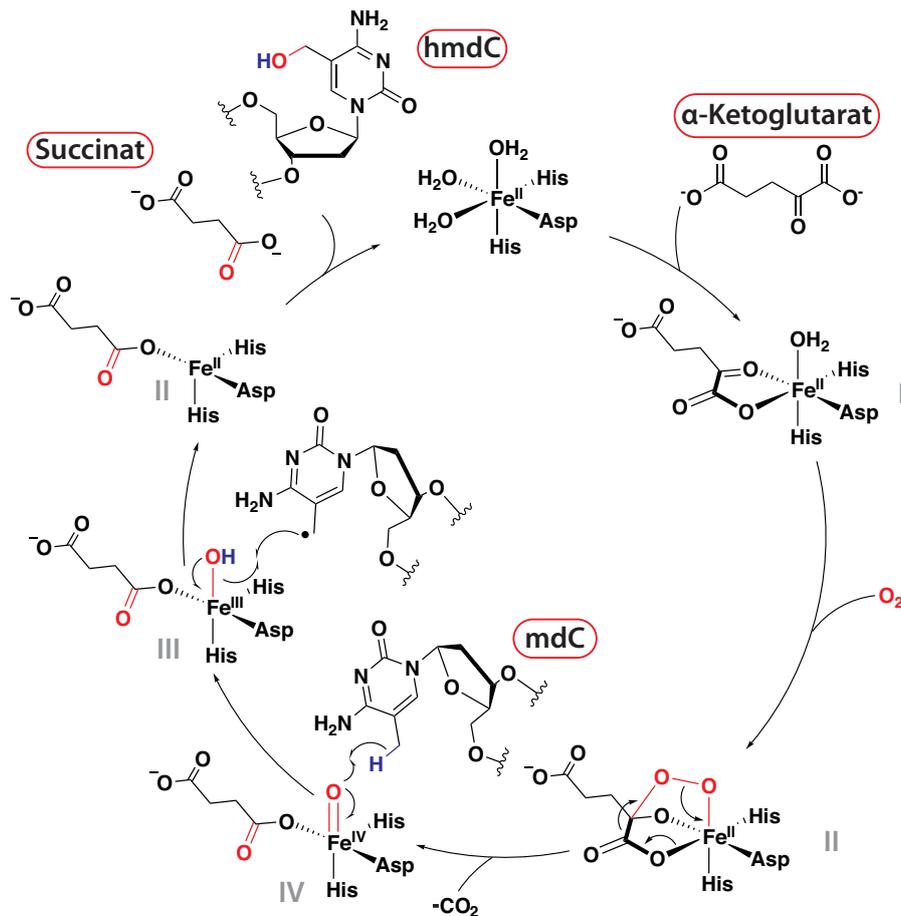


Abbildung 1.6: Mechanismus der Tet-vermittelten Hydroxymethylierung von mdC zu hmdC.

1 Einleitung

Zunächst koordiniert α -Ketoglutarat an das Eisen(II)-Zentrum unter Austausch zweier Aqua-Liganden. Molekularer Sauerstoff wird dann über den Austausch des dritten Aqualiganden aktiviert, wobei eine verbrückte, Acetal-ähnliche Zwischenstufe entsteht und der Sauerstoff kovalent an das carbonylische Kohlenstoffatom des α -Ketoglutarats gebunden ist. An dieser Stelle findet eine Decarboxylierung statt, wobei das Eisen(II)-Zentrum zu einem Eisen(IV)-Zentrum oxidiert wird. Daraufhin kann ein Wasserstoffatom der C5-Methylgruppe von mdC über einen radikalischen Mechanismus auf die Eisen-Oxo-Spezies übertragen werden, wobei es zu einer Reduktion des Eisen(IV)- zum Eisen(III)-Zentrum kommt. Über eine weitere radikalische Abstraktion der Eisen(III)-Hydroxy-Spezies kommt es zur finalen Hydroxymethylierung des transienten mdC-Radikals unter Ein-Elektronen-Reduktion des Eisen(III)-Zentrums zu Eisen(II), welches dann erneut Wasser als Liganden binden kann und für einen neuen katalytischen Zyklus verfügbar ist.

Die genaue Quantifizierung dieser Base in Mäusegeweben konnte unter anderem mit Hilfe der Isotopenverdünnungsmethode, das heißt der Zugabe von stabilen Isotopologen bekannter Konzentration, erreicht werden.^[103-105] In den meisten Organen von Mäusen ist hmdC mit einer Häufigkeit von 0.05% - 0.15% bezogen auf dC vorhanden, jedoch ändert sich der Anteil von hmdC in verschiedenen Regionen des Gehirns stark. Dort können bis zu einem Drittel aller mdCs hydroxymethyliert vorgefunden werden, außerdem akkumuliert die Menge an hmdC mit dem Alter.^[106,107] In der Grausubstanz des menschlichen Gehirns kann sogar bis zu 1.2% hmdC bezogen auf dG gefunden werden.^[108] Gleichzeitig konnte in der von *Wagner et al.* veröffentlichten Studie gezeigt werden, dass bei 5-Formyl-2'-desoxycytidin (fdC, *vide infra*) der gegenläufige Effekt beobachtet wurde, das heißt, dass die Menge an fdC mit zunehmenden Alter immer stärker sinkt. hmdC ist somit wohl eher eine permanente Base anstatt eines kurzlebigen Intermediats, was wiederum in Verfütterungsstudien von *Bachman et al.* bestätigt werden konnte.^[109] Neben der Verteilung von hmdC ist die Lokalisation dieser modifizierten Base im Genom von großem Interesse geworden. Dazu wurden neue Sequenzierungsmethoden entwickelt, mit denen Stellen im Genom mit hoher hmdC-Dichte gefunden werden konnten. Eine dieser verwendeten Methode ist die Sequenzierung über ein Restriktions-Endonuklease-System, welches auf einer enzymatischen Glykosylierungsreaktion der β -DNA Glykosyltransferase (β -GT) aus T4-Bakteriophagen beruht.^[110] Im DNA-Doppelstrang glykosyliert β -GT die freie Hydroxygruppe von hmdC, welche wiederum in die große Furche der DNA-Doppelhelix ragt.^[111,112] Dadurch wird das Erkennungspotential dieser Stelle im DNA-Doppelstrang geändert, wodurch das Restriktionsenzym MspI nichtmehr an diesen Stellen schneiden kann.^[110] DNA, welche lediglich dC, mdC oder hmdC enthält, kann jedoch weiterhin durch MspI prozessiert werden. Durch quantitative PCR von mit MspI verdauter, unverdauter und zuerst glukosylierter und dann verdauter DNA, können so Stellen im Genom, an denen hmdC sitzt, kartiert werden. Außerdem kann über ein zweites Experiment parallel ein Verdau über das Restriktionsenzym

HpaII erfolgen, welches bereits mdC und unmodifiziertes hmdC nichtmehr schneiden kann.^[113,114] Über einen Vergleich der unterschiedlichen Experimente kann so die Menge an dC, mdC und hmdC an verschiedenen Stellen im Genom relativ genau quantifiziert werden (**Abb. 1.7**).

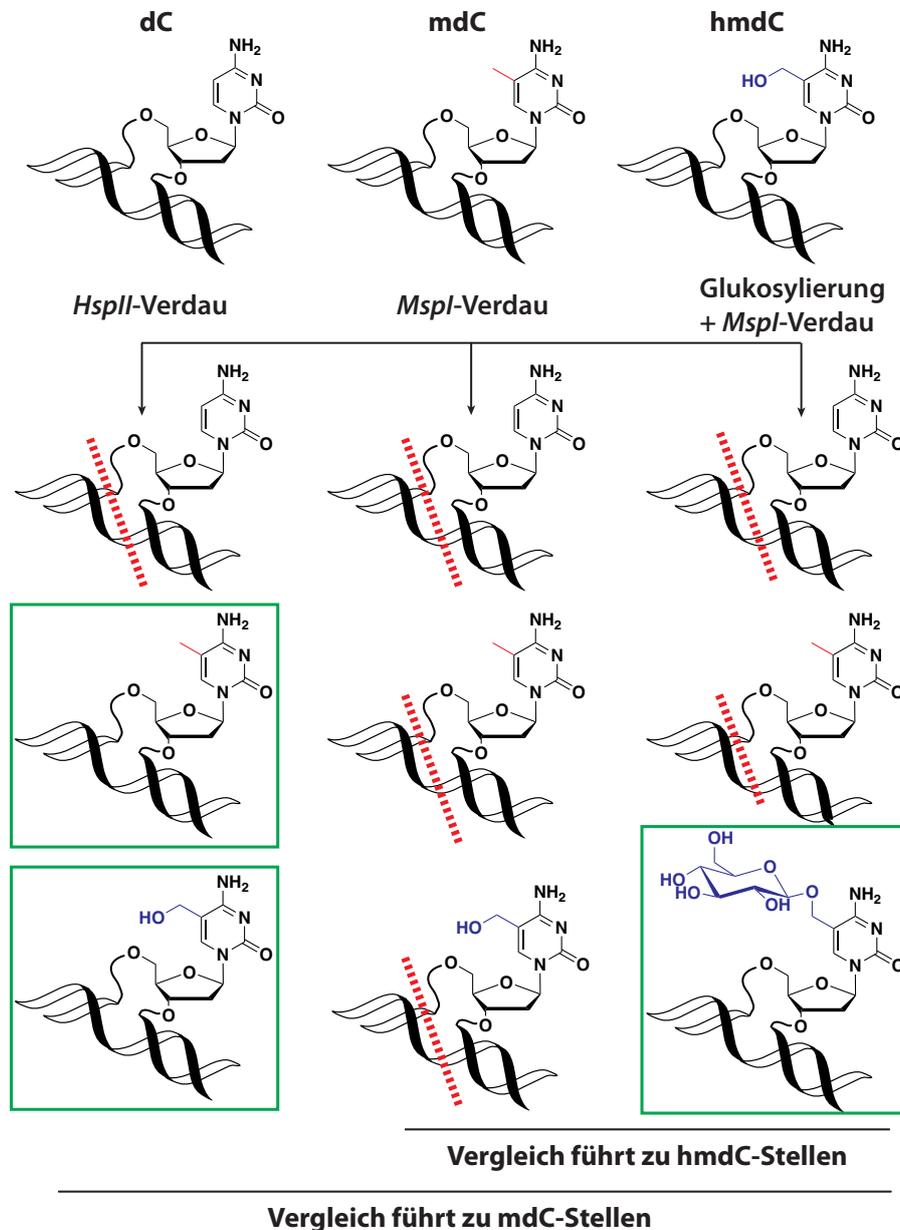


Abbildung 1.7: Sequenzierung von hmdC über *MspI* und *HpaII* Restriktionsenzyme. Durch vergleichende Analyse beider Methoden nach sequenziellen Experimenten verdauter, unverdauter und glukosylierter DNA, können Stellen, an denen sich hmdC oder mdC befindet von Stellen, an denen unmodifiziertes dC liegt, unterschieden werden.

Mit Hilfe weiterer Sequenzierungsmethoden, zum Beispiel dem *TAB-Seq* (*Tet-assisted bisulfite sequencing*), auf welches in diesem Kapitel nicht näher eingegangen wird, konnte man beobachten, dass mehr als 99% aller hmdCs in einem CpG, also in einem epigenetisch relevanten Kontext, vorkommen.^[115] Der größte Teil dieser hmdCs findet sich in euchromatidischen Regionen mit hoher Gendichte und

somit vermehrter Genexpressionsaktivität wieder.^[116,117] Während mdC die Genexpression blockiert, scheint hmdC ein Aktivator der Genexpression zu sein.^[118,119] Jedoch konnte mittlerweile auch gezeigt werden, dass hmdC in Promotor-Regionen ebenfalls einen repressiven Effekt auf die Genexpression haben könnte.^[120]

Auf der Suche nach weiteren Intermediaten der aktiven Demethylierung schien die Möglichkeit einer durch Tet hervorgerufenen Oxidation von hmdC zu fdC und 5-Carboxy-2'-desoxycytidin (cadC) eine logische Konsequenz zu sein. *In vitro* Studien konnten schon früh zeigen, dass durch das Fenton-Reagenz (Eisen(II) + H₂O₂), welches die Aktivität von Eisen(II)-Dioxygenasen nachstellt, hmdC zu fdC oxidiert werden kann.^[121] 2011 konnten schließlich die höher oxidierten Derivate fdC und cadC im Genom embryonaler Stammzellen entdeckt werden, wobei dadurch eine weitere Möglichkeit zur Kontrolle der Genexpression im Rahmen der aktiven Demethylierung erschlossen wurde.^[122-124]

1.3.4 5-Formyl-2'-desoxycytidin und 5-Carboxy-2'-desoxycytidin als Intermediate im epigenetischen Metabolismus

In genomischer DNA kann fdC mit einer Häufigkeit von ca. 0.02% aller dCs gefunden werden, cadC tritt mit 0.002% in etwa 10-Mal seltener auf, wobei die jeweiligen Mengen dieser Modifikationen wie bei hmdC stark von dem jeweils untersuchten Gewebe abhängen.^[92,122,125] Somit sind diese Modifikationen nur in kleinen Mengen vorhanden. Daher lag die Vermutung nahe, dass es sich um kurzlebige Intermediate während der aktiven Demethylierung handeln könnte. Jedoch konnte mittlerweile gezeigt werden, dass fdC positionsabhängig auch eine semi-stabile Base im Genom sein kann,^[126,127] die als epigenetischer Marker eigenständige Funktionen während der Transkription und Chromatin-Organisation haben könnte. *Kellinger et al.* konnten in einer systematischen Studie beobachten, dass fdC und cadC im Gegensatz zu dC und hmdC die Elongationsrate und Spezifität von RNA Polymerase II (Pol II) stark beeinträchtigen können.^[128] Durch fdC und cadC kommt es zu einer vorübergehenden Pausierung der Elongation, wobei diese als Signal für die Rekrutierung weiterer Transkriptionsfaktoren dienen kann. Es wurde bereits gezeigt, dass eine Verzögerung der Transkription unter anderem das Spleißen von *messenger* RNA (mRNA) einleiten kann oder dadurch sogar Faktoren gebunden werden, welche die Chromatinstruktur verändern und somit eine Transkriptionsregulation einleiten.^[129,130,131] Weiterhin konnte für fdC und cadC eine Vielzahl an sogenannten *Reader* Proteinen gefunden werden, die jeweils eine distinkte epigenetische Funktion im Umsatz dieser Modifikationen ausüben könnten.^[132,133] Darunter befanden sich neben verschiedenen Glykosylasen auch Chromatin- und Transkriptions-Regulatoren. Neben einer eigenständigen epigenetischen Funktion wurde auch früh postuliert, dass fdC die Struktur der Doppelhelix verändern könnte. Die Sequenzierung von fdC in genomischer DNA von mESCs konnte zeigen, dass es in spezifischen genomischen *loci* eine hohe Dichte aufeinanderfolgender fdCs gibt.^[134]

Außerdem oxidiert Tet mdC zu fdC präferiert an symmetrisch methylierten CpGs, also auf beiden gegenüberliegenden Strängen der DNA, was die Anzahl der aufeinanderfolgenden fdCs vergrößert.^[99] *Raiber et al.* synthetisierten ein selbstkomplementäres DNA-Dodecamer, welches drei aufeinanderfolgende fdCs im CpG-Kontext enthielt. Die Kristallstruktur zeigte dabei die ungewöhnliche Geometrie einer unterwundenen rechtsgängigen Helix, bei der die Formylgruppen von fdC in die große Furche zeigen und durch die starke Änderung der lokalen Furchengeometrie potentiell Protein-Erkennungsstellen entstehen.^[135] Jedoch sei hier erwähnt, dass die Kristallstruktur des unmodifizierten Dodecamers nicht bestimmt wurde, sodass kein direkter Vergleich zwischen den Strukturen gemacht werden kann. Dies ist vorallem problematisch, da während der Kristallisation des Dodecamers ein Puffer mit sehr hoher Salzkonzentration verwendet wurde und es bereits bekannt ist, dass es dadurch zu starken Verzerrungen der Helixstruktur, speziell an deren flexiblen Enden, kommen kann.^[136] Mittlerweile konnte mit Hilfe einer Kombination aus NMR-, UV-Vis und CD-Spektroskopie gezeigt werden, dass die von *Raiber et al.* beschriebene Struktur nicht in Lösung vorkommt, sondern dort eher eine normale B-DNA vorliegt.^[137] Somit ist fraglich, ob durch fdC größere strukturelle Veränderungen in der DNA-Konformation entstehen können, durch die dann die Bindung von bestimmten regulativen Faktoren ermöglicht wird. Jedoch weiß man, dass fdC, im Gegensatz zu mdC, die Flexibilität der DNA erhöht.^[138] Die Autoren dieser Studie schlagen vor, dass durch die oxidierten Derivate eine erhöhte Flexibilität in der DNA mit einer größeren Zugänglichkeit der auf Nukleosomen gewundenen DNA einher geht. Außerdem könnte dadurch die *Thymin DNA Glykosylase* (Tdg), welche fdC und cadC über den *base excision repair pathway* (BER) entfernen kann, an diese Stellen im Genom rekrutiert werden. In diesem Fall geht man bei fdC wieder von einem Intermediat der aktiven Demethylierung aus und es kommen somit verschiedene Mechanismen zur Entfernung dieser Modifikation in Frage, welche im nächsten Abschnitt genauer beleuchtet werden sollen.

1.4 Aktive Demethylierung von mdC

Neben einer passiven Demethylierung, bei der mdC durch Zellteilung bei gleichzeitiger Deaktivierung von Dnmt1 mit dC ausgetauscht wird, gibt es in der Zelle einen weiteren Mechanismus, der replikationsunabhängig ist. Durch Immunofärbung von mdC konnte von *Maeyer et al.* gezeigt werden, dass kurz nach der Befruchtung die Menge an mdC im väterlichen Genom der Zygote rapide abnimmt (siehe **Abbildung 1.5**).^[139] Diese schnelle Abnahme konnte nicht mit dem Vorhandensein eines einzigen, passiven Mechanismus erklärt werden, bis 2009 schließlich hmdC als sechste Base im Genom entdeckt wurde und seitdem als mögliches Intermediat im Rahmen der aktiven Demethylierung diskutiert wird. *Wossidlo et al.* konnten 2011 außerdem zeigen, dass die rapide mdC-Abnahme mit einer aktiven Bildung von hmdC einhergeht, was die Rolle der oxidierten mdC-Derivate bei der Reprogrammierung der Zellen

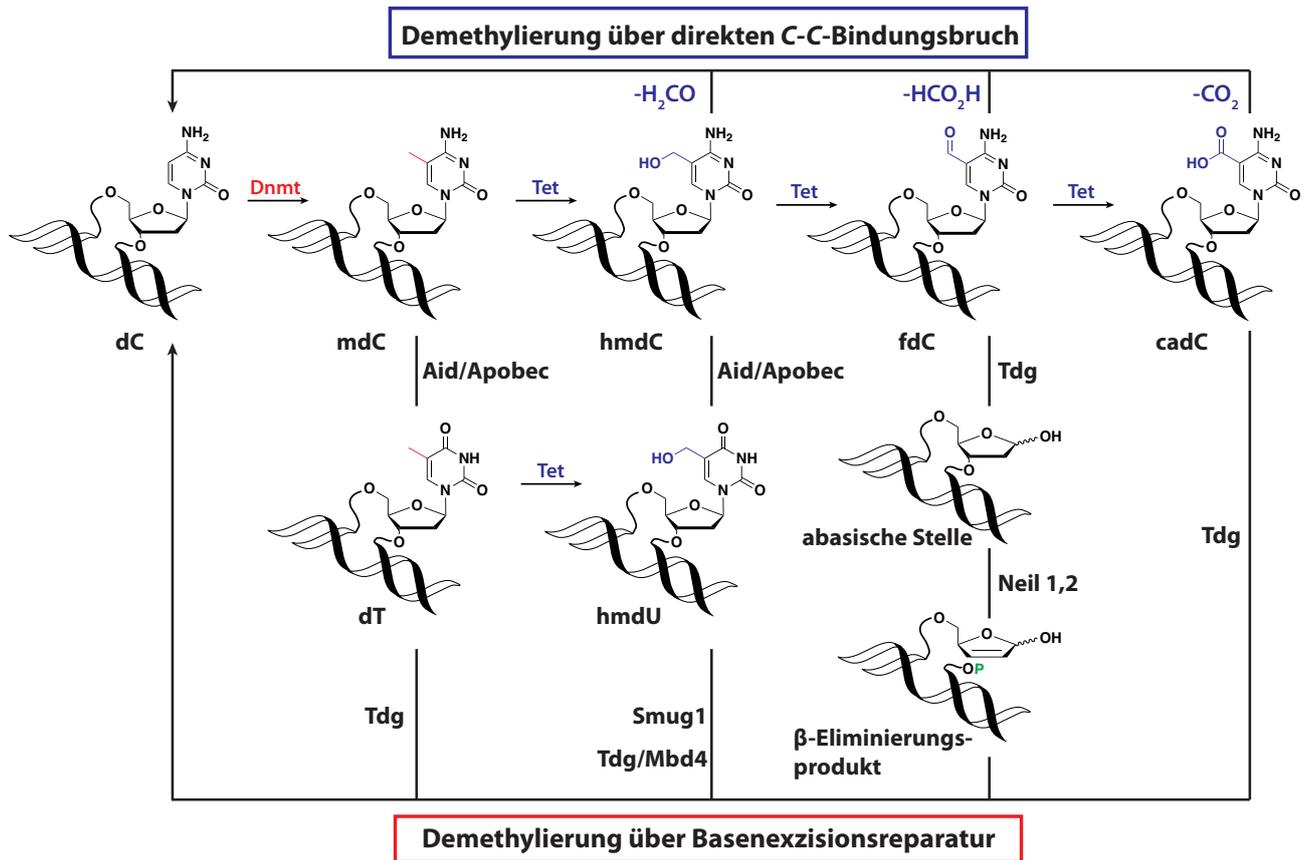


Abbildung 1.8: Überblick zu verschiedenen Wegen der aktiven Demethylierung. Oben ist die Entfernung der oxidativen Derivate über einen direkten *C-C*-Bindungsbruch gezeigt, bei dem das oxidierte Kohlenstoffatom entweder als Formaldehyd (hmdC), Ameisensäure (fdC) oder Kohlenstoffdioxid (CO_2) entfernt wird. Der untere Teil zeigt die Entfernung über BER, wobei fdC und cadC direkt ausgeschnitten und mdC oder hmdC zunächst enzymatisch Desaminiert werden und schließlich dT respektive hmdU über den BER weiter prozessiert werden.

unterstreicht.^[140] In den nächsten beiden Abschnitten sollen zwei Möglichkeiten der aktiven Demethylierung näher beleuchtet werden. Zum einen der Reaktionsweg über BER, welcher seit Entdeckung von fdC eine große Bedeutung in der Epigenetik erlangt hat und zum anderen die aktive Demethylierung über einen direkten *C-C*-Bindungsbruch an der *C5*-Position der modifizierten dC-Derivate. **Abbildung 1.8** gibt einen Überblick über mögliche Transformationen, die in der Zelle stattfinden können um mdC mit dC auszutauschen.

1.4.1 Basen-Exzisions-Reparatur gekoppelte aktive Demethylierung

Ein etablierter Mechanismus zur Entfernung von mdC ist die Prozessierung über Basenexzisionsreparatur, bei der intermediär abasische Stellen entstehen (**Abb. 1.8**). Mit einem geschätzten Auftreten von mehr als 10000 abasischen Stellen pro Zelle pro Tag kann man dieses kurzlebige Intermediat im Genom der DNA als den häufigsten Schaden bezeichnen.^[141] Die Bildung einer abasischen Stelle geht

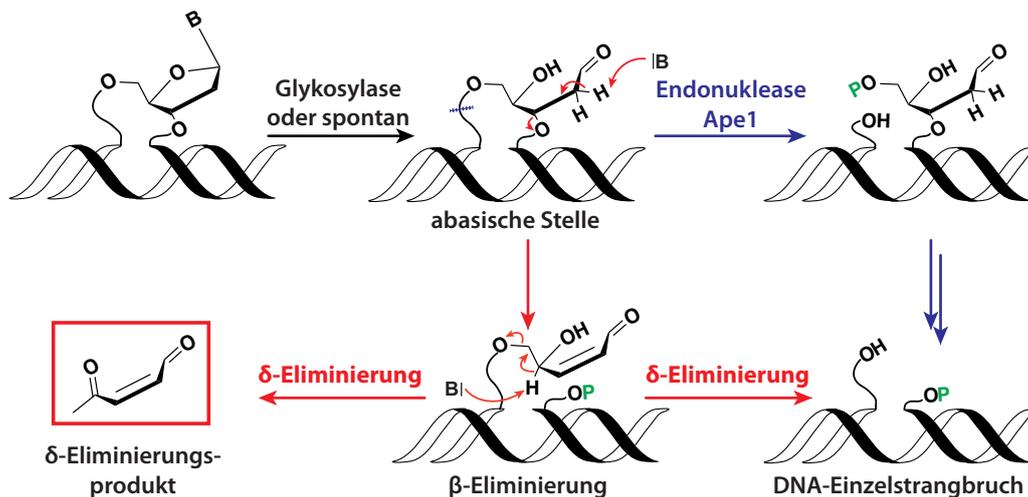


Abbildung 1.9: Übersicht zur Basenexzisionsreparatur. Gezeigt sind die verschiedenen Prozessierungswege, welche entweder über monofunktionale Glykosylasen und der Aktivität von Ape1 ablaufen oder über Bifunktionale Glykosylasen, bei der eine 2'-3'-ungesättigte Bindung entsteht.

zunächst mit dem Bruch der glykosidischen Bindung zur Base im DNA Doppelstrang einher, wobei dieser Vorgang entweder spontan oder eben aktiv im Rahmen der BER erfolgt.^[142–145] Für die Erkennung zur Prozessierung der verschiedenen Basen bedient sich die Zelle einer Vielzahl an spezifischen Glykosylasen, die entweder mono- und bifunktional sein können. Erstere trennen ausschließlich die glykosidische Bindung zwischen Ribose und DNA-Heterozyklus unter Bildung einer abasischen Stelle, welche dann über sogenannte Endonukleasen (zumeist *apurinic/apyrimidinic endonuclease* (Ape1)) zu einem freien 3'-OH und einem 5'-Phosphatester (5'-dRP) prozessiert wird (**Abbildung 1.9**).^[146] Bifunktionale Glykosylasen weisen neben einer Hydrolysefähigkeit der *N*-glykosidischen Bindung zwischen DNA-Base und Ribose außerdem eine sogenannte Lyasefunktion auf. Dabei wird ein Proton an der 2'-Position der abasischen Stelle unter Bildung einer 2'-3'-ungesättigten Bindung abstrahiert und es erfolgt ein Strangbruch.^[147,148] Die Anhäufung oder fehlende Prozessierung der abasischen Stellen kann zu Strangbrüchen oder sogar Doppelstrangbrüchen führen, ferner tritt mdC häufig in Clustern auf und die mdCs können an beiden gegenüberliegenden Stellen im Genom symmetrisch methyliert sein. Die BER kann also potentiell Doppelstrangbrüche erzeugen.^[149] Somit muss die aktive Demethylierung über die BER ein hochorchestrierter Mechanismus sein. Für die BER-gesteuerte Entfernungen von mdC oder dessen oxidierter Derivate hmdC, fdC und cadC gibt es viele Beispiele. 1995 konnte in einem *in vitro* Assay von *Jost et al.* gezeigt werden, dass eine mit Tdg co-eluierende Glykosylase mdC aus hemimethylierter DNA herauschneiden kann.^[150] Damals konnten sie weiterhin beobachten, dass bei der Aufreinigung des Enzyms aus Hühnchenembryos die Endonuklease Ape1 co-eluiert, was einen Hinweis darauf gab, dass eine BER-basierte Entfernung von epigenetisch relevanten Basen in einem Komplex mit den in diesem Prozess involvierten Faktoren stattfindet. Des weiteren konnte mit Mbd4 (*Methyl-CpG-binding domain*

protein 4) ein weiteres Enzym identifiziert werden, welches eine mdC-Glykosylase-Aktivität besaß.^[151] Jedoch sei erwähnt, dass die Aktivitäten der direkten mdC-Entfernung von sowohl Tdg als auch Mbd4 sehr schwach sind, die Reaktion langsam abläuft und ausschließlich *in vitro* abläuft.^[150,151,152] Eine direkte Entfernung von mdC, so wie sie in Pflanzen beschrieben ist, scheint daher unwahrscheinlich, jedoch könnte die Entfernung auch über einen Umweg geschehen. Die dC-Desaminasen Aid/Apobec können aktiv durch Desaminierung von mdC eine dT/dG-Fehlpaarung erzeugen, wobei dann dT über Tdg ausgeschnitten werden kann.^[153–157] Zwar gibt es sehr viele Hinweise auf die Beteiligung einer enzymatischen Desaminierung bei der aktiven Demethylierung von mdC zu dC, jedoch sei erwähnt, dass die Aid/Apobec-Proteine unmodifizierte dCs und einzelsträngige DNA stark bevorzugen, was ihre Aktivität bei einer aktiven Demethylierung wieder in Frage stellt.^[158]

Mit der Wiederentdeckung von hmdC im Jahre 2009 und der Entdeckung der durch Tet 1-3 vermittelten Weiteroxidation zu fdC und cadC konnten neue Erkenntnisse in der aktiven Entfernung von mdC gewonnen werden.^[88,89,122,123,124] Durch Immunofluoreszenz-Färbung wurde zunächst gezeigt, dass während der Ausbildung der Pluripotenz in embryonalen Stammzellen die Menge an hmdC steigt, während die Level an mdC sinken.^[140,159] Damit entsteht eine direkte Verbindung zwischen der Entfernung von mdC und der Bildung von hmdC. Diese Entfernung erfolgt zu einem großen Teil replikationsabhängig, also passiv, hmdC ist außerdem kein Substrat der *Maintenance* Methyltransferase Dnmt 1.^[160,161,162] Die hydroxymethylierte Form hmdC kann prinzipiell auch über das Aid/Apobec-System enzymatisch zu hmdU desaminiert werden, wobei eine hmdU-dG-Fehlpaarung entstünde, die dann über die BER zu einem unmodifizierten dC repariert werden könnte. *Hashimoto et al.* zeigten bereits, dass hmdU über Tdg/Mbd4 ausgeschnitten werden kann^[160] und weiterhin gibt es Hinweise, dass eine Desaminierung von hmdC über Aid eine wichtige Rolle bei der aktiven Demethylierung spielt. So zeigten unter anderem *Cortellino et al.*, dass Tdg mit Aid interagiert und hmdU in Paarung mit dG von Tdg ausgeschnitten wird. Hier sei jedoch erwähnt, dass die Aktivität von Aid bei größer werdendem Rest an der C5-Position von dC stark abnimmt.^[163] Die Glykosylase Smug1 hingegen kann hmdU auch im dA-Kontext prozessieren.^[164] Letzteres Szenario kann bei einer Oxidation von dT zu hmdU auftreten, welche in Stammzellen ebenfalls von den Tet-Proteinen enzymatisch katalysiert wird und somit einen weitere Möglichkeit der aktiven Demethylierung von mdC darstellt.^[165]

Mit der Entdeckung von fdC im Jahre 2011 konnte gezeigt werden, dass fdC, und zu einem kleinen Teil auch cadC, Substrate der Glykosylase Tdg sind.^[124,166,167] Ein Knockout dieser Glykosylase in mESC kann außerdem zu einer 10-fachen Anreicherung dieser beiden Modifikationen führen.^[124,127,134,168,169] Durch eine Rekonstitution der bei der Entfernung von fdC beteiligten Enzyme Tet, Tdg, Ape1, Pol β und Xrcc1-Lig3 *in vitro* wurde ebenfalls gezeigt, dass die BER-gestützte Entfernung ein hoch orchestrierter Vorgang ist.^[170] Ein 60-Basenpaar langes Fragment, welches ein mdC enthielt, konnte in diesem System

zu fdC/cadC aufoxidiert werden und schließlich wurde über die Entfernung der daraus resultierenden abasischen Stelle ein dC-Monophosphat eingebaut. Der DNA-Strang konnte dann ligiert werden, so dass ein komplettes 60 Basenpaar langes Oligodesoxynukleotid zurückerhalten wurde. In einer von *Spruijt et al.* durchgeführten Studie wurden zudem die Neil-Proteine als Binder von fdC ermittelt.^[132] Diese bifunktionalen Glykosylasen sind vor allem für die Entfernung von oxidativ geschädigten DNA-Basen verantwortlich, wie zum Beispiel 8-Oxo-dG oder 5-Hydroxy-Cytosin.^[141,171,172] *Schomacher et al.* konnten 2016 schließlich zeigen, dass Neil 1 und 2 in Zusammenarbeit mit Tdg die Prozessierung von fdC und cadC stark erhöhen.^[173] Ein Faktor hierfür ist, dass die Neil-Proteine in der Lage sind, Tdg von der abasischen Stelle zu entfernen, welches dort nach der Trennung der glykosidischen Bindung sehr stark gebunden bleibt, da Tdg eine durch das Produkt inhibierte DNA-Glykosylase ist.^[174] *Slyvka et al.* bestätigten die stimulierende Wirkung von Neil1 auf die Tdg-gestützte Exzision von fdC und cadC darüberhinaus *in vitro*.^[175]

Zwar ist die Beteiligung der BER und speziell von Tdg zum Teil gut verstanden und akzeptiert, jedoch gibt es auch Tendenzen, die zeigen, dass die aktive Demethylierung im Ein-Zell-Stadium von Maus-Embryos zwar von einer Tet3-vermittelten Oxidation, jedoch nicht von Tdg anhängig ist und somit ein alternativer Mechanismus zur aktiven Entfernung von mdC in den Fokus rückt.^[176] Eine weitere Möglichkeit der Entfernung von mdC und der oxidierten Derivate ist die direkte Trennung der C-C-Doppelbindung an Position C5, auf welche im nächsten Abschnitt noch näher eingegangen wird.

1.4.2 Demethylierung über einen C-C-Bindungsbruch

1999 wurde von *Bhattacharya et al.* eine direkte Entfernung der Methylgruppe von mdC zu dC durch Mbd2 beschrieben, welche laut ihren Ergebnisse spezifisch für CpG-lokaliserte mdCs erfolgte.^[177,178] Wasser und mdC wurden dabei als Edukte und Methanol und dC als Produkte vorgeschlagen. Diese Ergebnisse konnten jedoch nicht verifiziert werden, vielmehr wurde Mbd2 als Faktor identifiziert, der die Transkription unterdrückt und keine Demethylase Aktivität besitzt.^[179] Die direkte Entfernung einer Methylgruppe von einem aromatischen sp²-Kohlenstoff unter Freisetzung von Methanol erscheint weiterhin chemisch und thermodynamisch unplausibel, da in diesem Szenario am C5-Atom ein vinyliches Anion entstehen müsste, welches stark destabilisiert ist. In einer weiterführenden Studie konnte jedoch *in vitro* gezeigt werden, dass die DNA-Methyltransferasen Dnmt1, 3a und 3b in Abwesenheit von SAM mdC direkt zu dC demethylieren können.^[180] Jedoch wurden dabei keine weiteren mechanistischen Untersuchungen durchgeführt, welche die möglichen Produkte einer solchen Reaktion identifiziert hätten. Außerdem waren für diese Experimente oxidative Bedingungen nötig und es ist somit nicht klar, ob eine oxidierte Form von mdC für die beobachtete Reaktivität verantwortlich war. *Chen et al.* zeigten weiterhin, dass Dnmt3a und 3b hmdC direkt zu dC dehydroxymethylieren können und

gleichzeitig das dehydroxymethylierte Produkt remethylieren können. Diese Reaktion ist somit reversibel und würde ein attraktives System zur Entfernung und Wiederherstellung des Methylierungsmusters darstellen, ein *in vivo* Beweis dieses Reaktionsweges steht aber weiterhin aus.^[181] *Liutkeviciute et al.* beschrieben ebenfalls ein System, bei dem eine Methyltransferase in Abwesenheit von SAM nicht nur verschiedene Aldehydfunktionalitäten an die C5-Position von dC anbringen kann, sondern auch eine Reversierung der Hydroxymethylierung von hmdC zu dC vollzogen wird.^[182] Hier ist aber wie bei der von *Chen et al.* beschriebenen, direkten Demethylierung von mdC zu dC der Ausschluss von SAM als Methylgruppendonor notwendig, was aufgrund der hohen intrazellulären SAM-Konzentration biologisch irrelevant erscheint. Jedoch kann man aus diesen Experimenten deduzieren, dass die Oxidation der Methylgruppe für eine direkte Demethylierung über einen C-C-Bindungsbruch nötig sein könnte. Diese oxidativen Demethylierungen sind in der Natur ebenfalls bekannt und zum Teil gut beschrieben. Im *Thymine-Salvage-Pathway* wird über die Isoorotat-Decarboxylase (IDCase) Isoorotat zu Uracil enzymatisch decarboxyliert.^[183] Zuvor wird Thymin durch die *Thymin-7-hydroxylase*, welche wie die Tet-Proteine eine α -Ketoglutarat-abhängige Eisen(II)-oxygenase ist, zu hmU, 5-Formyluracil und schließlich zu 5-Carboxyuracil oxidiert.^[184] Trotz einer eher schwachen Aktivität konnte mittlerweile auch eine IDCase-vermittelte Decarboxylierung von cadC zu dC beobachtet werden.^[185] Methylierte Lysinreste können ebenfalls über eine Oxidationskaskade demethyliert werden. Dafür sind vor allem die sogenannten *Jumoni*- und Amin-Oxidase-Enzyme verantwortlich.^[186–189] Die Entfernung von Methylgruppen an exozyklischen Aminen von DNA- oder RNA-Basen wie zum Beispiel bei m⁶A erfolgt ebenfalls über eine Oxidations-Demethylierungs-Kaskade.^[190–192] Der große Unterschied zur durch die IDCase beschriebenen Decarboxylierung oder der allgemeinen Demethylierung von epigenetisch relevanten dC-Derivaten liegt jedoch in der Verknüpfung der Methylgruppe. Eine an einem Stickstoff gebundene Formyl- oder Carboxylgruppe, wie es bei den methylierten Lysinresten oder exozyklischen Aminen der Fall ist, lässt sich chemisch gesehen relativ leicht hydrolysieren. Eine an einem Csp²-Zentrum gebundene Hydroxymethyl-, Formyl- oder Carboxylgruppe ist hingegen stabil. In der Cholesterinbiosynthese gibt es jedoch einen prominenten Fall, bei dem erneut durch eine Oxidationskaskade eine Methylgruppe entfernt werden kann (siehe **Abbildung 1.10 A**). Die dafür verantwortliche Lanosterin-Demethylase oxidiert zunächst in zwei Schritten mit Hilfe von zwei Äquivalenten Sauerstoff und NADPH die C-14- α -Methylgruppe zu einer Formylgruppe und mit einem weiteren Äquivalent Sauerstoff/NADPH wird mit Hilfe eines nukleophilen Eisen(II)-Peroxyanion unter Bildung einer Doppelbindung die Methylgruppe in Form von Formaldehyd frei.^[92,193] Im Gegensatz zur möglichen Entfernung einer Formylgruppe wie zum Beispiel bei fdC läuft die Reaktion hier an einem Csp³-Zentrum ab. Das heißt, dass eine oxidative Entfernung von Methylgruppen, welche an C-Atomen gebunden sind, vermutlich nur über ein Csp³-Zentren ablaufen. *Schiesser et al.* zeigten 2012, dass eine direkte Decarboxylierung von cadC zu

dC in Stammzelllysaten möglich ist.^[194] Erste Experimente und eine weiterführende *in vitro*-Studie zeigten, dass für den Ablauf dieser Reaktion eine Sättigung der C5-C6-Bindung am dC nötig ist, somit also ein Csp²-intermediär zu einem Csp³-Zentrum umgewandelt werden muss.^[195] Diese Sättigung kann durch Angriff eines Helfer-Nukleophils an der C6-Position, wie zum Beispiel einem Cystein, erfolgen, welches dann nach der Deformylierungs- oder Decarboxylierungsreaktion unter Rearomatisierung des Heterozyklus eliminiert werden kann (**Abb. 1.10 B**). Erneut konnten *Liutkeviciute et al.* zeigen, dass Dnmt 3a und 3b cadC *in vitro* direkt zu dC decarboxylieren können.^[196] Die Reaktion läuft wie zuvor nur unter Ausschluss von SAM ab und somit ist auch dieser Mechanismus *in vivo* fragwürdig. Eine enzymatische Aktivität zur direkten Decarboxylierung von cadC zu dC konnte darüberhinaus noch nicht entdeckt werden. Schließlich konnte vor kurzem eine direkte Deformylierung von fdC zu dC *in vivo* durch Isotopenverfolgungsexperimente beobachtet werden.^[197] Auch hier ist die dafür verantwortliche Entität noch nicht bekannt, jedoch unterstützen diese Experimente das Konzept einer aktiven Demethylierung, welche nicht über die Erzeugung von für die Zelle potentiell schädlichen abasischen Stellen abläuft.

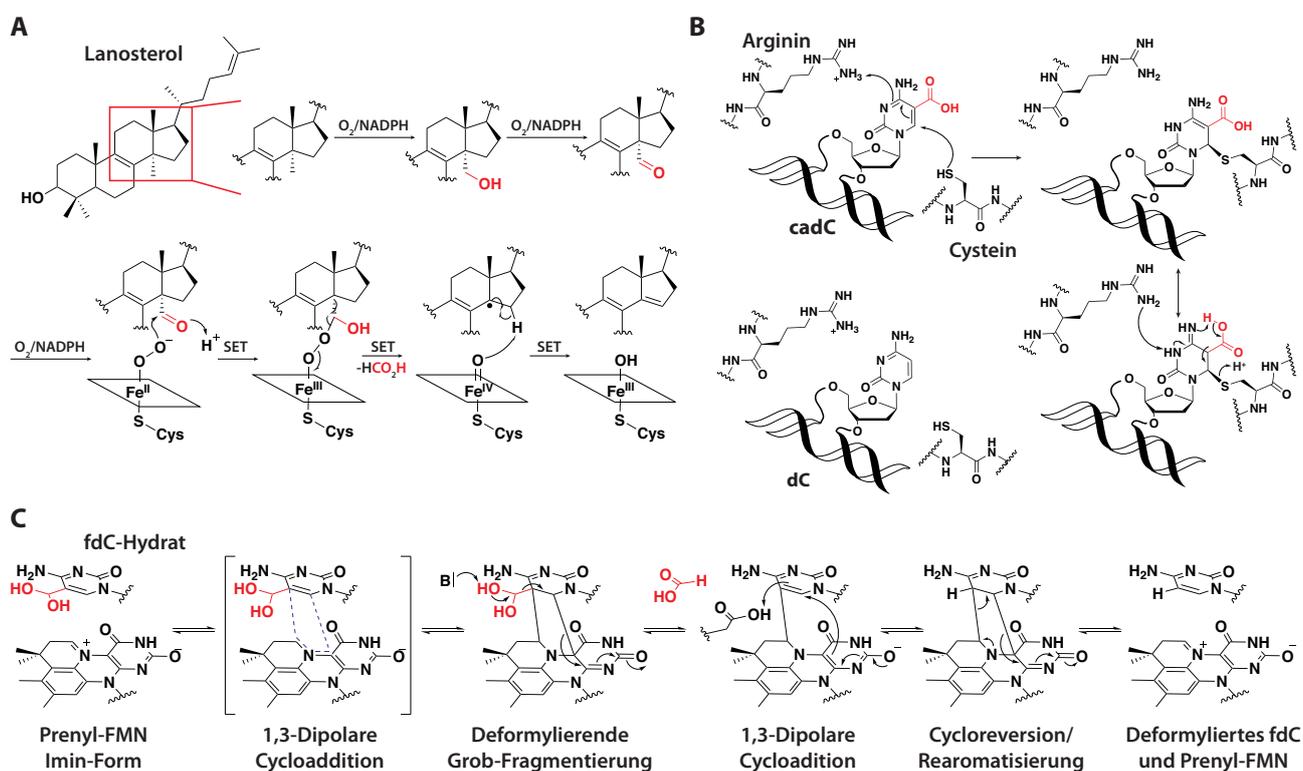


Abbildung 1.10: **A** Oxidative Demethylierung der C14 α -Position in Lanosterol. Die Entfernung der Formylgruppe in Form von Ameisensäure erfolgt durch eine Eisen-Oxo-Spezies. Frei bearbeitet nach *Carell et al.*^[92] **B** Möglicher Mechanismus der direkten Decarboxylierung von cadC zu dC. Cystein fungiert als Helfernukleophil, um die C5-C6-Bindung zu sättigen. Nach erfolgter Imin-Enamin-Tautomerie kann über einen sechs-Gliedrigen-Übergangszustand eine Decarboxylierung erfolgen. Arginin fungiert als Protonen-Shuttle und nach Rearomatisierung des Systems wird das Cystein-Helfernukleophil wieder abgespalten. **C** Möglicher Mechanismus der Deformylierung über ein prenyliertes Flavinmononukleotid.

1 Einleitung

Ein bis dato absolut neues Konzept einer möglichen Deformylierung wurde kürzlich von *Thomas Carell* diskutiert. 2015 wurde von *Payne et al.* ein neuartiges, prenyliertes Flavinmononukleotid charakterisiert (siehe **Abbildung 1.10 C**), welches als Cofaktor bei den sogenannten UbiX/Fdc Proteinen eine Decarboxylierungsreaktion verschiedener Zimtsäurederivate katalysieren konnte.^[198] **Abbildung 1.10 C** zeigt, wie mit Hilfe dieses Cofaktors eine Deformylierungsreaktion an fdC prinzipiell ablaufen könnte. Ein an der Formylgruppe hydratisiertes fdC könnte dabei zunächst über eine 1,3-Dipolare Cycloaddition ein pyrrolidin-artiges Intermediat erzeugen, welches nach einer deformylierenden, *Grob*-artigen Fragmentierung wieder zerfallen kann. An diesem Punkt wäre das C5-C6-ungesättigte Nukleotid mit der C5-Position kovalent am prenylierten FMN gebunden, wobei die Formylgruppe in Form von Ameisensäure frei würde. Durch erneute 1,3-dipolare Cycloaddition könnte die Doppelbindung über einen Aminosäurerest protoniert werden und in einer darauffolgenden Cycloreversion und Rearomatisierung, welche die Triebkraft dieser Reaktion wäre, würde der Flavin-Cofaktor zurückgewonnen und deformyliertes dC erhalten werden. Voraussetzung für das Vorhandensein eines solchen Mechanismus ist natürlich, dass dieser oder ein ähnlicher Cofaktor ebenfalls in eukaryotischen Zellen vorhanden ist. Durch *in vitro* Experimente könnte man jedoch zunächst ermitteln, ob eine solche neuartige Reaktivität prinzipiell möglich ist und damit ein weiterer potentieller Weg zur aktiven Demethylierung von mdC eröffnet werden würde.

2 Aufgabenstellung

Der Hauptfokus dieser Arbeit lag auf der Untersuchung zweier Mechanismen, welche zur Entfernung von fdC oder cadC im Rahmen der aktiven Demethylierung von mdC zu dC diskutiert werden. Dabei handelt es sich, wie zuvor erwähnt, auf der einen Seite um die BER-gestützte Entfernung von fdC und cadC in mESCs und auf der anderen Seite um eine mögliche direkte Demodifikation von fdC zu dC (**Abb. 2.1 A**). Dazu sollte zunächst ein 2'-(*R*)-Fluor-markiertes fdC an mESCs verfüttert werden, welches dann in die DNA eingebaut und im Zuge der aktiven Demethylierung zu 2'-(*R*)-Fluor-markiertem dC umgesetzt werden könnte (**Abb. 2.1 B**). Durch Verwendung der 2'-(*R*)-Fluor-Markierung wird das markierte fdC um 18u schwerer als natürlich vorkommendes fdC. Somit lassen sich etwaige metabolische Umwandlungen dieser Modifikation leicht mittels UHPLC-MS/MS-Techniken analysieren. Außerdem ist bei Verwendung einer 2'-(*R*)-Fluorierung ein möglicher Reparaturweg über die BER ausgeschlossen, da in diesem Falle ein während der Deglykosylierung am 1'-*C*-Atom intermediär auftretendes Oxycarbeniumion sehr stark destabilisiert sein würde. Damit erlaubt dieses Molekül die getrennte Betrachtung eines möglichen C-C-Bindungsbruchs.

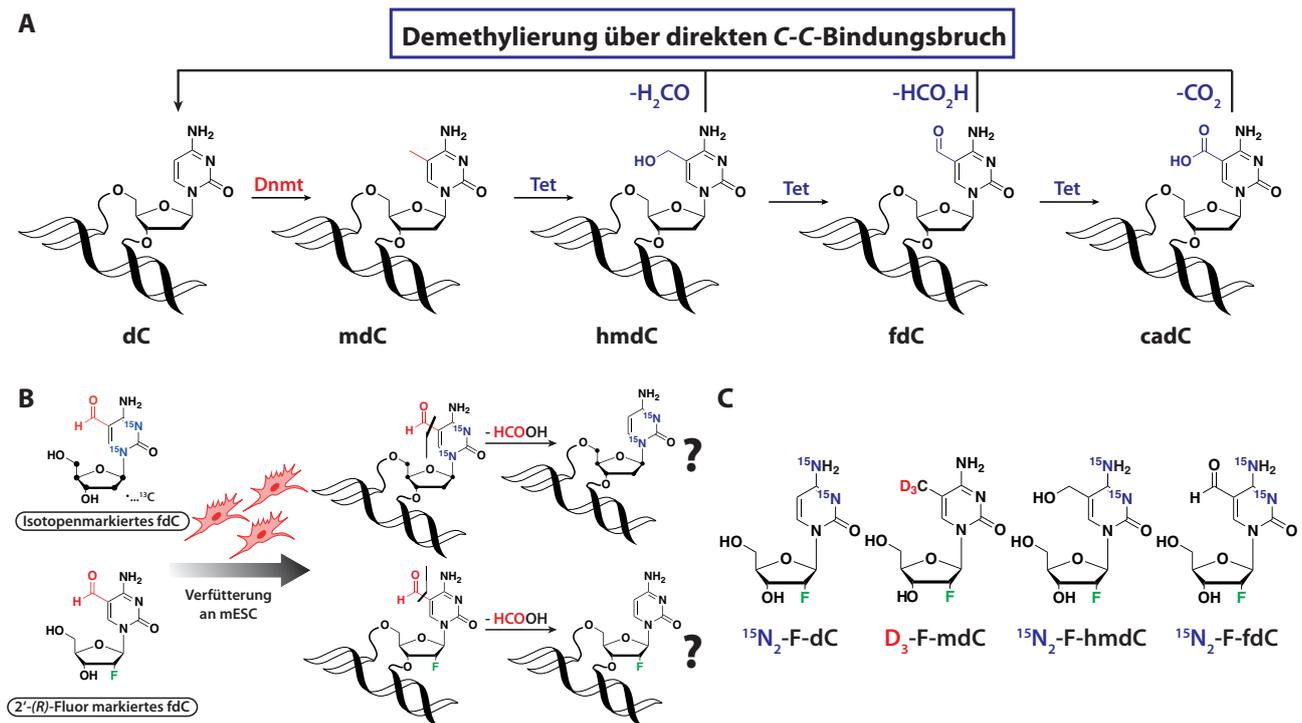


Abbildung 2.1: **A** Übersicht zu den möglichen Mechanismen der aktiven Demethylierung via C-C-Bindungsbruch. **B** Verfütterungsstudien von istopenmarkiertem fdC und 2'-(*R*)-fluoriertem fdC an Zellen. Nach Isolation und enzymatischem Verdau der genomischen DNA auf Nukleosideebene kann eine mögliche Demodifizierungsreaktion von fdC zu dC mittels UHPLC-MS/MS analysiert werden. **C** Strukturen der stabilen Isotopologe von F-dC, F-mdC, F-hmdC und F-fdC, welche für die exakte Quantifizierung aller Bestandteile eines DNA-Hydrolysats nach Verfütterung von F-fdC benötigt werden.

Zur exakten Quantifizierung des nach Verfüterung der fluorierten Testmoleküle erhaltenem DNA-Nukleoside sollten außerdem die Synthese stabiler Isotopologe durchgeführt werden, welche als interne Standards während der UHPLC-MS/MS-Messung fungieren sollten (**Abb. 2.1 C**).

Da eine 2'-(*R*)-Fluor-Gruppe in einem biologischen Kontext nicht vorkommt und somit die Reaktivität des DNA-Heterozyklus beeinflusst sein könnte, sollte des weiteren ein isopenmarkiertes, natürliches fdC synthetisiert und an Zellen verfüttert werden (**Abb. 2.1 B**). Dieses Nukleosid sollte eine Markierung sowohl am Zucker, als auch an der Nukleobase aufweisen, damit bei der späteren UHPLC-MS/MS-Analyse sichergestellt wird, dass Zucker- und Basenfragment während der möglichen Metabolisierung im Kontext der aktiven Demethylierung intakt bleiben.

Diese ersten Experimente sollten zunächst helfen, eine mögliche Demodifizierungsreaktion von fdC zu dC einzugrenzen. In einem zweiten Teilprojekt dieser Arbeit sollte dann untersucht werden, ob und, wenn ja, welche Chemie an der C5-Position stattfindet. Dazu eignen sich C5-deuterierte dC-Spezies, da diese nach einer Dnmt-vermittelten Methylierung zu mdC und fortlaufender Umsetzung in dem in **Abbildung 2.2** gezeigten Reaktionsweg einen Deuterium-Wasserstoff-Austausch zeigen würden.

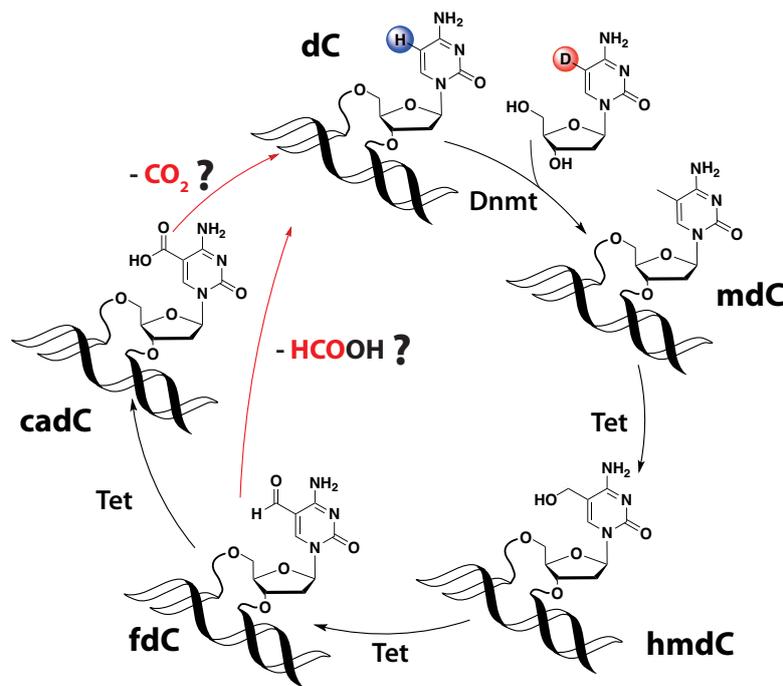


Abbildung 2.2: Durch Verfüterung von C5-markierten dC-Derivaten könnte über einen D/H-Austausch (hier blau umkreist) eine mögliche Beteiligung eines Mechanismus, bei dem ein direkter C-C-Bindungsbruch während der aktiven Demethylierung von mdC zu dC stattfindet, nachgewiesen werden.

Dieser kann über UHPLC-MS/MS verfolgt werden und somit wichtige Hinweise auf eine Deformylierungs- oder Decarboxylierungsreaktion *in vivo* geben. In diesem Zusammenhang sollte der Hauptfokus auf der Synthese einer C5-deuterierten dC-Spezies liegen, welche zum einen eine möglichst hohe Deuterium-Anreicherung an C5-Position aufweisen sollten und zum anderen sollte das dadurch entstehende,

dedeuterte Produkt eindeutig von natürlich vorkommenden dC-Spezies im Nucleosidgemisch unterscheidbar sein.

Die Untersuchung zur Beteiligung der BER-basierten Entfernung von fdC beziehungsweise cadC aus genomischer DNA im Rahmen der aktiven Demethylierung kann über das gemeinsame, kurzlebige Intermediat, der abasischen Stelle, erfolgen (Abb. 2.3 A).

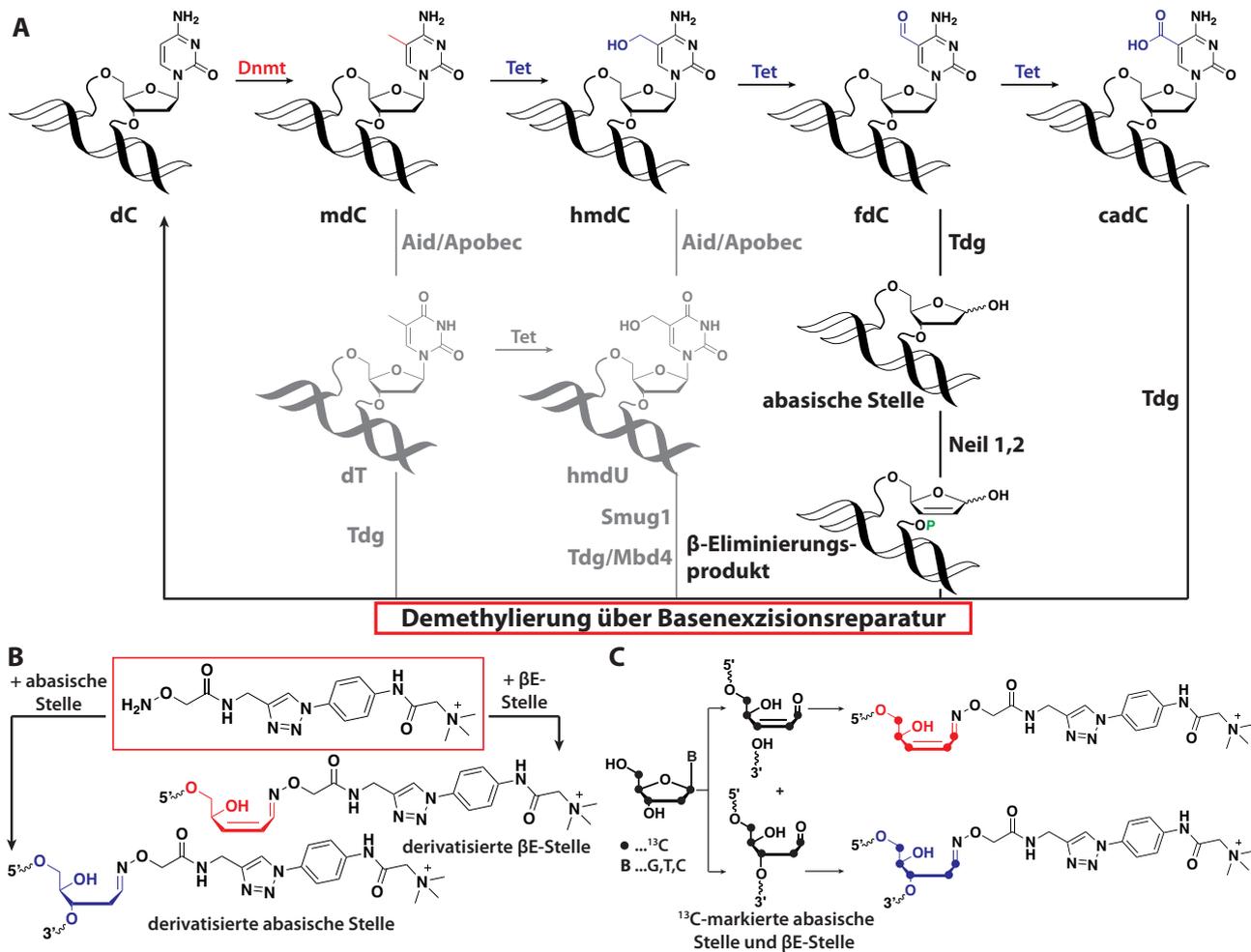


Abbildung 2.3: **A** Aktive Demethylierung im Rahmen der BER. Tdg erzeugt abasische Stellen, welche im weiteren Verlauf durch bifunktionelle Glykosylasen zu β -eliminierten abasischen Stellen umgesetzt werden können. **B** Struktur des Hydroxylamin-Reagenz, mit dem eine sensitive und quantitative Analyse von abasischen und β -eliminierten abasischen Stellen ermöglicht wird. Außerdem sind die dabei in genomischer DNA erhaltenen untersuchten Addukte abgebildet. **C** Durch Verfütterung von isotonenmarkierten Nucleosiden kann der Ursprung der abasischen Stelle zurückverfolgt werden. Mit Hilfe des ^{13}C -markierten Zuckers werden die erhaltenen Addukte um 5u schwerer als ihre natürlichen Derivate.

Dazu sollte ein neuartig entwickeltes Hydroxylamin-Testmolekül angewendet werden, mit welchem die abasischen Stellen in genomischer DNA derivatisiert werden können (Abb. 2.3 B). Dadurch wird die sehr schlecht ionisierbare, underivatisierte abasische Stelle in eine massenspektrometrisch gut zu untersuchende Spezies überführt. Durch Synthese und Zugabe von isotonenmarkierten Standards zum

DNA-Hydrolysat sollte weiterhin eine möglichst sensitive und exakte Quantifizierung der abasischen Stellen sichergestellt werden. Im Gegensatz zu den bisher etablierten Methoden kann durch die Isotopenverdünnungsmethode die genaue Identität der durch das Testmolekül derivatisierten, aldehydischen abasischen Stellen ermittelt werden. Dadurch wird zum ersten Mal auch die gleichzeitige Quantifizierung von β -eliminerten abasischen Stellen ermöglicht, welche sich in ihrer Masse und ihrer Retentionszeit während der UHPLC-MS/MS-Messung deutlich von den abasischen Stellen unterscheiden. Durch Zugabe von an der Ribose ^{13}C -markierten Nucleosiden dG, dT und dC sollte weiterhin der Ursprung der quantifizierten abasischen Stellen ermittelt werden (**Abb. 2.3 C**). Dabei wird eine Unterscheidung von spontan oder durch oxidative Schädigung erzeugten (größtenteils Purine) zu den epigenetisch relevanten abasischen Stellen, die von dC und dessen Derivaten erzeugt werden, ermöglicht.

In einem weiteren Teilprojekt sollte untersucht werden, ob eine direkte Derivatisierung der abasischen Stellen *in cellulo* möglich ist, um diese kurzlebigen Intermediate direkt während einer aktiven Demethylierung abfangen zu können. Da erste Versuche, bei denen das Testmolekül direkt an Zellen verfüttert wurde, fehlschlagen, sollte die reaktive Hydroxylaminfunktion mit einer photolabilen Schutzgruppe maskiert werden. Hierbei sollte eine Schutzgruppe verwendet werden, welche zum einen die Zellmembran passieren kann und zum anderen unter milden Bedingungen wieder abgespalten wird. Letzterer Punkt ist insofern wichtig, da die meisten etablierten Schutzgruppen mit Hilfe von UV-Licht entschützt werden. Dadurch können jedoch artifiziell abasische Stellen generiert werden, die ein hohes Hintergrundsignal bei der Untersuchung dieses Intermediats hervorrufen würden. Letztendlich sollte durch die direkte *in cellulo* Anwendung des Derivatisierungsreagenz ein tieferer Einblick in die dynamische Umsetzung der epigenetisch relevanten Basen während der aktiven Demethylierung gewährt werden.

3 Veröffentlichte Arbeiten

3.1 Deformylierung von 5-Formyldesoxycytidin zu Desoxycytidin in genomischer DNA als alternativer aktiver Demethylierungsweg über einen C-C-Bindungsbruch

„5-Formylcytosine to cytosine conversion by C-C bond cleavage *in vivo*“ Katharina Iwan[#], René Rahimoff[#], Angie Kirchner[#], Fabio Spada[#], Arne S. Schröder, Olesya Kosmatchev, Shqiponja Ferizaj, Jessica Steinbacher, Edris Parsa, Markus Müller, Thomas Carell, *Nature Chem. Bio.* **2018**, *14*, 72 – 78.

[#]: Diese Autoren haben zu gleichen Teilen beigetragen.

Prolog

Wie genau die aktive Demethylierung von mdC zu dC im Rahmen der Genexpressionskontrolle aktiv vonstatten geht, ist bis heute noch nicht eingehend geklärt und damit Gegenstand kontroverser Diskussionen. Diese wurden vor allem beflügelt, nachdem 2009 mit der Wiederentdeckung des hydroxymethylierten DNA-Nukleotids hmdC in genomischer DNA ein Meilenstein in der Epigenetik gelang. Ferner wurde gezeigt, dass diese Modifikation aktiv von mdC mit Hilfe der Tet-Proteine generiert werden kann und somit nicht nur, wie lange zuvor geglaubt, ein Produkt oxidativer Schädigung ist. Kurz darauf wurden weiterhin die höher oxidierten Derivate fdC und cadC im Genom embryonaler Stammzellen entdeckt. Erste Hinweise deuteten darauf hin, dass diese Derivate durch die Glykosylase Tdg unter Bildung einer abasischen Stelle entfernt und diese dann letztendlich im Rahmen der BER durch unmodifiziertes dC wieder aufgefüllt werden könnten. Neben diesem BER-basierten Vorgang wurde außerdem vorgeschlagen, dass eine Entfernung dieser oxidierten Derivate auch prinzipiell durch einen direkten C-C-Bindungsbruch ablaufen könnte. Dabei geht man von einer direkten Decarboxylierung von cadC beziehungsweise Deformylierung von fdC zu dC aus, wobei dieser weitere Weg der direkten Demethylierung eine Erzeugung von potentiell mutagenen abasischen Stellen vermeiden könnte.

In der vorliegenden Studie wurde gezeigt, dass eine direkte Demodifikation von fdC zu dC tatsächlich *in vivo* möglich ist. Durch Verfütterung von isotoopenmarkierten oder an der 2'-Position der Ribose mit Fluor derivatisierten fdC Nukleosiden an Zellkulturen konnte mit Hilfe von Verfolgungsexperimenten beobachtet werden, dass aus dem zugegebenen fdC-Derivat ein dC-Derivat entsteht. Die Experimente wurden dabei so geplant, dass eine Verfälschung der Beobachtungen durch unbekannte, BER-basierte Prozesse ausgeschlossen werden konnte und somit die Demodifikation auf einen direkten C-C-Bindungsbruch zurückzuführen sind. Weiterhin wurde gezeigt, dass die Demodifizierung nicht im Cytosol vor dem Einbau in die genomische DNA sondern vermutlich direkt im Genom stattfindet. Außerdem wurde ausgeschlossen, dass eine Deformylierung der verwendeten fdC Derivate spontan geschieht, was wiederum auf einen aktiven Demodifizierungsmechanismus innerhalb der Zelle hindeutet.

Die beobachtete Reaktion war ferner unabhängig von einer Tet-vermittelten Oxidation zu cadC und kann neben embryonalen Stammzellen auch in verschiedenen somatischen Zellen stattfinden.

Autorenbeitrag

In dieser Arbeit war ich für die Entwicklung und Durchführung der Syntheseroute des $^{13}\text{C}_5$ - $^{15}\text{N}_2$ -fdC Derivats zuständig. Ferner wurde das von *Arne Schröder* entwickelte Protokoll zur Synthese von 2'-(*R*)-fdC verwendet und alle Nukleoside vor der Verfütterung an Zellkulturen von mir mittels HPLC mehrmals aufgereinigt. Neben unmarkierten, flouierten fdC Derivaten war ich für die Entwicklung und Durchführung der Syntheseroute von $^{15}\text{N}_2$ -2'-(*R*)-fdC, $^{15}\text{N}_2$ -2'-(*R*)-hmdC, $^{15}\text{N}_2$ -2'-(*R*)-dC und D₃-2'-(*R*)-mdC verantwortlich, womit letztlich eine exakte Quantifizierung der verschiedenen Metaboliten ermöglicht wurde. UHPLC-LC-MS/MS Messungen und Methodenentwicklungen wurden von *Katharina Iwan* durchgeführt und ausgewertet. Arbeiten in der Zellkultur und Probenvorbereitungen wurden von *Angie Kirchner* und *Fabio Spada* durchgeführt. Alle Experimente und erhaltene Daten wurden in enger Zusammenarbeit mit *Katharina Iwan*, *Angie Kirchner*, *Fabio Spada*, *Arne Schröder*, *Markus Müller* und *Thomas Carell* geplant und diskutiert.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags; Copyright 2018 *Nature Publishing Group*.

5-Formylcytosine to cytosine conversion by C–C bond cleavage *in vivo*

Katharina Iwan^{1,2} , René Rahimoff^{1,2} , Angie Kirchner^{1,2} , Fabio Spada^{1,2} , Arne S Schröder¹, Olesea Kosmatchev¹, Shqiponja Ferizaj¹, Jessica Steinbacher¹, Edris Parsa¹, Markus Müller¹  & Thomas Carell^{1*} 

Tet enzymes oxidize 5-methyl-deoxycytidine (mdC) to 5-hydroxymethyl-dC (hmdC), 5-formyl-dC (fdC) and 5-carboxy-dC (cadC) in DNA. It was proposed that fdC and cadC deformylate and decarboxylate, respectively, to dC over the course of an active demethylation process. This would re-install canonical dC bases at previously methylated sites. However, whether such direct C–C bond cleavage reactions at fdC and cadC occur *in vivo* remains an unanswered question. Here we report the incorporation of synthetic isotope- and (R)-2'-fluorine-labeled dC and fdC derivatives into the genome of cultured mammalian cells. Following the fate of these probe molecules using UHPLC–MS/MS provided quantitative data about the formed reaction products. The data show that the labeled fdC probe is efficiently converted into the corresponding labeled dC, most likely after its incorporation into the genome. Therefore, we conclude that fdC undergoes C–C bond cleavage in stem cells, leading to the direct re-installation of unmodified dC.

Modification of genomic cytosine modulates the interaction of DNA-binding factors with the genome, thus affecting gene expression and chromatin structure^{1,2}. The primary and most prevalent modification is methylation to mdC, which in mammals is catalyzed by the DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b, at least partly in co-operation with the catalytically inactive Dnmt3l³. Because Dnmt1 is a maintenance methyltransferase that copies the methylation pattern during replication, the information that this pattern conveys is inherited through cell division. Genomic mdC can be iteratively oxidized to hmdC^{4,5}, fdC^{6,7} and cadC^{7,8} by the Ten-eleven translocation (Tet) family of α -ketoglutarate-dependent dioxygenases (Fig. 1a). These oxidized cytosine derivatives are prominently detected in DNA isolated from neuronal tissues^{5,9,10} and in the genome of embryonic stem cells (Fig. 1b), in which their levels change during differentiation^{4,7,11}. For example, hmdC can reach levels of up to 1.3% per deoxyguanosine (dG) in DNA isolated from brain¹². Although the presence of mdC and hmdC is believed to influence the transcriptional activity of genes^{13,14}, no clear function has yet been assigned to the higher oxidized modifications fdC and cadC. Recent reports, however, show that fdC is a stable¹⁵, or at least semi-stable¹⁶, base in the genome. These discoveries and the identification of specific reader proteins that recognize fdC and cadC suggest that they might have regulatory purposes as well^{17–20}. So far, however, fdC and cadC are mainly considered to be intermediates of an active demethylation process that allows cells to replace mdC by a canonical dC nucleotide^{20–22}. One such scenario involves fdC and cadC as substrates of the thymine-DNA glycosylase (Tdg), which cleaves the corresponding glycosidic bond. This converts fdC and cadC into abasic sites, which are further processed through base excision repair (BER) as depicted in Figure 1a. This Tdg-initiated process establishes an active demethylation pathway, ultimately incorporating unmodified dC nucleotides at former fdC and cadC sites^{8,23}. A problem associated with this mechanism is that the removal of every mdC creates a potentially harmful single-strand break intermediate. If an mdC is close to the first in the opposite DNA strand, even double strand breaks may be generated. In addition to these concerns, it was shown

that both maternal and paternal genomes of mouse zygotes undergo active demethylation independently of Tdg²⁴. To explain such an excision-independent demethylation and provide an alternative to generating harmful repair intermediates, it was suggested that fdC and cadC may directly deformylate and decarboxylate, respectively, under C–C bond cleavage (Fig. 1a)^{9,25}. Indeed, chemical pathways that allow such a direct deformylation and decarboxylation of fdC and cadC have been described^{26,27}. These pathways involve addition of a helper nucleophile to the C6 position of fdC and cadC in a Michael-addition-type reaction followed by deformylation or decarboxylation and final elimination of the helper nucleophile²⁶. The chemistry is therefore quite similar to the known reaction mechanisms employed by the Dnmt proteins³. Although chemically plausible, it is unclear whether such direct C–C bond cleavage reactions occur within the genome²⁸. This process would provide a new and harmless way to convert mdC back into dC in the genome without forming potentially harmful abasic site intermediates.

Here we report a sensitive MS-based isotope tracing study investigating whether a C–C bond cleavage reaction occurs on fdC bases (Fig. 1c). We supplemented the medium of cultured mammalian cells with synthetic isotope- and fluorine-labeled fdC derivatives to metabolically integrate the nucleosides as reporter molecules into their genome. After isolation of the genomic DNA, the levels of the modified dC derivatives were measured by ultra-high performance liquid chromatography coupled to tandem mass spectrometry (UHPLC–MS/MS), thereby tracing isotopically or fluorine-labeled dC derivatives.

RESULTS

Isotopically labeled fdC is directly converted into dC

We started the study with a [¹³C₅][¹⁵N₂]fdC (**1**; Fig. 2a), in which all five C-atoms of the ribose ring were exchanged with ¹³C and the two in-ring nitrogen atoms replaced with ¹⁵N (Supplementary Note). This provides compound **1**, which is seven mass units heavier than naturally occurring fdC and hence easily distinguishable by MS. The large mass difference allows exact tracing of all transformations that may take place on this base with high sensitivity, because the natural

¹Center for Integrated Protein Science Munich CiPSM at the Department of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany.

²These authors contributed equally to this work. *e-mail: Thomas.carell@lmu.de

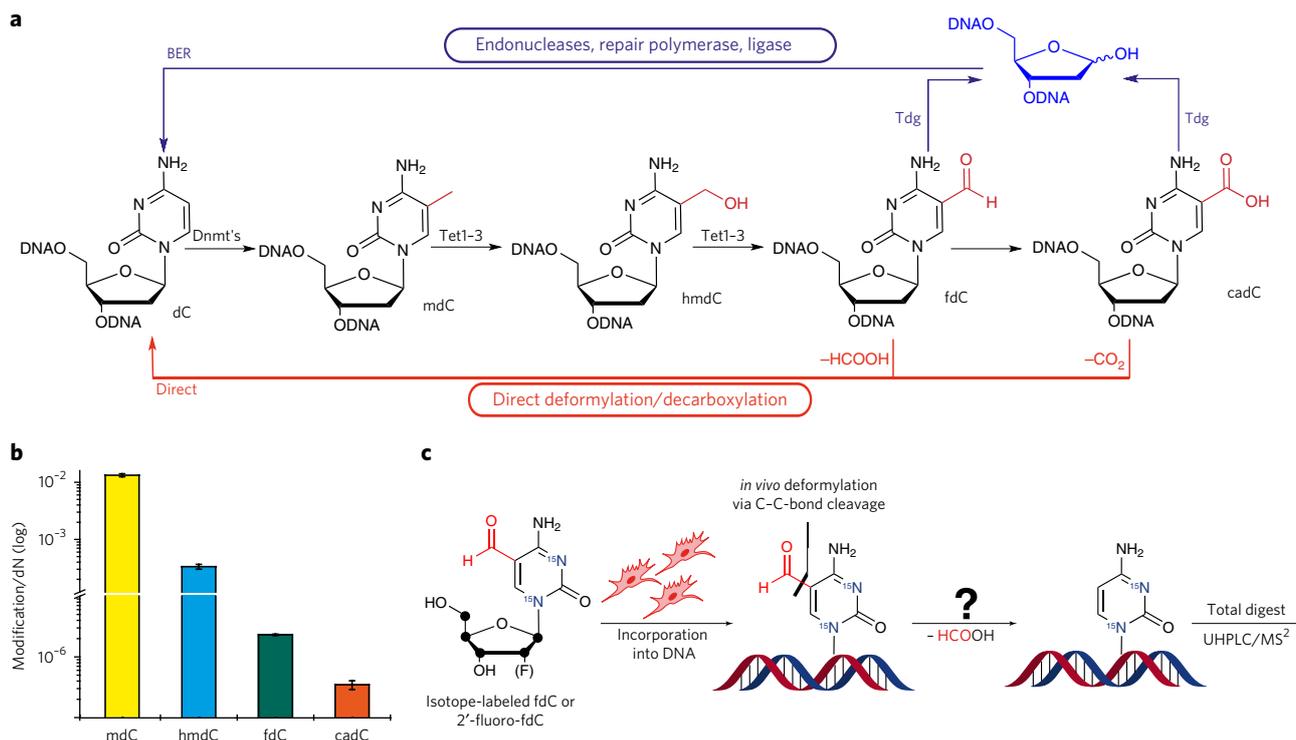


Figure 1 | Isotope tracing studies. (a) Suggested pathways of active demethylation. Thymine-DNA glycosylase (Tdg)-based cleavage of the glycosidic bond of 5-formyl-dC (fdC) and 5-carboxy-dC (cadC) results in an abasic site, which initiates a BER process that leads to the replacement of fdC and cadC by canonical dC (blue). Deformylation of fdC (–HCOOH) and decarboxylation (–CO₂) of cadC provides dC directly (red). (b) UHPLC-coupled MS/MS experiments allow exact quantification of various dC derivatives in mESCs. Mean and s.d. of technical triplicates from two independent cultures are shown. (c) Schematic depiction of the feeding experiment using synthetic isotope and fluorine-labeled fdC derivatives, which are metabolically integrated into the genome. ● = ¹³C atoms.

abundance of such highly isotopically modified dC derivatives is essentially null. Possible transformations are the deformylation of **1** to [¹³C₅][¹⁵N₂]dC (**2**) and its deamination to [¹³C₅][¹⁵N₂]dU (**3**), followed by methylation of **3** to [¹³C₅][¹⁵N₂]dT (**4**). Alternatively, compound **1** can deaminate to [¹³C₅][¹⁵N₂]fdU (**5**) and, finally, the deformylated compound **2** can be methylated to [¹³C₅][¹⁵N₂]mdC (**6**). Analysis of the MS pattern of **1** showed that cleavage of the glycosidic bond is the dominant fragmentation pathway. This leads to a clearly detectable fingerprint mass transition of $m/z = 263.1$ to $m/z = 142.1$ (Fig. 2b). Detection of the demodified product dC **2** would be possible on the basis of its mass transition from $m/z = 235.1$ to $m/z = 114.0$. For the experiment, we added **1** to the medium of mouse embryonic stem cells (mESCs) under priming conditions. After 3 d, the genomic DNA was isolated using a standard protocol and digested to the individual nucleosides. The obtained nucleoside mixture was analyzed by UHPLC coupled to a triple quadrupole mass spectrometer. We noted that **1** was indeed metabolically incorporated into the genome of mESCs. The mass transition of **1** ($m/z = 263.1$ to $m/z = 142.1$) was clearly detectable at a retention time of 5.50 min under our conditions (Fig. 2b). By using the mass transitions specific for all the expected natural dC derivatives, we were also able to detect next to **1**, natural mdC, hmdC and fdC (Fig. 1b).

Analysis of the nucleoside mixtures revealed the presence of a new isotope-labeled dC derivative at a retention time of 1.95 min displaying the expected mass transition ($m/z = 235.1$ to $m/z = 114.0$) for **2**, showing that [¹³C₅][¹⁵N₂]fdC is indeed demodified (Fig. 2c). To unequivocally prove the identity of **2**, an even heavier isotopically modified dC isotopologue, [¹³C₉][¹⁵N₃]dC (**7**), with a characteristic MS transition of $m/z = 240.1$ to $m/z = 119.1$ (Supplementary Results, Supplementary Fig. 1), was used as an internal standard. Compound

7 was added to the nucleoside mixture and it eluted at the same retention time as **2** (Fig. 2c), confirming that the UHPLC–MS/MS signal at 1.95 min is caused by **2**. Exact quantification of the conversion showed that when **1** was supplied to mESC cultures at 50 μM for 3 d, steady state incorporation levels of about $3\text{--}5 \times 10^{-7}$ of [¹³C₅][¹⁵N₂]fdC per deoxynucleoside (dN) in genomic DNA were reached (Fig. 2d). We observed higher levels of product **2** (up to a factor of 10), as shown in Figure 2d.

Whereas we can exclude spontaneous deformylation of **1** based on previous stability studies²⁶, **2** can in principle form either by C–C bond cleavage in the genome or at the level of the soluble nucleoside/nucleotide pool. Conversion of **1** in the soluble pool to **2** would then be followed by metabolic incorporation of the **2**-triphosphate into the genome. It is known that soluble **2** is the substrate for other metabolic processes such as deamination to **3** (catalyzed by cytidine deaminase and deoxycytidylate deaminase), which is followed by methylation by thymidylate synthase to give **4** (refs. 29,30). To distinguish the two possible conversion scenarios (genomic DNA or soluble pool), we reasoned that if **1** is converted into **2** in the soluble pool, we would find compounds **3** and, particularly, **4** in the genome.

To investigate the behavior of soluble dC, we cultured mESCs in the presence of an isotopically labeled dC derivative, [¹³C₉][¹⁵N₃]dC, and indeed detected the expected presence of the corresponding isotopically labeled deamination products [¹³C₉][¹⁵N₂]dT and [¹³C₉][¹⁵N₂]dU (**8** and **9**, respectively) in the genome (Fig. 2e). In contrast, when **1** was supplied to mESC cultures, we detected next to **2** only the direct deamination product **5** in the genome, but not **3** and **4** (Fig. 2d). Even after three consecutive days of feeding compound **1** to mESCs, we were unable to detect even traces of **4** in the genome. This argues against formation of **2** in the soluble pool. We next

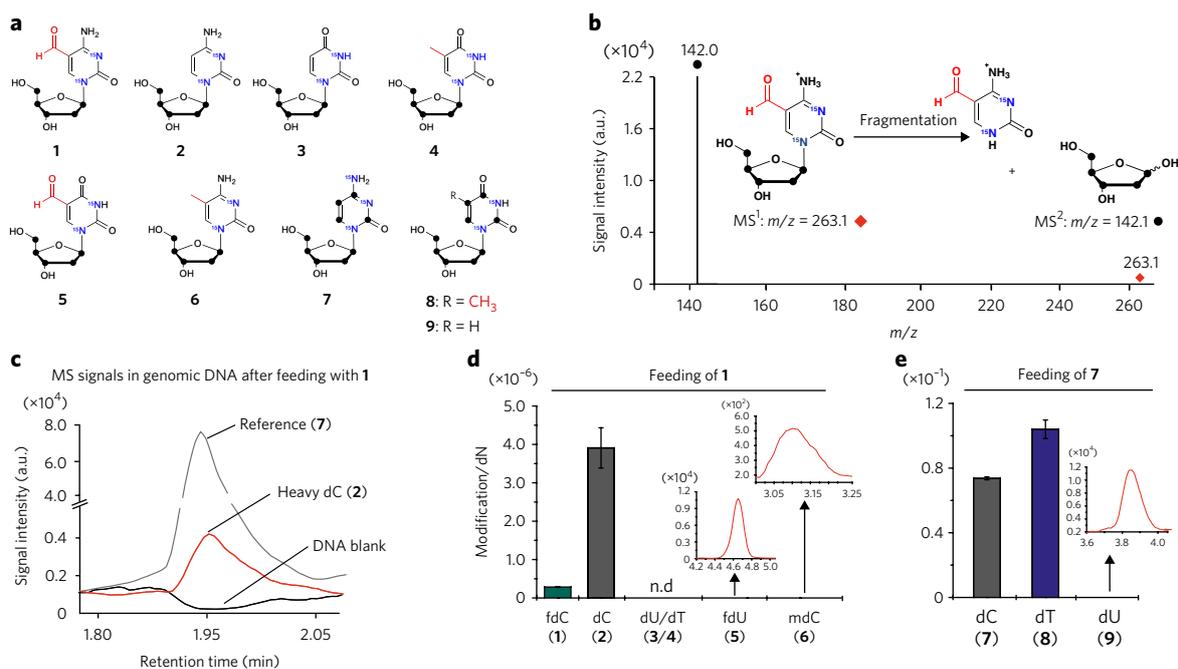


Figure 2 | Conversion of isotopically labeled fdC into dC in mESCs. (a) Overview of the compounds that may be detected after feeding of **1** or **7** to mESCs. (b) Feeding of **1** to mESCs results in incorporation of the isotopologue into the genomic DNA, as proven by its fingerprint MS transition. a.u., arbitrary units. (c) Analysis of gDNA after feeding of **1** shows the presence of labeled dC **2**. (d) Quantitative data obtained upon feeding **1** to mESCs. Mean and s.d. of technical triplicate measurements from a single culture are shown. (e) Quantitative data obtained upon feeding of **7** to mESCs. Mean and s.d. of technical triplicates from two independent cultures are shown. For a schematic overview of the dC or fdC metabolic pathways see **Supplementary Figure 3**.

analyzed the soluble nucleoside/nucleotide pool directly for the content of **2** after feeding of **1** (**Supplementary Fig. 2**). To this end, we fed **1** to mESCs over 3 d. The cells were washed extensively and finally resuspended in 50% (v/v) MeCN to extract soluble metabolites. After further purification by solid-phase extraction, the nucleotides were dephosphorylated to nucleosides. Analysis of this solution by UHPLC-MS/MS did not give any signal for **2**. All these control experiments suggest that **1** undergoes C-C bond cleavage to **2** directly in the genome and not in the soluble pool, although the latter scenario cannot be fully ruled out because of the complexity of the metabolic pathways (**Supplementary Fig. 3**). Interestingly, we also noted the presence of the remethylated product **6** in the genome of mESCs fed with **1**, but because of the low signal intensity we were unable to obtain quantitative data (**Fig. 2d**).

2'-fluorinated cytosines detect biochemical conversions

To reach higher sensitivity, we experimented with various other isotope-labeled fdC derivatives, and finally found that 2'-fluorinated dC derivatives **10-17** are excellent probe molecules (**Fig. 3a** and **Supplementary Fig. 4**). The exchange to an F-atom makes the compound 18 atom units heavier. The compounds also have a slightly shifted retention time (**Fig. 3b**) and give sharp signals in the UHPLC-MS/MS analysis because of a glycosidic bond that is more labile in the MS-fragmentation step. Furthermore, the 2'-(R)-configured compounds are well tolerated by the cells used for this study. The F-substituent does affect the ability of the molecule to undergo further biochemical conversions, but the effect is small. (R)-2'-F-dC (**10**), for example, is efficiently methylated by DNA methyltransferases³¹ and (R)-2'-F-mdC (**11**) is also oxidized to (R)-2'-F-hmdC (**12**) by the Tet enzymes (**Fig. 3a,c**), although a reduced speed of oxidation is observed³².

To show that the fluorinated compounds are valid probe molecules, we first added **10** to the mESC culture at 0.5 μ M, 1.0 μ M or 2.5 μ M for 3 d. Under these conditions, UHPLC-MS/MS analysis

of the isolated genomic DNA showed a clear dose-dependent integration of **10** into the genome, up to 1×10^{-3} per dN. We next searched for other 2'-fluorinated pyrimidine nucleosides and detected a dose-dependent presence of (R)-2'-F-dU (**13**) and (R)-2'-F-dT (**14**), which was formed by deamination of **10** to **13** followed by methylation to **14** (**Fig. 3c**). In addition, we detected a dose-dependent formation of (R)-2'-F-mdC and (R)-2'-F-hmdC, confirming that compound **10** is biochemically converted, as expected (**Fig. 3c**).

To quantify the levels of methylation, we synthesized the isotope-labeled compounds (R)-2'-[D₃]-F-mdC (**18**), (R)-2'-[¹⁵N₂]-F-dC (**19**) and (R)-2'-[¹⁵N₂]-F-fdC (**20**) and used them as internal standards for quantification (**Fig. 3a**). Upon feeding mESCs with 1 μ M **10** for 3 d, we detected around 3% ($\pm 0.5\%$) of **11** relative to **10**, which is similar to proportions observed for the natural bases (**Supplementary Fig. 5**).

(R)-2'-F-fdC is converted into (R)-2'-F-dC in mESCs

To study the direct C-C bond cleavage process, we again cultured mESCs in the presence of (R)-2'-F-fdC (**15**; 350 μ M for 3 d), isolated the DNA and analyzed the nucleoside composition. Next to genomic (R)-2'-F-fdC (5.7×10^{-7} /dN), we detected (R)-2'-F-dC at a level of 7.3×10^{-6} /dN (**Fig. 3d**). Because the nucleosides **13** and **14** were not detected, we suspected again that the observed reaction of **15** to **10** occurs directly within the genome. The detection limit of **13** and **14** is, however, around 5 fmol, and so the compounds may just escape observation.

To substantiate the conclusion that genomic **15** undergoes intragenomic C-C bond cleavage to **10**, we first assessed whether **15** can spontaneously deformylate. To investigate this possibility, we heated an aqueous solution of **15** to 60 °C for 3 d; however, we did not detect compound **10**. We also incubated **15** in culture medium for 3 d and were unable to detect any **10**. Finally, we added a 28-mer oligonucleotide containing a single **15** to culture medium for 3 d, then re-isolated the DNA strand and searched for **10**. Formation

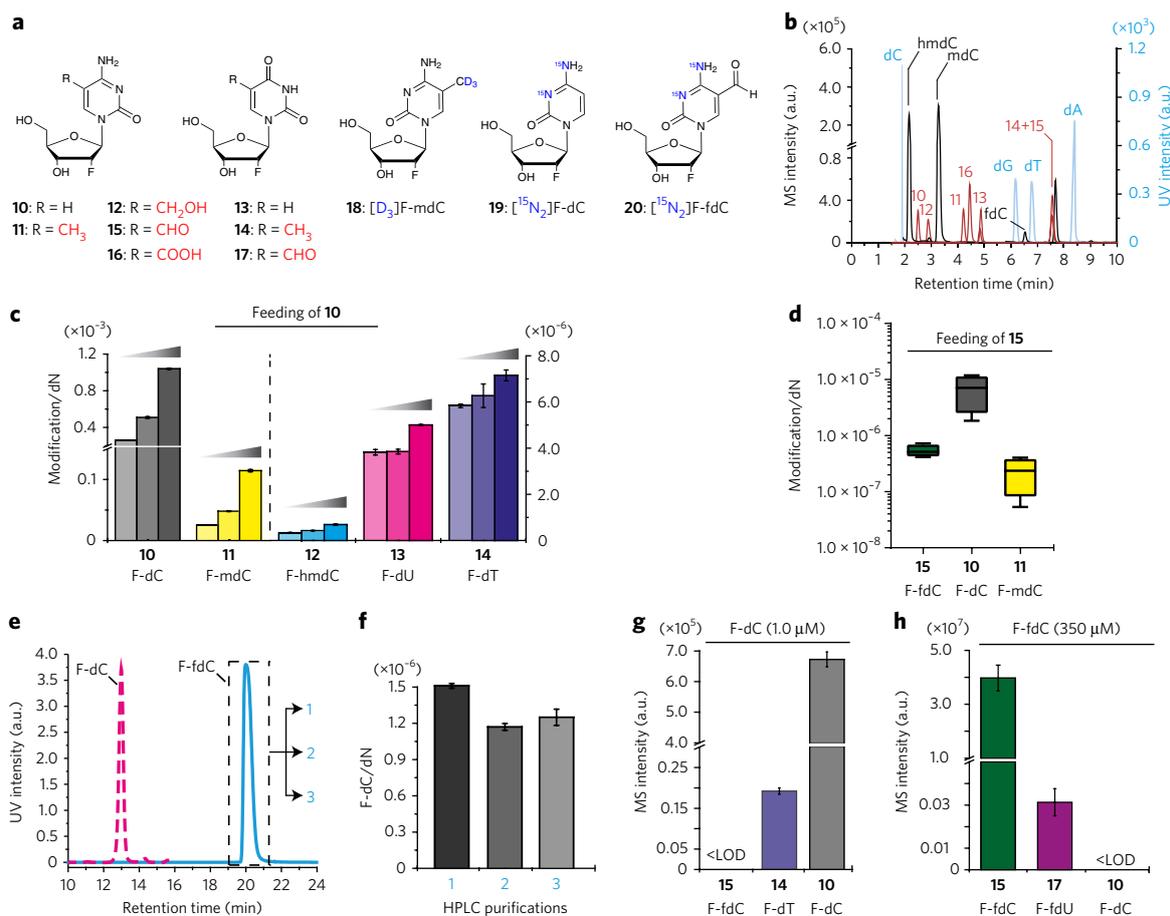


Figure 3 | F-fdC is converted into F-dC within the genome. (a) Chemical structures of 2'-fluorinated dC and dU derivatives that were investigated and internal standards used. (b) Resulting UV and MS traces of the nucleosides under investigation. Light blue, canonical bases; black, natural dC derivatives; red, fluorinated bases. (c) Quantitative data of fluorinated pyrimidine derivatives after feeding with **10** for 3 d at different concentrations (0.5 μM, 1.0 μM and 2.5 μM). A DNA sample from a single culture was measured as technical triplicates. (d) Quantitative data of fluorinated pyrimidine derivatives after feeding with **15** at 350 μM for 3 d. Technical triplicates from four independent cultures were measured. (e) HPLC of three consecutive purifications of **15** (1–3, blue line). Material from the dashed-line box was collected. The purple dashed line marks the position where a peak of contaminating F-dC would be expected. (f) Quantitative data after feeding of **15** that has been purified three consecutive times (from e). The levels of the deamination product **10** remain the same, ruling out any contribution from a possible contamination. (g, h) Quantitative data of fluorinated pyrimidine derivatives in the soluble pool after feeding **10** at 1 μM (g) and **15** at 350 μM (h) for 3 d. Technical triplicates from single cultures were measured. In c, e–h mean values with s.d. are shown.

of **10** was again not detected. Together, these experiments exclude background deamination.

We next analyzed the possibility that **15** is contaminated with traces of **10**. To this end, the purity of **15** was checked by MS, and indeed, **10** was not found. To exclude the presence of even traces of **10** below the detection limit, we performed three consecutive HPLC purifications of **15**. This compound **15** is well separable from **10** because **10** elutes 7.50 minutes earlier than **15** during the HPLC purification (Fig. 3e). Feeding of the material **15** obtained from three consecutive purifications resulted in unchanged values of genomic **10**, arguing against the possibility that the detected **10** is an accumulated impurity (Fig. 3f).

To further substantiate that the C–C bond cleavage does not occur in the soluble nucleoside/nucleotide pool, we added **10** to the mESC culture for 3 d. UHPLC–MS/MS analysis of the soluble pool allowed us to detect **10**, **14** and, in traces, **13** (Fig. 3g). However, when we repeated the study with **15** (Fig. 3h), we detected just **15** in the soluble pool plus the deaminated compound (*R*)-2'-F-fdU (**17**), but not **10**. We next determined the medium concentration of **10** that would be needed to reach the measured value for genome-integrated **10** (7.3×10^{-6} /dN; Fig. 3d) and found that a concentration of

5–10 nM would be required (Supplementary Fig. 6) in the soluble pool. With a detection limit of 30 amol for **10** (40 μL injection), this is a concentration at which **10** is unambiguously detectable.

All of these control experiments support the idea that the C–C bond cleavage to F-dC occurs within the genome. Interestingly, upon feeding of **15** to mESCs, we also detected the methylated derivative **11**, demonstrating that the demethylated product **10** is methylated to **11** in the genome. Using the isotopically labeled internal standards (*R*)-2'-[D₃]F-mdC, (*R*)-2'-[¹⁵N₂]F-dC and (*R*)-2'-[¹⁵N₂]F-fdC (Fig. 3a), we found that remethylation of **10** results in levels of 2.8% (±0.3%) F-mdC (Fig. 3d), which is only slightly lower compared to the level observed with direct feeding of **10** (Supplementary Fig. 5).

To study the time dependence of the C–C bond cleavage process, we fed compound **15** and measured the genome-integrated levels of **15**, **10** and **11**. Already at 0.5 h, we detected a stable incorporation of **15** (Fig. 4a). The C–C bond cleaved product **10** appeared after about 1 h and the levels increased steadily (Fig. 4b). After about 4 h, we saw the first remethylated product **11** (Fig. 4c). If **10** was a contamination in the preparation of **15**, we would expect faster incorporation kinetics. When we fed both **15** and **10** simultaneously, a steady increase in the level of **10** was already observed after 5 min

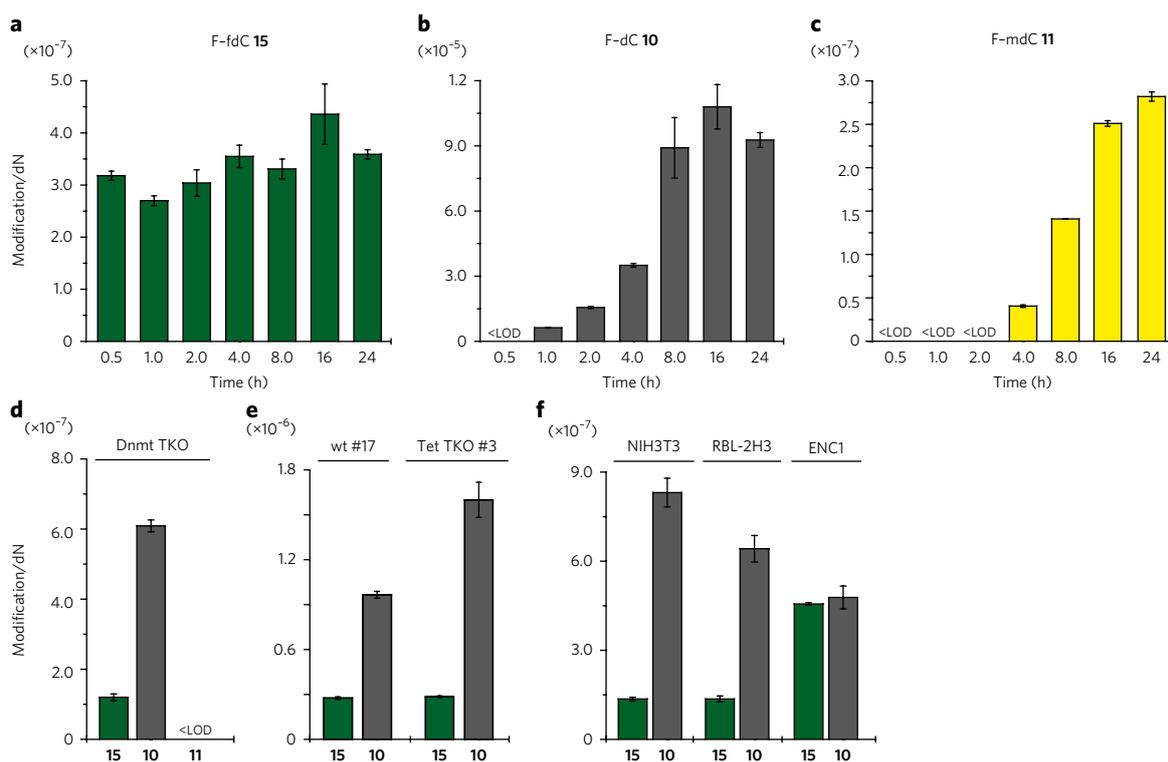


Figure 4 | Demodification of 2'-fluorinated fdC is a rapid process, does not require Dnmt or Tet enzymes and occurs also in somatic cell types.

(a–c) Time-course study showing the genomic build-up of F-fdC **15** (a), F-dC **10** (b) and F-mdC **11** (c) upon metabolic labeling with **15**. LOD, limit of detection. (d–f) Genomic levels of **15** and **10** upon metabolic labeling of Dnmt triple knockout (TKO) mESCs (d), Tet TKO and corresponding wild-type mESC lines (e) and various somatic cell lines (f) with **15**. Genomic levels of **11** are also shown in d. In all panels, mean values and s.d. of technical triplicate measurements from single representative experiments are shown. Data from two additional independent experiments are shown in **Supplementary Figures 8 and 9**.

(**Supplementary Fig. 7**), confirming that our probe nucleosides are quickly incorporated into the genome. These data show that the C–C bond cleavage is a rapid process.

Dnmt or Tet enzymes are not required for demodification

To investigate whether the remethylation of **10** is driven by the known DNA methyltransferases and whether these are involved in demodification, we added **15** (350 μ M for 3 d) to mESCs deficient in all active DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b (Dnmt triple knockout (TKO)) and analyzed the DNA. In this experiment, the demodified product **10** is again detected, but the methylated product **11** is not seen (**Fig. 4d** and **Supplementary Fig. 8**), showing that DNA methyltransferases are responsible for methylation of **10** and are not required for demodification of **15**. We finally investigated whether the Tet enzymes are involved in demodification of **15**. However, repeating the feeding experiment with mESCs lacking all three members of the Tet family (Tet TKO) resulted in full demodification activity (**Fig. 4e** and **Supplementary Fig. 9**). The fact that the conversion of **15** to **10** does not change in the absence of Tet proteins is particularly noteworthy. Because Tet enzymes were shown to accept **11** as a substrate and convert it into **12**, **15** and (R)-2'-F-cadC **16** (ref. 32), we can exclude the possibility that the observed C–C bond cleavage is in fact a Tet-dependent decarboxylation of **16**. Indeed, this result implies either that **15** is directly deformylated to **10** or that factors other than Tet enzymes can oxidize **15** to **16**, which would then be decarboxylated to **10**. Ascorbic acid has been shown to increase Tet enzymatic activity *in vitro* and the levels of oxidized mdC derivatives *in vivo*^{33–35}. Ascorbic acid treatment of mESC cultures fed with **15** indeed resulted in increased levels of naturally occurring fdC and cadC, but had no effect on conversion of **15** into unmodified product **10**

(**Supplementary Fig. 10**). This further supports that demodification of **15** to **10** does not depend on the enzymatic activity of the Tet enzymes.

Finally, we tested whether conversion of **15** into **10** occurs in nonpluripotent cells by feeding **15** to cell lines representing a variety of cell types (**Fig. 4f** and **Supplementary Fig. 11**). Albeit to various degrees, we detected the conversion of **15** to **10** in all these cell lines, arguing that this is rather widespread in mammalian cell types. In summary, our data prove that fdC is converted into dC *in vivo* through C–C bond cleavage and strongly suggest that this conversion is an intragenomic process.

DISCUSSION

In recent years, several mechanisms for active erasure of cytosine methylation from the genome have been proposed²⁵. Among these, the best-established mechanism entails Tet-mediated iterative oxidation of mdC to fdC or cadC followed by the replacement of these higher oxidized derivatives with unmodified dC through BER³⁶. Considering the frequent occurrence of mdC in high-density clusters and prevalent symmetrical configuration at CpG sites in vertebrate genomes, a BER-based erasure mechanism poses a substantial risk of creating clustered single and double strand breaks with potentially deleterious consequences. It is possible that excision of fdC and cadC by Tdg, processing of the abasic site and insertion of unmodified dC are orchestrated by a single multimolecular complex, thus allowing tight control of strand breaks. Alternatively, it is also conceivable that to minimize the potentially deleterious consequences of BER, complementary mechanisms are in place to remove fdC and cadC that do not involve DNA repair. In this context, it should be kept in mind that Tet3-dependent active demethylation of maternal and paternal genomes in the mouse zygote may not require

Tdg²⁴. With an isotope-tracing experiment using labeled dC-derived nucleosides in combination with highly sensitive UHPLC-MS/MS detection, we show here that in mammalian cells fdC is converted to dC while the glycosidic bond is kept intact. Evidence is provided to support the theory that the C–C bond cleavage reaction happens when fdC is located inside the genome. This establishes an intragenomic demodification process independent of DNA repair. We also show that this process does not require any of the Tet-family enzymes. Therefore, unless other factors are able to oxidize fdC to cadC, this demodification process is likely a direct deformylation of fdC. Although we have firmly established the occurrence of a C–C bond cleavage of fdC or cadC to dC, the mechanism of this process remains to be defined, including the identification of the factors that mediate the demodification reactions.

We would like to emphasize that in our approach the probe nucleosides are randomly incorporated into the genome through DNA replication. Consequently, we cannot determine the sequence and genomic context wherein the demodification of fdC or cadC to dC takes place. Obviously, this would require a sequencing approach that allows identification of the converted dC bases. In addition, we detected conversion of 2'-fluorinated fdC to 2'-fluorinated dC in mESCs and different somatic cell types. This indicates that the ability to carry out the demodification reaction may be widespread in mammalian cells and tissues rather than being restricted to events of active genomic demethylation known to occur in specific developmental and tissue contexts. Assuming that the deformylation of fdC to dC establishes an active demethylation pathway, we need to emphasize that deformylation reactions and deformylases are widespread in nature. A prominent example is the enzyme lanosterol 14 α -demethylase (CYP51A1), which oxidizes the C14 α methyl group of lanosterol to a formyl group to achieve deformylation under concomitant introduction of a double bond (dehydrating deformylation). This enzyme, a P450-type monooxygenase, contains a heme cofactor that seems to utilize a nucleophilic Fe peroxyl anion species to attack the substrate^{37,38}. Another well studied enzyme that catalyzes deformylation is the aldehyde-deformylating deoxygenase, which again uses a nucleophilic metal bond peroxyl anion radical as the attacking species. This enzyme shortens fatty acid chains by oxidizing the terminal methyl group to a formyl group, which is followed by deformylation^{39,40}. In contrast to fdC, these deformylation reactions take place on formyl groups attached to saturated C atoms, whereas in fdC the formyl group is linked to an aromatic heterocycle. Such structures are known for decarboxylations, and they are catalyzed by the enzymes orotate⁴¹ and isoorotate⁴² decarboxylase. Indeed, it was suggested that the isoorotate decarboxylase could be a blueprint for a putatively existing cadC decarboxylase⁴³. Our data now support the idea that fdC and possibly also cadC are converted to dC by a direct C–C bond cleavage. The questions as to when and where these reactions occur *in vivo* now require the identification of putative catalytic factors.

Received 8 May 2017; accepted 31 October 2017;
published online 27 November 2017

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

References

- Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
- Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
- Jeltsch, A. & Jurkowska, R.Z. Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm *Nucleic Acids Res.* **44**, 8556–8575 (2016).
- Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
- Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
- Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
- Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
- He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
- Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* **5**, e15367 (2010).
- Münzel, M., Globisch, D. & Carell, T. 5-Hydroxymethylcytosine, the sixth base of the genome. *Angew. Chem. Int. Ed. Engl.* **50**, 6460–6468 (2011).
- Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
- Wagner, M. *et al.* Age-dependent levels of 5-methyl-, 5-hydroxymethyl-, and 5-formylcytosine in human and mouse brain tissues. *Angew. Chem. Int. Ed. Engl.* **54**, 12511–12514 (2015).
- Su, M. *et al.* 5-Formylcytosine could be a semipermanent base in specific genome sites. *Angew. Chem. Int. Ed. Engl.* **55**, 11797–11800 (2016).
- Raiber, E.A. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2015).
- Song, C.X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
- Kellinger, M.W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **19**, 831–833 (2012).
- Zhu, C. *et al.* Single-cell 5-formylcytosine landscapes of mammalian early embryos and ESCs at single-base resolution. *Cell Stem Cell* **20**, 720–731.e5 (2017).
- Hill, P.W., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324–333 (2014).
- Wu, X., Inoue, A., Suzuki, T. & Zhang, Y. Simultaneous mapping of active DNA demethylation and sister chromatid exchange in single cells. *Genes Dev.* **31**, 511–523 (2017).
- 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).
- Guo, F. *et al.* Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* **15**, 447–459 (2014).
- Wu, S.C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
- Schiesser, S. *et al.* Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
- Liutkevičiūtė, Z. *et al.* Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. *J. Am. Chem. Soc.* **136**, 5884–5887 (2014).
- Schiesser, S. *et al.* Mechanism and stem-cell activity of 5-carboxylcytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed. Engl.* **51**, 6516–6520 (2012).
- Jekunen, A. & Vilpo, J.A. 5-Methyl-2'-deoxycytidine. Metabolism and effects on cell lethality studied with human leukemic cells in vitro. *Mol. Pharmacol.* **25**, 431–435 (1984).
- Vilpo, J.A. & Vilpo, L.M. Biochemical mechanisms by which reutilization of DNA 5-methylcytosine is prevented in human cells. *Mutat. Res.* **256**, 29–35 (1991).
- Schröder, A.S. *et al.* Synthesis of (R)-configured 2'-fluorinated mC, hmC, fC, and caC phosphoramidites and oligonucleotides. *Org. Lett.* **18**, 4368–4371 (2016).
- Schröder, A.S. *et al.* 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes. *Chem. Commun. (Camb.)* **52**, 14361–14364 (2016).
- Blaschke, K. *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500**, 222–226 (2013).

34. Minor, E.A., Court, B.L., Young, J.I. & Wang, G. Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J. Biol. Chem.* **288**, 13669–13674 (2013).
35. Yin, R. *et al.* Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J. Am. Chem. Soc.* **135**, 10396–10403 (2013).
36. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.* **18**, 517–534 (2017).
37. Hargrove, T.Y. *et al.* Substrate preferences and catalytic parameters determined by structural characteristics of sterol 14 α -demethylase (CYP51) from *Leishmania infantum*. *J. Biol. Chem.* **286**, 26838–26848 (2011).
38. Lepesheva, G.I. *et al.* CYP51: A major drug target in the cytochrome P450 superfamily. *Lipids* **43**, 1117–1125 (2008).
39. Aukema, K.G. *et al.* Cyanobacterial aldehyde deformylase oxygenation of aldehydes yields n-1 aldehydes and alcohols in addition to alkanes. *ACS Catal.* **3**, 2228–2238 (2013).
40. Jia, C. *et al.* Structural insights into the catalytic mechanism of aldehyde-deformylating oxygenases. *Protein Cell* **6**, 55–67 (2015).
41. Fujihashi, M., Mnpotra, J.S., Mishra, R.K., Pai, E.F. & Kotra, L.P. Orotidine monophosphate decarboxylase—a fascinating workhorse enzyme with therapeutic potential. *J. Genet. Genomics* **42**, 221–234 (2015).
42. Smiley, J.A., Angelot, J.M., Cannon, R.C., Marshall, E.M. & Asch, D.K. Radioactivity-based and spectrophotometric assays for isoorotate decarboxylase: identification of the thymidine salvage pathway in lower eukaryotes. *Anal. Biochem.* **266**, 85–92 (1999).
43. Xu, S. *et al.* Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Res.* **23**, 1296–1309 (2013).

Acknowledgments

Tet TKO mESC lines were kindly provided by G.-L. Xu (Shanghai Institutes for Biological Sciences) and R. Jaenisch (Whitehead Institute, MIT, Cambridge). We are grateful to M. Okano and H. Niwa (both at Kumamoto University, Japan) for providing the Dnmt TKO mESC line and the Oct4-YFP reporter cell line, respectively. A.S.S. is supported by a fellowship from the Fonds der Chemischen Industrie. We thank the Deutsche Forschungsgemeinschaft for financial support through the programs: SFB749 (TP A4), SFB1032 (TP A5), SPP1784 and CA275-11/1. We thank the European Union Horizon 2020 program for funding the ERC Advanced project EPiR (741912). Further support is acknowledged from the Excellence Cluster CiPSM (Center for Integrated Protein Science).

Author contributions

K.I. developed and performed the UHPLC–MS/MS studies. R.R. and A.S.S. synthesized the fluorinated and isotopically labeled nucleosides. A.K. designed and performed cell culture work. F.S. designed, supervised and performed cell culture work. O.K. and J.S. analyzed feeding studies of isotopically labeled dC. S.F. contributed to experiments for the analysis of soluble nucleoside pools. M.M. supervised the biochemical work, interpreted and discussed results. T.C. designed and supervised the study. All members discussed results, interpreted data and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to T.C.

Supplementary Tables

Supplementary Table 1: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with fluorinated nucleosides. 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 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 1.5-3.0 min								
F-dC	246.09	Wide	112.06	Wide	70	15	3	Positive
$[^{15}\text{N}_2]$ -F-dC	248.08	Wide	114.04	Wide	70	15	3	Positive
$[^{15}\text{N}_2]$ -cadC	274.08	Wide	158.03	Wide	40	5	5	Positive
cadC	272.09	Wide	156.04	Wide	40	5	5	Positive
F-hmdC	276.10	Wide	142.06	Wide	50	10	3	Positive
$[^{15}\text{N}_2]$ -F-hmdC	278.09	Wide	144.06	Wide	50	10	3	Positive
$[^{15}\text{N}_2, \text{D}_2]$ -hmdC	262.12	Wide	146.07	Wide	25	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	25	27	1	Positive
$[\text{D}_3]$ -mdC	245.13	Wide	129.09	Wide	50	60	1	Positive
mdC	242.11	Wide	126.07	Wide	50	60	1	Positive
Time segment 3.0-4.7 min								
F-hmdU	275.07	Wide	255.06	Wide	80	3	7	Negative
F-dU	245.06	Wide	225.06	Wide	80	3	5	Negative
$[\text{D}_3]$ -F-mdC	263.12	Wide	129.09	Wide	80	15	3	Positive
F-mdC	260.10	Wide	126.07	Wide	80	15	3	Positive
F-cadC	290.08	Wide	156.04	Wide	80	5	5	Positive
Time segment 4.7-10 min								
F-fdC	274.08	Wide	140.05	Wide	90	15	3	Positive
$[^{15}\text{N}_2]$ -F-fdC	276.08	Wide	142.04	Wide	90	15	3	Positive
F-fdU	273.05	Wide	253.05	Wide	30	3	5	Negative
F-dT	259.07	Wide	239.07	Wide	70	3	5	Negative
dT	243.10	Wide	127.05	Wide	20	5	5	Negative
$[^{15}\text{N}_2]$ -fdC	258.09	Wide	142.04	Wide	30	5	5	Positive
fdC	256.09	Wide	140.05	Wide	30	5	5	Positive
$[^{15}\text{N}_2]$ -fdC	142.04	Wide	98.04	Wide	20	13	7	Positive
fdC	140.05	Wide	97.04	Wide	20	13	7	Positive

Supplementary Table 2: Compound-dependent LC-MS/MS ranges of the corresponding linear equations.

compound	n (ULOQ)	n (LLOQ)	A/A* (ULOQ)	A/A* (LLOQ)
F-fdC	25.1 fmol	0.390 fmol	0.2267	0.003345
F-dC	800 fmol	3.13 fmol	4.074	0.01512
F-mdC	49.9 fmol	0.780 fmol	0.04174	0.0005833

Supplementary Table 3: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with 1. 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 the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

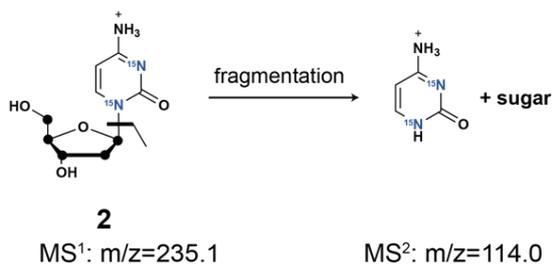
compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 1.5-3.3 min								
sh-cadC	279.10	Wide	158.03	Wide	50	5	5	Positive
$[^{15}N_2]$ -cadC	274.08	Wide	158.03	Wide	20	5	5	Positive
cadC	272.09	Wide	156.04	Wide	20	5	5	Positive
sh-hmdC	265.12	Wide	144.06	Wide	60	27	1	Positive
$[^{15}N_2, D_2]$ -hmdC	262.12	Wide	146.07	Wide	20	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	20	27	1	Positive
sh-mdC	249.12	Wide	128.06	Wide	60	60	1	Positive
$[D_3]$ -mdC	245.13	Wide	129.09	Wide	20	60	1	Positive
mdC	242.11	Wide	126.07	Wide	20	60	1	Positive
sh-dC	235.11	Wide	114.04	Wide	60	10	2	Positive
dC	228.10	Wide	112.05	Wide	15	1	1	Positive
$[^{13}C_9, ^{15}N_3]$ -dC	240.10	Unit	119.10	Unit	70	10	2	Positive
Time segment 3.3-4.15 min								
$[D_2]$ -hmdU	259.09	Wide	216.08	Wide	80	7	5	Negative
hmdU	257.08	Wide	214.07	Wide	80	7	5	Negative
sh-dU	234.08	Wide	192.09	Wide	120	5	5	Negative
$[^{15}N_2]$ -dU	229.06	Wide	185.06	Wide	80	5	5	Negative
dU	227.07	Wide	184.06	Wide	120	5	5	Negative
Time segment 4.15-9 min								
$[^{15}N_2]$ -8-oxo-dG	289.08	Wide	173.04	Wide	10	9	7	Positive
8-oxo-dG	284.10	Wide	168.05	Wide	10	9	7	Positive
sh-fdC	263.10	Wide	142.04	Wide	80	5	5	Positive
$[^{15}N_2]$ -fdC	258.09	Wide	142.04	Wide	80	5	5	Positive
fdC	256.09	Wide	140.05	Wide	10	5	5	Positive
sh-fdU	262.07	Wide	141.01	Wide	80	15	5	Negative
$[^{15}N_2]$ -fdU	257.06	Wide	141.01	Wide	80	15	5	Negative
fdU	255.06	Wide	139.01	Wide	10	15	5	Negative
sh-dT	250.11	Wide	129.04	Wide	20	5	5	Positive
dT	243.10	Wide	127.05	Wide	10	45	3	Positive

Supplementary Table 4: Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA fed with 7. 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 the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

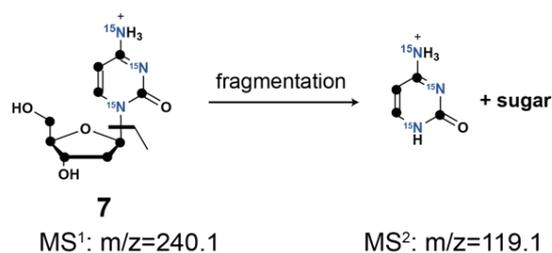
compound	Precursor ion (m/z)	MS1 Resolution	Product Ion (m/z)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
Time segment 1.5-3.3 min								
$[^{13}\text{C}_9, ^{15}\text{N}_3]$ -dC	240.10	Unit	119.10	Unit	70	30	5	Positive
dC	228.1	Wide	112.10	Wide	70	1	3	Positive
Time segment 3.3-4.8 min								
dU	227.07	Wide	184.06	Wide	80	5	5	Negative
$[^{13}\text{C}_9, ^{15}\text{N}_2]$ -dU	238.09	Wide	193.10	Wide	80	5	5	Negative
Time segment 4.8-12.0 min								
dT	243.1	Wide	127.05	Wide	70	40	3	Positive
$[^{13}\text{C}_9, ^{15}\text{N}_2]$ -dT	254.12	Unit	133.06	Unit	70	3	5	Positive
$[D_3]$ -dT	246.10	Unit	130.05	Unit	30	3	5	Positive

Supplementary Figures

MS-signal for *in vivo* deformylated fdC

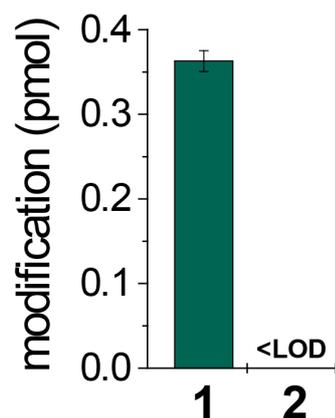


MS-signal for internal standard

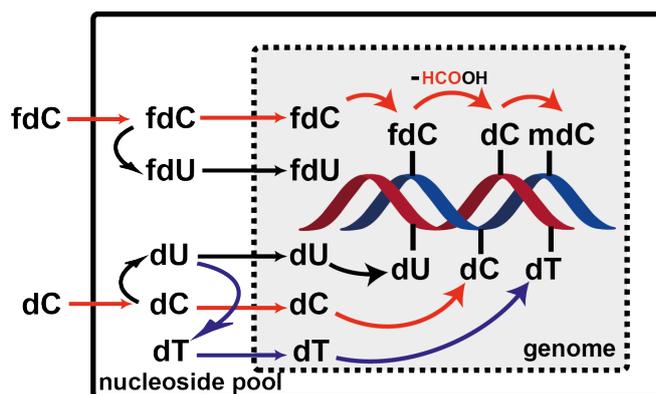


Supplementary Figure 1: Characteristic MS transition of [$^{13}\text{C}_5$][$^{15}\text{N}_2$]-dC and [$^{13}\text{C}_9$][$^{15}\text{N}_3$]-dC.

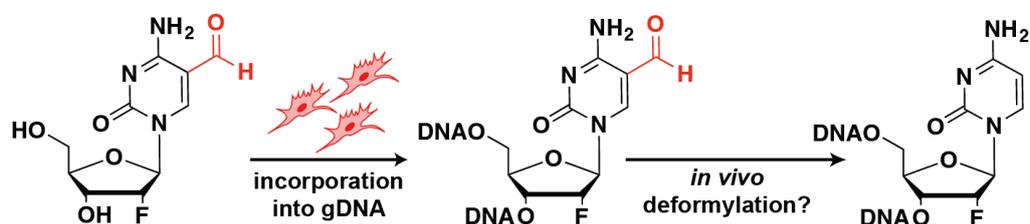
Fingerprint MS-fragmentation pathways and detected MS signals of **2** and **7** used to track isotopically labelled dC and its derivatives in mESC cultures.



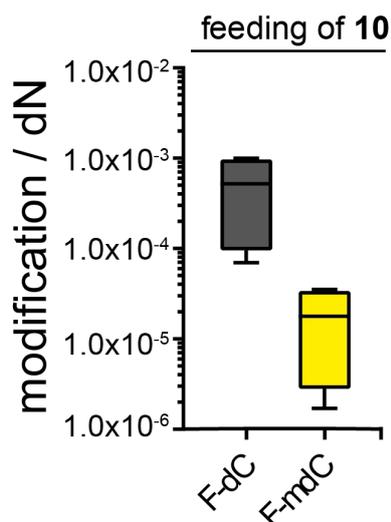
Supplementary Figure 2: Quantitative data of isotopically labelled pyrimidine derivatives in the soluble nucleoside pool of mESCs after feeding [$^{13}\text{C}_5$][$^{15}\text{N}_2$]-fdC shows its incorporation and no formation of [$^{13}\text{C}_5$][$^{15}\text{N}_2$]-dC. LC-MS/MS analysis of the soluble nucleoside pool from J1 mESCs cultured in the presence of 50 μM **1** under priming conditions for three days. Compound **1** is incorporated but no formation of **2** is detected in the soluble nucleoside pool. Mean values and s.d. of technical triplicates from a single culture are shown.



Supplementary Figure 3: Schematic overview of the possible metabolic conversions of fdC and dC.

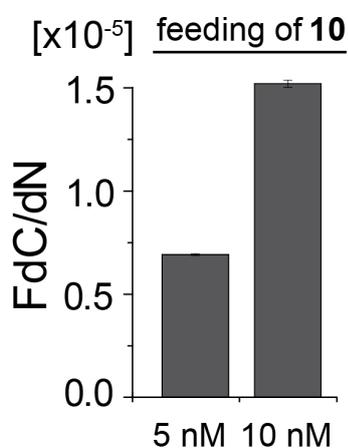


Supplementary Figure 4: Schematic overview of the feeding experiment with (*R*)-2'-F-fdC to mESCs. Feeding of **15** results in the incorporation of (*R*)-2'-F-fdC into genomic DNA and is converted *in vivo* to (*R*)-2'-F-dC by a C-C bond cleavage reaction.



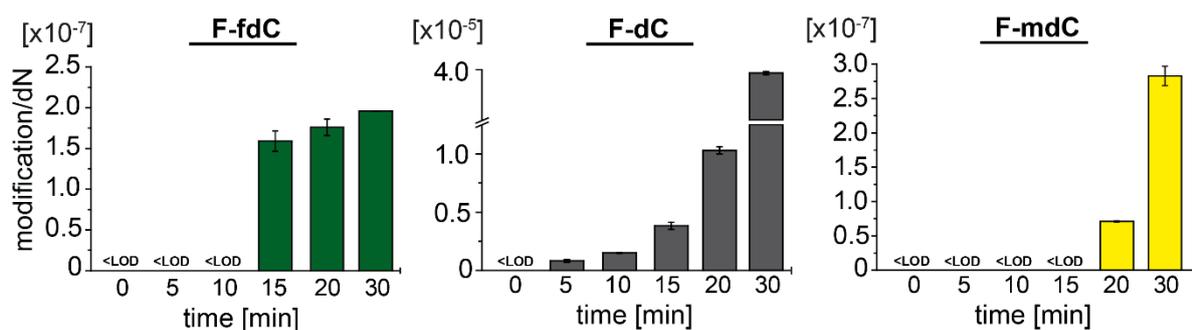
Supplementary Figure 5: Quantitative data of fluorinated pyrimidine derivatives after treatment with 10 show incorporation of 10 and 3% re-methylation.

LC-MS/MS analysis of genomic DNA from J1 mES cells cultured in the presence of $1.0 \mu\text{M}$ (*R*)-2'-F-dC under priming conditions for three days. DNA samples of four independent cultures were measured, each as technical triplicate. Mean values with s.d. are depicted.



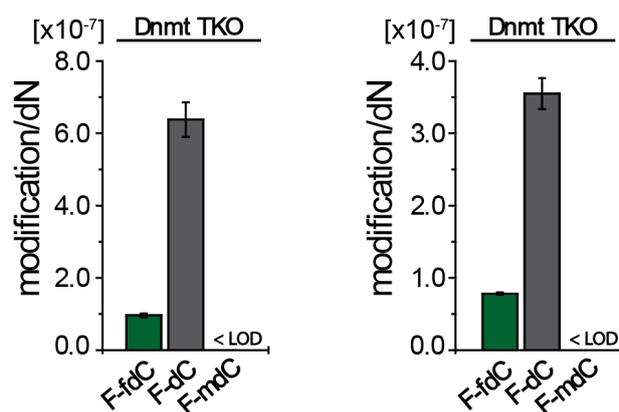
Supplementary Figure 6: Quantification data of incorporated (*R*)-2'-F-dC after feeding of 10 at two concentrations. Levels similar to those resulting from feeding of (*R*)-2'-F-dC were observed.

Analysis of genomic DNA from J1 mES cells cultured for three days under priming conditions in the presence of 5 and 10 nM (*R*)-2'-F-dC. Mean values and s.d. of technical triplicates from single cultures are shown.



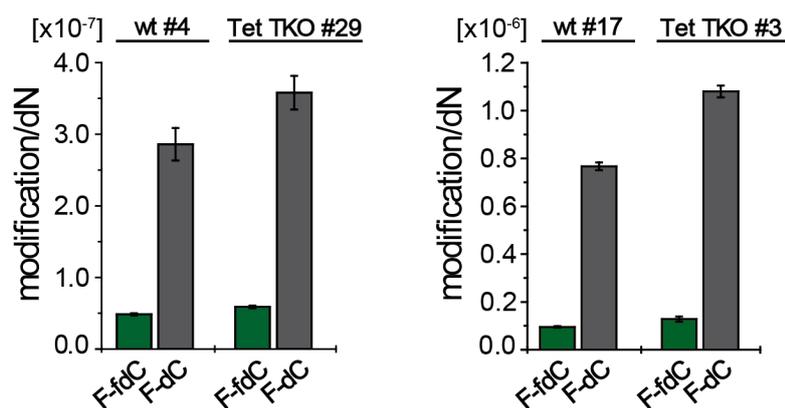
Supplementary Figure 7: Time dependent co-feeding with 15 and 10 show incorporation of 10 already after 5 min, 15 after 15 min and 11, resulting from both compounds, after 20 min.

LC-MS/MS analysis of the genomic DNA from J1 mESCs cultured under priming conditions and after 48 h supplemented with both (*R*)-2'-F-fdC (350 μ M) and (*R*)-2'-F-dC (1.0 μ M) for 0, 5, 10, 15, 20 and 30 min. Mean values and s.d. of technical triplicates from single cultures are shown.



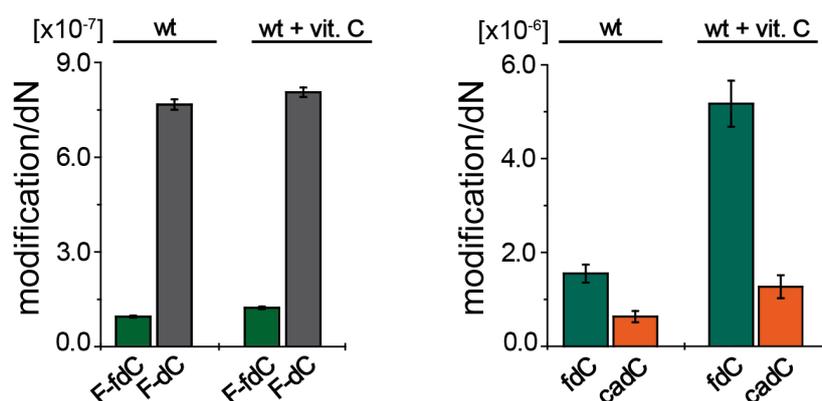
Supplementary Figure 8: Catalytically active DNA methyltransferases are required for methylation of (*R*)-2'-F-dC, but not for deformylation of (*R*)-2'-F-fdC (related to Fig. 4d).

LC-MS/MS analysis of genomic DNA from Dnmt TKO J1 mESCs cultured in the presence of (*R*)-2'-F-fdC under priming conditions. Shown are mean values and s.d. from two independent cultures, each measured as technical triplicates.



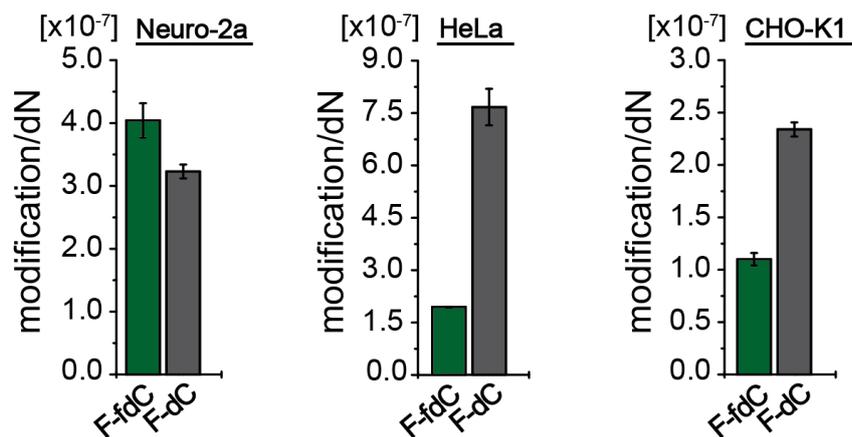
Supplementary Figure 9: Tet enzymes are not required for deformylation of (*R*)-2'-F-fdC (related to Fig. 4e).

LC-MS/MS analysis of genomic DNA from wt and Tet TKO mESCs cultured in the presence of (*R*)-2'-F-fdC under priming conditions. Left panel: cell lines described in (34). Right panel: cell lines reported in (33) (same cell lines as for experiment shown in Figure 4e). Shown are mean values and s.d. from two independent cell culture experiments, each measured as technical triplicates.



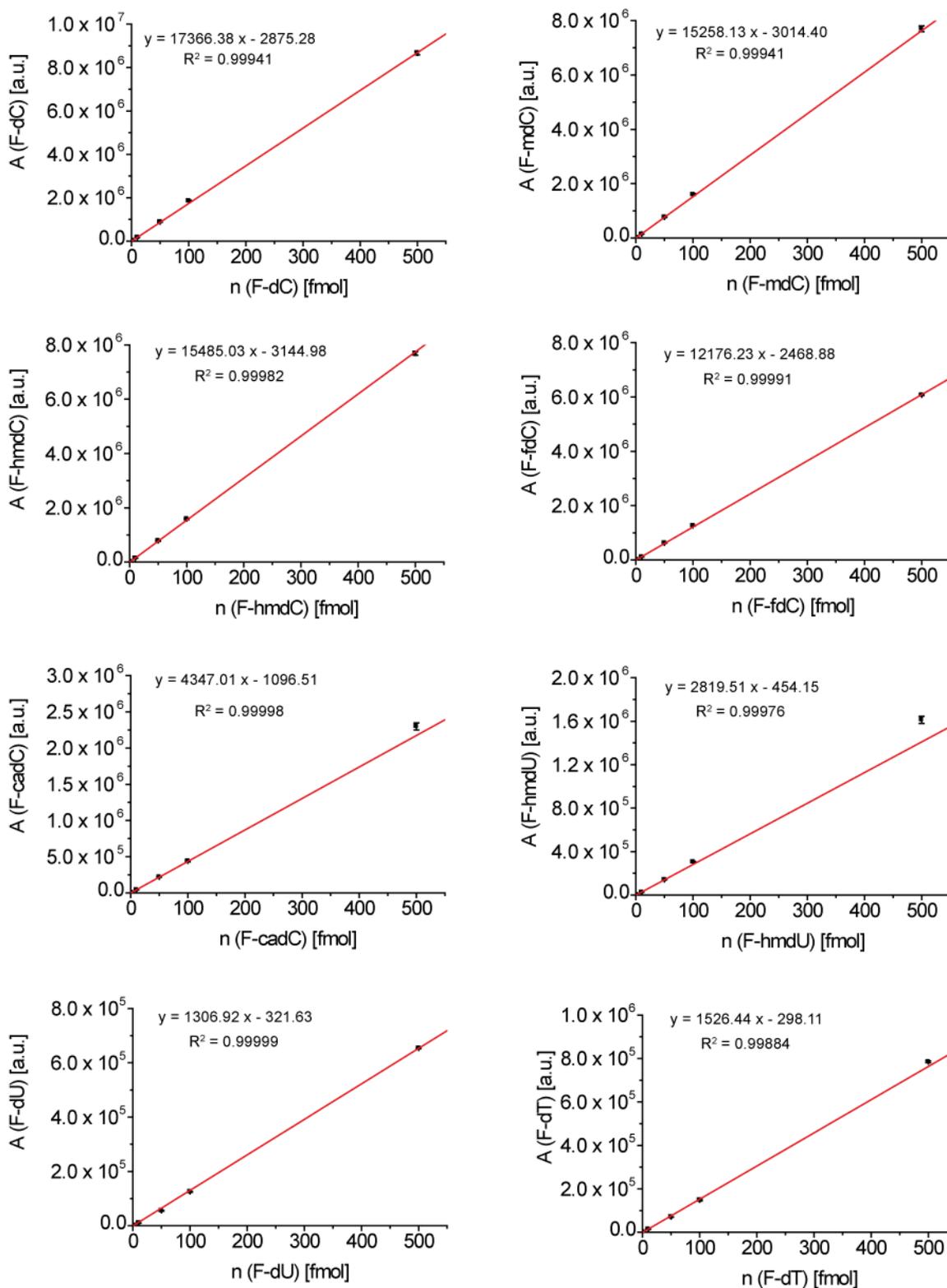
Supplementary Figure 10: Deformylation of (*R*)-2'-F-fdC is not affected by treatment with ascorbic acid-2-phosphate.

LC-MS/MS analysis of genomic DNA from wt mESCs cultured in the presence of (*R*)-2'-F-fdC under priming conditions with or without ascorbic acid-2-phosphate. The cell lines used in this experiment were reported in (33) and are the same as used for experiment shown in Figure 4e and Supplementary Figure 9. Shown are mean values and s.d. from two independent cell culture experiments, each measured as technical triplicates.

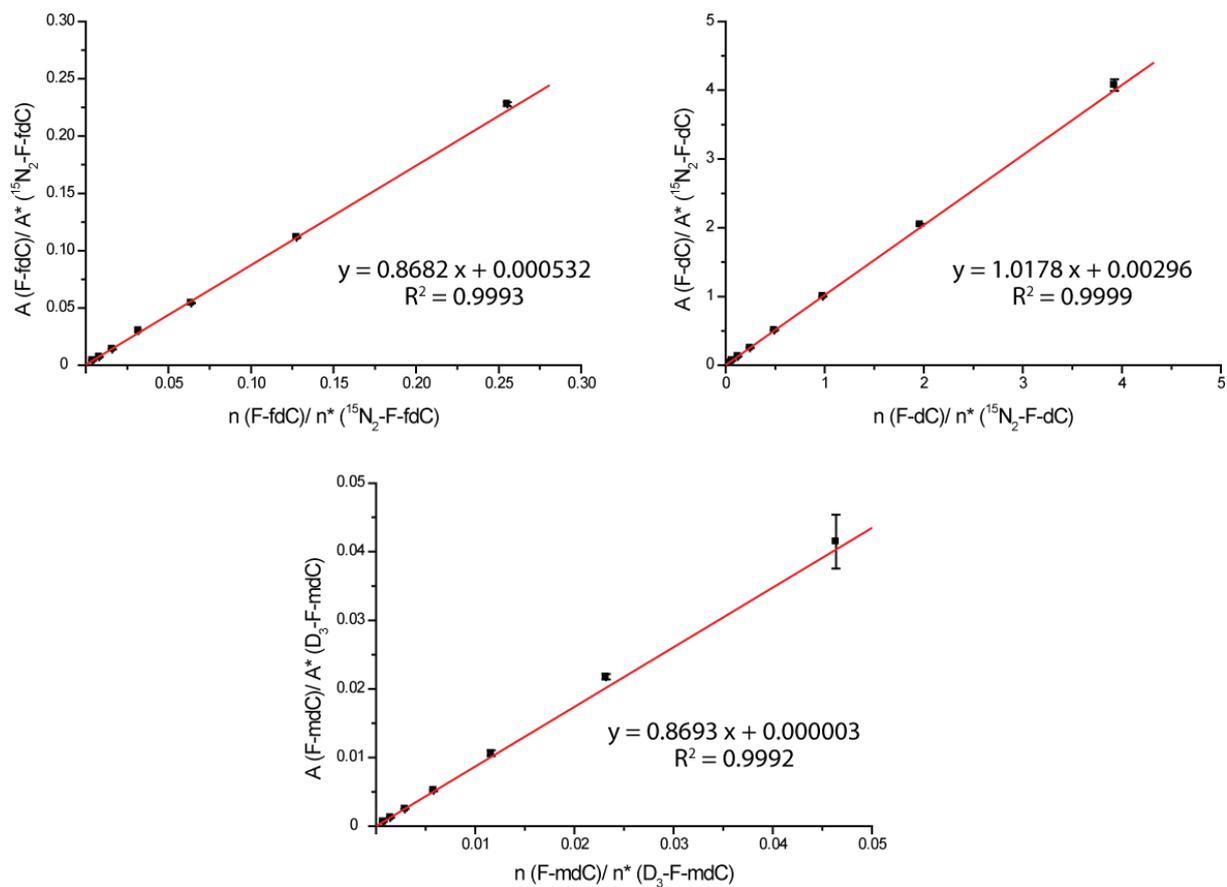


Supplementary Figure 11: Deformylation of (*R*)-2'-F-fdC to (*R*)-2'-F-dC occur also in non-pluripotent mouse and human cell lines (related to Fig. 4f).

LC-MS/MS analysis of genomic DNA isolated from Neuro-2a, HeLa and CHO-K1 cells cultured in the presence of (*R*)-2'-F-fdC. Albeit at variable levels, incorporation of (*R*)-2'-F-fdC into the genome as well as its conversion to (*R*)-2'-F-dC are detected in all cell lines. Shown are mean values and s.d. from three independent cell culture experiments, each measured as technical triplicates.



Supplementary Figure 12: Representative external calibration curves for all investigated nucleosides.



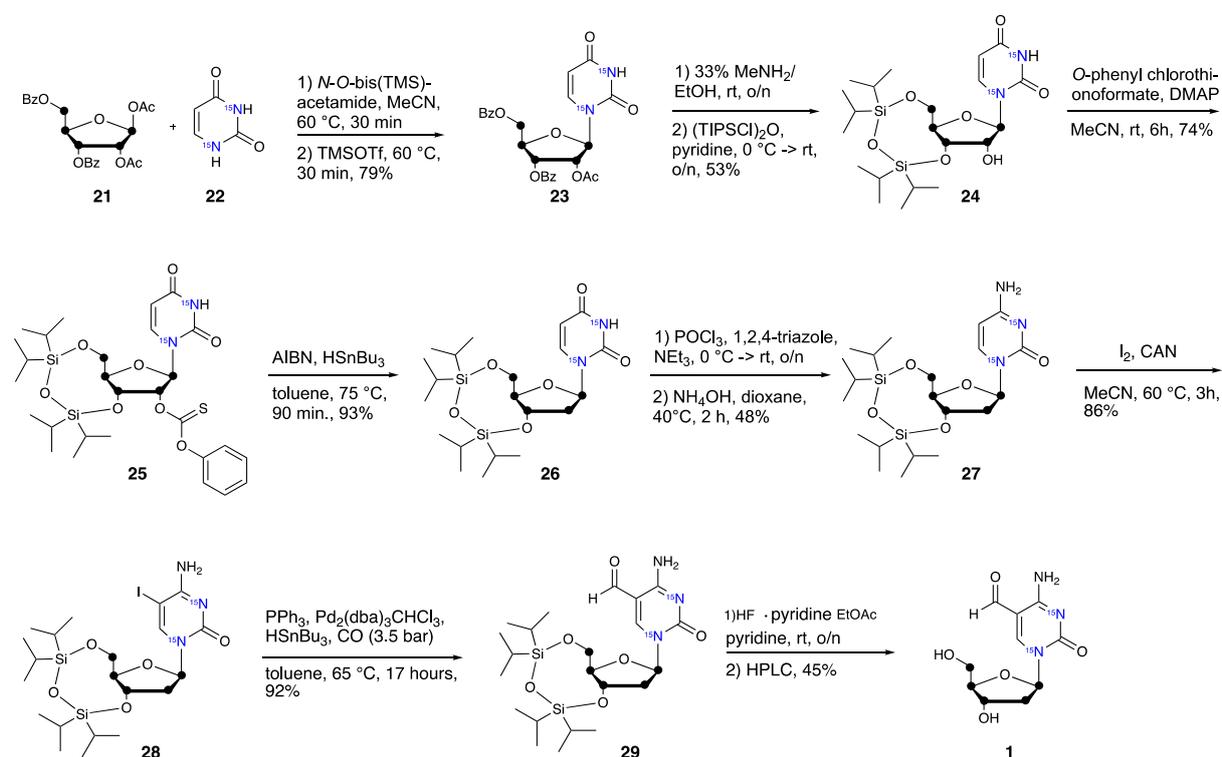
Supplementary Figure 13: Internal calibration curves for the exact quantification of (R)-2'-F-fdC, (R)-2'-F-dC and (R)-2'-F-mdC with the corresponding linear equation and coefficient of determination.

Synthetic procedures

General information

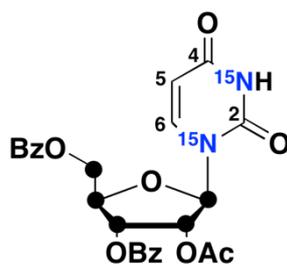
Isotopically labelled **7** was bought from *B.A.C.H. UG*, 2'-deoxy-2'-(*R*)-fluorocytidine **10** was purchased from *Carbosynth*. Compounds **8** and **9** resulted from the feeding experiment of **7**, compounds **11 - 14** resulted from the feeding of **10**. **15** and **30** were synthesized as described in the literature.^{1,2} Identity of compounds used in this study were confirmed by standard spectroscopic methods such as NMR and MS.

Synthesis of 2'-deoxy-5-formyl(1',2',3',4',5'-¹³C₅, ¹⁵N₂)cytidine (**1**)



Scheme 1: Synthesis of ¹³C₅-¹⁵N₂-labelled fdC **1**.

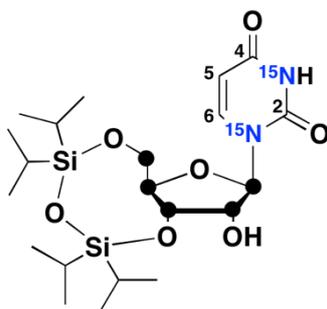
2'-O-acetyl-3',5'-di-O-benzoyl(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-uridine (23)



The ¹³C ribose fragment **21** and ¹⁵N₂ urea **22** were synthesized according to the literature.^{3,4} Bis-benzoyl and bis-acetyl protected ¹³C ribose **21** (2.00 g, 4.47 mmol, 1.0 eq), ¹⁵N₂-Uracil (510 mg, 4.47 mmol, 1.0 eq) and *N,O*-bis(trimethylsilyl)acetamide (3.28 mL, 13.4 mmol, 3.0 eq) were suspended in dry MeCN (25 mL) and heated to 60 °C whereas a clear solution developed after 30 minutes of stirring. Subsequently TMS triflate (2.83 mL, 15.7 mmol, 3.5 eq) was added at 60 °C and the mixture was stirred for an additional 30 minutes. After the reaction was cooled to room temperature, MeCN was evaporated *in vacuo*, the residue was taken up in sat. aqueous NaHCO₃ (75 mL) and extracted with DCM (3 x 75 mL). The combined organic layers were dried over Na₂SO₄ and the solvents were removed by rotary evaporation. Purification of the crude material through column chromatography (20% EtOAc/Hex → 50% EtOAc/Hex) yielded 1.75 g (3.49 mmol, 79%) of **23** as a white foam.

¹H-NMR (400 MHz, CDCl₃, ppm): δ = 8.08 - 8.05 (m, 4H, Bz-H), 7.67- 7.57 (m, 2H, Bz-H), 7.51 - 7.46 (m, 4H, Bz-H), 7.49 (td, ³J = 7.8 Hz, ¹J_{N-H} = 3.2 Hz, 1H, 5-H), 6.41 (d, ¹J_{C-H} = 169.1 Hz, 1H, 1'-H), 5.96 (d, ¹J_{C-H} = 150.3 Hz, 1H, 3'-H), 5.74 (d, ¹J_{C-H} = 153.0 Hz, 2'-H), 5.57 5.74 (m, 1H, 6-H), 4.97 (d, ¹J_{C-H} = 154.6 Hz, 1H, 4'-H), 4.82 (d, ¹J_{13C-H} = 150.1 Hz, 1H, 5'-H), 2.06 (s, 3H, Ac-CH₃). **¹³C-NMR (101 MHz, CDCl₃, ppm):** δ = 169.7 (Ac-C=O), 166.0 (Bz-C=O), 165.3 (Bz-C=O), 162.4 (d, ¹J_{C-N} = 9.5 Hz, C-4), 150.1 (t, ¹J_{C-N} = 19.4 Hz, C-2), 139.2 (d, ¹J_{C-N} = 12.5 Hz, C-5), 133.9 (2 x C_{Ar}), 133.8 (2 x C_{Ar}), 129.8 (2 x C_{Ar}), 129.6 (2 x C_{Ar}), 128.8 (2 x C_{Ar}), 128.7 (2 x C_{Ar}), 103.5 (d, ¹J_{C-N} = 6.9 Hz, C-6) 87.2 (ddd, J = 44.4 Hz, 13.7 Hz, 3.7 Hz, 1'-C), 80.4 (dd, J = 43.1 Hz, 38.3 Hz, 4'-C), 73.0 (dd, J = 44.3, 39.9, 2'-C), 70.9 (td, J = 39.9 Hz, 39.4 Hz, 3.7 Hz, 3'-C), 63.7 (d, J = 43.1 Hz), 20.5 (Ac-CH₃). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -224.4, -224.5. **HRMS (ESI⁺):** calc. for C₂₀¹³C₅H₂₆N¹⁵N₂O₉⁺ [M+NH₄]⁺: 519.1772, found: 519.1784. **HRMS (ESI⁻):** calc. for C₂₀¹³C₅H₂₁¹⁵N₂O₉ [M-H]⁻: 500.1361, found: 500.1359. **IR (ATR):** ν (cm⁻¹) = 1712 (s), 1688 (s), 1679 (s), 1631 (w), 1601 (w), 1584 (w), 1450 (m), 1369 (m), 1315 (w), 1249 (s), 1220 (s), 1176 (m), 1107 (s), 1067 (s), 1025 (s), 806 (m), 762 (w), 707 (s), 686 (m).

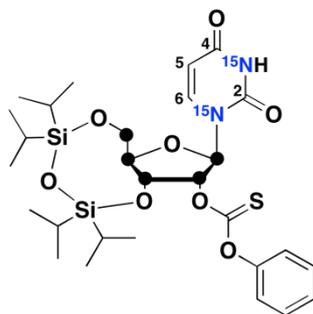
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-uridine (24)



First, **23** (1.75 g, 3.49 mmol, 1.0 eq) was deprotected by stirring the compound in a solution of MeNH₂ in ethanol (10 mL, 33%) overnight. The ethanolic solution was concentrated *in vacuo* and resuspended in water (25 mL). The aqueous phase was extracted with DCM (5 x 25 mL) until no UV absorption on a TLC plate was visible anymore. Subsequently, the water was removed by rotary evaporation and the deprotected nucleoside was dried on high vac. Then, pyridine (3.0 mL) was added and the light red, clear solution was cooled to 0 °C. 1,3-dichloro-1,1,3,3-tetraisopropylidisiloxane (1.34 μL, 4.19 mmol, 1.2 eq) was added dropwise and the mixture was then warmed to room temperature and stirred for 14 hours. MeOH (3.0 mL) was added and the volatiles were removed *in vacuo*. The crude product mixture was then charged with saturated solution of aqueous NH₄Cl (50 mL), extracted twice with EtOAc (50 mL) and the organic layers dried over Na₂SO₄. After removal of the solvent *in vacuo*, pure product was isolated after column chromatography (30% EtOAc/*i*Hex --> 40% EtOAc/*i*Hex), which yielded 920 mg (1.86 mmol, 53%) of **24** as an off white foam.

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 9.23 (d, ²J_{N-H} = 90.9 Hz, 1H, N-H), 7.72 (td, ³J = 8.1 Hz, ²J_{N-H} = 2.2 Hz, 1H, 5-H), 5.88 (d, ¹J_{C-H} = 174.6 Hz, 1H, 1'-H), 5.69 (ddd, ³J = 7.8 Hz, ³J_{N-H} = 4.6 Hz, ³J_{N-H} = 2.8 Hz, 1H, 6-H), 4.47 - 3.87 (m, 2'-H, 3'-H, 4'-H, 5'-H), 1.10 - 1.01 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 163.3 (d, ¹J_{C-N} = 9.5 Hz, C-4), 150.0 (t, ¹J_{C-N} = 18.4 Hz, 2-C), 139.9 (d, ¹J_{C-N} = 12.7 Hz, C-5), 105.4 (d, ²J_{C-N} = 34.7 Hz, C-6), 90.9 (dd, ¹J_{C-C} = 42.6 Hz, ¹J_{C-N} = 11.0 Hz, 1'-C), 81.9 (t, J = 42.3 Hz, 2'-C), 75.1 (t, ¹J_{C-C} = 40.6, 3'-C), 68.9 (t, J = 39.9 Hz, 4'-C), 60.2 (d, J = 39.9 Hz), 17.4 - 16.8 (Si-C(CH₃)), 13.4 - 12.8 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -224.4, -224.5. **HRMS (ESI⁺):** calc. for C₁₆¹³C₅H₃₉¹⁵N₂O₇Si₂⁺ [M+H]⁺: 494.2399, found: 494.2412. **HRMS (ESI⁻):** calc. for C₁₆¹³C₅H₃₇¹⁵N₂O₇Si₂⁻ [M-H]⁻: 492.2253, found: 492.2252. **IR (ATR):** ν (cm⁻¹) = 2944 (m), 2867 (m), 1688 (s), 1679 (s), 1650 (m), 1631 (w), 1462 (m), 1444 (m), 1383 (w), 1255 (m), 1198 (w), 1135 (m), 1096 (s), 1028 (s), 976 (s), 919 (w), 884 (s), 852 (m), 807 (m), 761 (m), 694 (s).

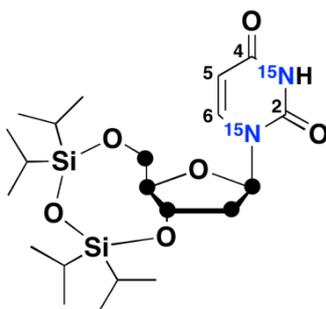
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-2'-O-phenoxythiocarbonyl-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-uridine (25)



Compound **25** was synthesized according to the literature with slight modifications.⁵ **24** (280 mg, 0.57 mmol, 1.0 eq) was dissolved in MeCN (8 mL) and *O*-phenyl chlorothionoformate (120 μ L, 0.86 mmol, 1.5 eq) and 4-(diethylamino)pyridine (202 mg, 1.65 mmol, 2.9 eq) were added at room temperature. After the yellowish mixture, which showed a colourless precipitate after 30 minutes, was stirred for a total of six hours, the reaction was aborted through addition of MeOH (10 mL) and the volatiles were removed under reduced pressure. Subsequently, a saturated aqueous NaHCO₃ solution (25 mL) was given to the crude reaction mixture and extracted with DCM (3 x 25 mL). Combined organic layers were then dried over Na₂SO₄ and the solvent was removed *in vacuo*. Purification through column chromatography (25% EtOAc/*i*Hex) yielded 262 mg (0.42 mmol, 74%) of pure product **25** as a colourless oil.

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 8.27 (d, ²*J*_{N-H} = 91.1 Hz, 1H, N-H), 7.71 (dt, ³*J* = 8.1 Hz, ²*J*_{N-H} = 2.2 Hz, 1H, 5-H), 7.42 (t, ³*J* = 8.0 Hz, 2H, H_{Ar}), 7.30 (t, ³*J* = 7.6 Hz, 1H, H_{Ar}), 7.12 (d, ³*J* = 8.3 Hz, 2H, H_{Ar}), 6.15 (td, ¹*J*_{C-H} = 164.2 Hz, ³*J* = 5.5 Hz, 1H, 2'-H), 6.07 (d, ¹*J*_{C-H} = 174.8 Hz, 1H, 1'-H), 5.72 (ddd, ³*J* = 7.7 Hz, ³*J*_{N-H} = 4.9 Hz, ³*J*_{N-H} = 2.5 Hz, 1H, 6-H), 4.55 (ddt, ¹*J*_{C-H} = 142.4 Hz, ³*J* = 8.4 Hz, 4.0 Hz, 1H, 4'-H), 4.45 - 3.90 (m, 3H, 3'-H, 5'-H), 1.12 - 1.03 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 193.7 ((O)₂C=S), 162.5 (d, ¹*J*_{C-N} = 9.5 Hz, C-4), 153.4 (C_{Ar}), 149.3 (t, ¹*J*_{C-N} = 18.5 Hz, 2-C), 139.4 (d, ¹*J*_{C-N} = 12.8 Hz, C-5), 129.6 (2 x C_{Ar}), 126.7 (C_{Ar}), 121.7 (2 x C_{Ar}), 102.3 (d, ²*J*_{C-N} = 7.5 Hz, C-6), 88.6 (dd, ¹*J*_{C-C} = 53.5 Hz, ¹*J*_{C-N} = 12.2 Hz, 1'-C), 83.7 (t, *J* = 41.9 Hz, 2'-C), 82.2 (t, ¹*J*_{C-C} = 42.3, 3'-C), 68.1 (t, *J* = 41.0 Hz, 4'-C), 59.4 (d, *J* = 43.1 Hz), 17.4 - 16.8 (Si-C(CH₃)), 13.4 - 12.4 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -223.5, -223.6. **HRMS (ESI⁺):** calc. for C₂₃¹³C₅H₄₃¹⁵N₂O₈SSi₂⁺ [M+H]⁺: 630.2382, found: 630.2393. **IR (ATR):** ν (cm⁻¹) = 2944 (w), 2866 (w), 2359 (w), 1722 (w), 1709 (m), 1691 (m), 1679 (m), 1649 (w), 1641 (w), 1631 (w), 1591 (w), 1462 (w), 1443 (w), 1382 (w), 1274 (m), 1202 (s), 1137 (w), 1098 (s), 1030 (s), 925 (m), 883 (m), 853 (m), 808 (m), 766 (m), 689 (s).

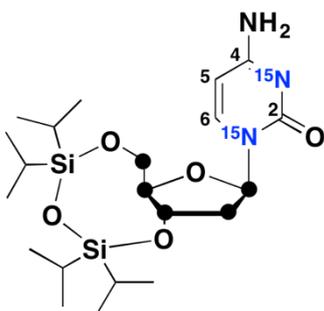
3',5'-(1,1,3,3-tetraisopropylidisiloxane-1,3-diyl)-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-uridine (26)



The deoxygenation of **25** was carried out as described in the literature with slight modifications.⁵ First, **25** (260 mg, 0.41 mmol, 1.0 eq), AIBN (14 mg, 0.08 mmol, 0.2 eq) and tributyltin hydride (165 μ L, 0.62 mmol, 1.5 eq) were dissolved in toluene (8 mL). The mixture was freeze-pump-thaw degassed (3 x) and then heated to 75 °C for 90 minutes. The solvent was removed *in vacuo* and the residue was taken up in water (25 mL). The aqueous layer was then extracted with DCM (3 x 25 mL) and the combined organic layers were dried over Na₂SO₄. Flash-chromatography of the crude (25% EtOAc/*i*Hex --> 30% EtOAc/*i*Hex) yielded **26** as a white foam (180 mg, 0.38 mmol, 93%).

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 8.43 (d, ²J_{N-H} = 91.1 Hz, 1H, N-H), 7.71 (dt, ³J = 8.1 Hz, ²J_{N-H} = 2.2 Hz, 1H, 5-H), 6.19 (dd, ¹J_{C-H} = 175 Hz, ³J = 6.5 Hz, 1H, 1'-H), 5.70 - 5.67 (m, 1H, 6-H), 4.57 (d, ¹J_{C-H} = 142.1 Hz, 3'-H), 4.27 - 3.64 (m, 3H, 4'-H, 5'-H), 2.63 - 2.14 (m, 2H, 2'H), 1.10 - 1.01 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 162.9 (d, ¹J_{C-N} = 9.4 Hz, C-4), 149.8 (t, ¹J_{C-N} = 18.5 Hz, 2-C), 139.6 (d, ¹J_{C-N} = 12.9 Hz, C-5), 101.6 (d, ²J_{C-N} = 7.0 Hz, C-6), 85.1 (t, ¹J_{C-C} = 41.9 Hz, 4'-C), 84.3 (dd, ¹J_{C-C} = 36.4 Hz, ¹J_{C-N} = 10.6 Hz, 1'-C), 67.0 (t, ¹J_{C-C} = 39.2 Hz, 3'-C), 60.0 (d, *J* = 37.0 Hz, 5'-C), 39.9 (t, ¹J_{C-C} = 37 Hz, 2'-C), 17.5 - 16.7 (Si-C(CH₃)), 13.4 - 12.4 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -224.1, -224.2. **HRMS (ESI⁺):** calc. for C₁₆¹³C₅H₃₉¹⁵N₂O₆Si₂⁺ [M+H]⁺: 478.2450, found: 478.2451. **HRMS (ESI⁻):** calc. for C₁₆¹³C₅H₃₇¹⁵N₂O₆Si₂⁻ [M-H]⁻: 476.2304, found: 476.2306. **IR (ATR):** ν (cm⁻¹) = 2944 (w), 2866 (w), 1680 (s), 1446 (w), 1384 (w), 1246 (m), 1094 (m), 1065 (m), 1029 (s), 967 (m), 919 (w), 883 (s), 853 (m), 806 (m), 761 (m), 693 (s).

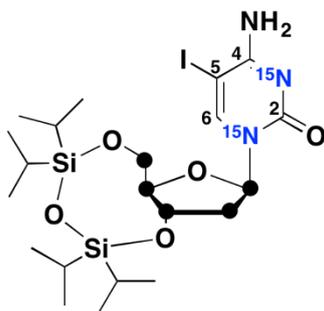
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-cytidine (27)



For the amination reaction, triazole (398 mg, 5.76 mmol, 9.0 eq) was dissolved in MeCN (12 mL) and cooled to 0°C. Then, POCl₃ (120 μL, 1.27 mmol, 2.0 eq) was added dropwise and the milky suspension was kept at 0 °C for ten minutes. Triethyl amine (753 μL, 5.50 mmol, 8.6 eq) was subsequently given to the slurry and stirred for an additional 20 minutes. Finally, **26** (304 mg, 0.64 mmol, 1.0 eq) was added and the reaction was allowed to warm to room temperature overnight. After TLC showed complete consumption of the starting material, the crude was suspended in DCM (25 mL) and washed with brine (25 mL). The organic phase was dried over Na₂SO₄ and the volatiles were removed *in vacuo*. This residue was then diluted with dioxane (15 mL), charged with NH₄OH (4 mL, 25%) and stirred at 40 °C for three hours. The mixture was subsequently neutralized with saturated aqueous NH₄Cl (50 mL) and extracted with EtOAc (2 x 25 mL). Column chromatography (5% MeOH/DCM) yielded 147 mg (0.31 mmol, 48%) of **27** as a colourless glass.

¹H-NMR (800 MHz, CDCl₃, ppm): δ = 7.90 (ddd, ³J = 7.3 Hz, ²J_{N-H} = 2.7 Hz, 1.1 Hz, 1H, 5-H), 6.17 (d, ¹J_{C-H} = 174.7 Hz, ³J = 7.7 Hz, 1H, 1'-H), 5.64 (dd, ³J = 7.4 Hz, ²J_{N-H} = 3.4 Hz, 6-H), 4.46 (d, ¹J_{C-H} = 137.8 Hz, 1H, 3'-H), 4.25 (d, ¹J_{C-H} = 149.2 Hz, 1H, 4'-H), 4.10 (d, ¹J_{C-H} = 139.7 Hz, 1H, 5'-H), 3.86 (d, ¹J_{C-H} = 143.6 Hz, 1H, 5'-H), 2.61 (d, ¹J_{C-H} = 134.2 Hz, 1H, 2'-H), 2.61 (d, ¹J_{C-H} = 134.2 Hz, 1H, 2'-H), 2.41 (d, ¹J_{C-H} = 143.2 Hz, 1H, 2'-H), 1.10 - 0.98 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (201 MHz, CDCl₃, ppm):** δ = 164.6 (d, ¹J_{C-N} = 6.2 Hz, C-4), 154.6 (t, ¹J_{C-N} = 12.2 Hz, 2-C), 140.3 (d, ¹J_{C-N} = 12.8 Hz, C-5), 92.0 (d, ²J_{C-N} = 7.0 Hz, C-6), 84.1 (d, ¹J_{C-C} = 40.4 Hz, 1'-C), 83.9 (t, ¹J_{C-C} = 42.8 Hz, 4'-C), 65.7 (t, ¹J_{C-C} = 36.9, 3'-C), 59.9 (d, J = 43.0 Hz, 5'-C), 38.9 (t, ¹J_{C-C} = 36.9 Hz, 2'-C), 16.5 - 15.8 (Si-C(CH₃)), 12.4 - 11.4 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -217.7, -217.9. **HRMS (ESI⁺):** calc. for C₁₆¹³C₅H₄₀N¹⁵N₂O₅Si₂⁺ [M+H]⁺: 477.2609, found: 477.2610. **HRMS (ESI⁻):** calc. for C₁₆¹³C₅H₃₈N¹⁵N₂O₅Si₂⁻ [M-H]⁻: 475.2464, found: 475.2464. **IR (ATR):** ν (cm⁻¹) = 3337 (w), 2944 (w), 2867 (w), 1622 (m), 1465 (m), 1394 (w), 1267 (w), 1180 (w), 1120 (m), 1095 (s), 1030 (s), 937 (w), 883 (s), 782 (m), 694 (s).

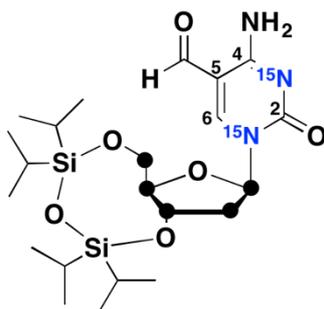
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-5-iodo-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-cytidine (28**)**



Ceric ammonium nitrate (320 mg, 0.58 mmol, 2.0 eq), iodine (147 mg, 0.58 mmol, 2.0 eq) and **27** (139 mg, 0.29 mmol, 1.0 eq) were dissolved in MeCN (20 mL) and heated to 60 °C for two and a half hours. The dark brownish solution was then poured into a mixture of brine (25 mL) and saturated aqueous Na₂S₂O₃ solution (5 mL). After extraction with DCM (3 x 25 mL), the organic phases were dried over Na₂SO₄ and then removed *in vacuo*. Purification via column chromatography (4% MeOH/DCM) yielded **28** as an off white powder (149 mg, 0.25 mmol, 86%).

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 8.07 (sd, ²J_{N-H} = 2.4 Hz, 1H, 5-H), 5.98 (dd, ¹J_{C-H} = 175.8 Hz, ³J = 6.8 Hz, 1H, 1'-H), 4.49 (d, ¹J_{C-H} = 126.2 Hz, 1H, 3'-H), 4.29 (d, ¹J_{C-H} = 146.0 Hz, 1H, 4'-H), 4.13 (d, ¹J_{C-H} = 137.9 Hz, 1H, 5'-H), 3.86 (d, ¹J_{C-H} = 136.7 Hz, 1H, 5'-H), 2.63 (d, ¹J_{C-H} = 133.4 Hz, 1H, 2'-H), 2.44 (d, ¹J_{C-H} = 136.0 Hz, 1H, 2'-H), 1.11 - 0.99 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 163.7 (d, ¹J_{C-N} = 5.2 Hz, C-4), 154.6 (t, ¹J_{C-N} = 10.5 Hz, 2-C), 146.5 (d, ¹J_{C-N} = 12.6 Hz, C-5), 85.7 - 84.9 (1'-C + 4'-C), 66.5 (t, ¹J_{C-C} = 39.2, 3'-C), 59.7 (d, J = 43.2 Hz, 5'-C), 55.3 (C-6), 39.9 (t, ¹J_{C-C} = 37.0 Hz, 2'-C), 17.8 - 16.9 (8 x Si-C(CH₃)), 13.5 - 12.3 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -209.7, -209.9. **HRMS (ESI⁺):** calc. for C₁₆¹³C₅H₃₉N¹⁵N₂O₅Si₂⁺ [M+H]⁺: 603.1576, found: 603.1581. **HRMS (ESI⁻):** calc. for C₁₆¹³C₅H₃₇N¹⁵N₂O₅Si₂⁻ [M-H]⁻: 601.1430, found: 601.1427. **IR (ATR):** ν (cm⁻¹) = 3435 (w), 3328 (w), 2944 (w), 2867 (w), 1631 (m), 1600 (m), 1570 (m), 1463 (m), 1386 (w), 1262 (w), 1235 (w), 1155 (w), 1116 (w), 1094 (s), 1044 (m), 1023 (s), 696 (m), 937 (w), 919 (w), 884 (m), 871 (m), 852 (w), 787 (w), 775 (m), 697 (m).

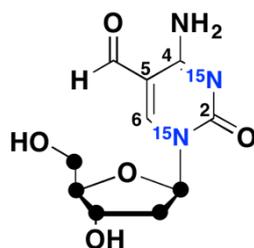
3',5'-(1,1,3,3,-tetraisopropylidisiloxane-1,3-diyl)-5-formyl-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-cytidine (29)



A high pressure autoclave glass vessel was charged with PPh₃ (13 mg, 0.05 mmol, 0.6), Pd₂(dba)₃CHCl₃ (8 mg, 0.01 mmol, 0.1 eq) and **28** (48 mg, 0.08 mmol, 1.0 eq) and suspended in toluene (5 mL). The vessel was flushed twice with CO and then set to a final CO pressure of 50 psi. The mixture was then heated to 65 °C and tributyltin hydride (27 μL diluted with 300 μL toluene) was added through a syringe pump with an addition speed of 20 μL per hour. After the addition was finished, the volatiles were removed *in vacuo* and the yellowish crude mixture was purified by flash column chromatography (20% EtOAc/*i*Hex --> 40% EtOAc/*i*Hex --> 50% EtOAc/*i*Hex), whereas 37 mg (0.073 mmol, 92%) of product **29** could be obtained as a light yellowish oil.

¹H-NMR (400 MHz, CDCl₃, ppm): δ = 9.50 (s, 1H, CHO), 8.53 (sd, ²J_{N-H} = 2.7 Hz, 1H, 5-H), 6.01 (dd, ¹J_{C-H} = 177.6 Hz, ³J = 6.7 Hz, 1H, 1'-H), 4.54 - 3.63 (m, 4H, 3'-H, 4'-H, 5'-H), 2.78 (d, ¹J_{C-H} = 135.1 Hz, 1H, 2'-H), 2.44 (d, ¹J_{C-H} = 137.5 Hz, 1H, 2'-H), 1.10 - 0.96 (m, 28H, Si(CH₃)₂, Si-CH). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 187.5 (CHO), 162.9 (d, ¹J_{C-N} = 5.9 Hz, C-4), 153.3 (t, ¹J_{C-N} = 7.9 Hz, 2-C), 152.7 (d, ¹J_{C-N} = 14.3 Hz, C-5), 104.8 (C-6), 86.1 - 84.9 (1'-C + 4'-C), 65.9 (t, ¹J_{C-C} = 38.2, 3'-C), 59.4 (d, J = 42.9 Hz, 5'-C), 39.5 (t, ¹J_{C-C} = 36.9 Hz, 2'-C), 17.5 - 16.8 (8 x Si-C(CH₃)), 13.5 - 12.4 (4 x Si-C(CH₃)). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -206.6, -206.8. **HRMS (ESI⁺):** calc. for C₁₇¹³C₅H₄₀N¹⁵N₂O₆Si₂⁺ [M+H]⁺: 505.2559, found: 505.2567. **IR (ATR):** ν (cm⁻¹) = 3393 (w), 2944 (m), 2866 (m), 2361 (w), 1658 (s), 1498 (m), 1463 (m), 1437 (m), 1412 (w), 1387 (w), 1308 (w), 1228 (m), 1183 (m), 1160 (m), 1118 (s), 1092 (s), 1064 (s), 1030 (s), 970 (m), 939 (m), 919 (w), 883 (m), 778 (m), 749 (m), 720 (s), 693 (s).

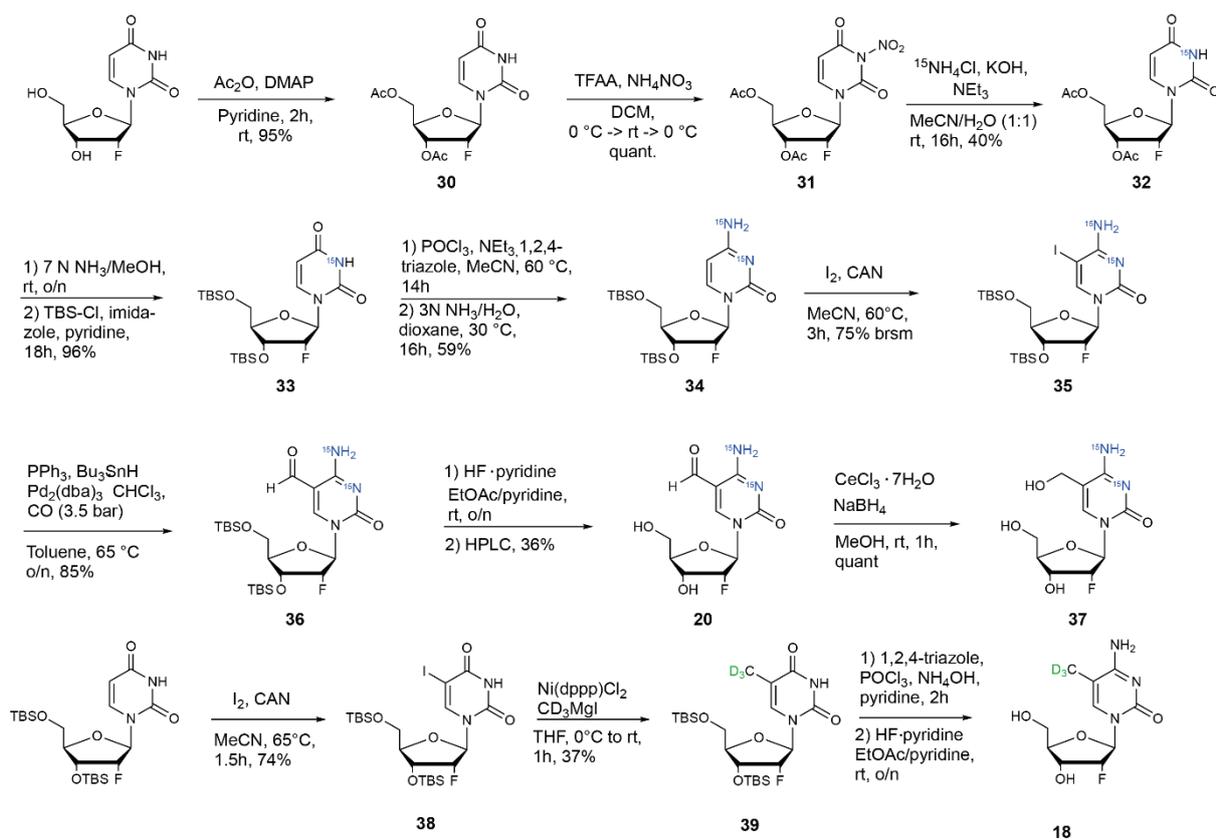
5-formyl-2'-deoxy-(1',2',3',4',5'-¹³C₅, N¹,N³-¹⁵N₂)-cytidine (1)



29 (30 mg, 0.06 mmol, 1.0 eq) was dissolved in EtOAc (1 mL) and pyridine (24 μ L) and HF•pyridine (16 μ L, 0.6 mmol, 10 eq) were added. After the mixture was stirred for 12 hours at room temperature, a white precipitate formed. Excess of HF was quenched with TMSOMe (60 μ L) and the colourless slurry was stirred for an additional 30 minutes. Then, the mixture was centrifuged (6000 rpm à 10 mins.), the supernatant removed and the colourless powder washed with EtOAc (5 mL) for three consecutive times. HPLC purification (0% --> 25% MeCN in 30 minutes) yielded **1** (7 mg, 0.026 mmol, 45%) as a colourless solid.

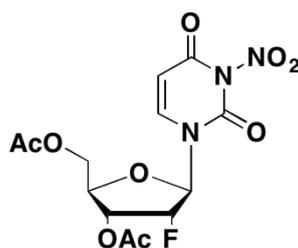
¹H-NMR (400 MHz, D₂O, ppm): δ = 9.50 (s, 1H, CHO), 8.84 (sd, ²J_{N-H} = 2.9 Hz, 1H, 5-H), 6.15 (dt, ¹J_{C-H} = 175.4 Hz, ³J = 5.8 Hz, 1H, 1'-H), 4.61 (d, ¹J_{C-H} = 151.9 Hz, 1H, 3'-H), 4.30 - 3.93 (m, 2H, 4'-H + 5'-H), 3.76 - 3.53 (m, 1H, 5'-H), 2.74 (d, ¹J_{C-H} = 136.0 Hz, 1H, 2'-H), 2.53 (d, ¹J_{C-H} = 134.5 Hz, 1H, 2'-H). **¹³C-NMR (151 MHz, D₂O, ppm):** δ = 190.5 (CHO), 161.7 (d, ¹J_{C-N} = 7.4 Hz, C-4), 154.7 (d, ¹J_{C-N} = 14.2 Hz, C-5), 154.0 (t, ¹J_{C-N} = 11.3 Hz, 2-C), 105.4 (C-6), 87.8 - 86.5 (1'-C + 4'-C), 69.3 (t, ¹J_{C-C} = 36.5, 3'-C), 60.5 (d, J = 41.5 Hz, 5'-C), 40.0 (t, ¹J_{C-C} = 36.0 Hz, 2'-C). **¹⁵N-NMR (41 MHz, D₂O, ppm):** δ = -209.9, -210.2. **HRMS (ESI⁺):** calc. for C₅¹³C₅H₁₃N¹⁵N₂NaO₅⁺ [M+Na]⁺: 285.0856, found: 285.0857.

Synthesis of isotopically labelled 2'-(*R*)-fluorocytidine derivatives



Scheme 2: Synthesis of $^{15}\text{N}_2$ -labelled 2'-(*R*)-F-fdC **20**, 2'-(*R*)-F-hmdC **37** and $[\text{D}_3]$ -2'-(*R*)-F-mdC **18**.

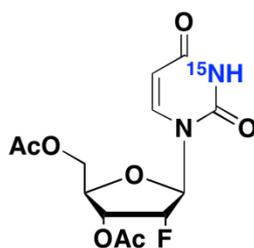
3',5'-di-O-acetyl-2'-deoxy-2'-(R)-fluoro-3-nitrouridine (31)



NH_4NO_3 (1.45 g, 18.1 mmol, 2.0 eq) was cautiously powdered with a mortar and pestle and subsequently suspended in DCM (50 mL). The resulting mixture was cooled to 0 °C and trifluoroacetic anhydride (5.10 g, 36.2 mmol, 4.0 eq) was added dropwise. After 30 minutes of warming up to room temperature, a brownish slurry developed which was stirred for another two hours until almost all of the remaining NH_4NO_3 was dissolved (note that complete dissolution of NH_4NO_3 could not be observed even after extended stirring). The resulting brownish-yellow solution was then again cooled to 0 °C and **30** (3.00 g, 9.06 mmol, 1.0 eq) was added in small portions. The reaction mixture was stirred for additional three hours at 0 °C and was quenched through the addition of PBS-Buffer (100 mL, pH = 7.4). The aqueous layer was extracted with DCM (3 x 75 mL), combined organic layers were dried over Na_2SO_4 and the solvent was removed *in vacuo*. The crude residue was purified by column chromatography (1% MeOH/DCM to 2.5% MeOH/DCM) and **31** (3.60 g, 9.59 mmol, quant.) was yielded as a colourless powder.

$^1\text{H-NMR}$ (599 MHz, CDCl_3 , ppm): δ = 7.47 (d, 3J = 8.3 Hz, 1H, 6-H), 5.89 (d, 3J = 8.3 Hz, 1H, 5-H), 5.81 (dd, $^3J_{\text{H-F}}$ = 17.2 Hz, 3J = 1.4 Hz, 1H, 1'-H), 5.41 (ddd, $^2J_{\text{H-F}}$ = 51.6 Hz, 3J = 4.8 Hz, 3J = 1.4 Hz, 1H, 2'-H), 5.13 (ddd, $^3J_{\text{H-F}}$ = 18.0 Hz, 3J = 4.8 Hz, 3J = 3.5 Hz, 1H, 3'-H), 4.47 - 4.43 (m, 2H, 5'-H + 4'-H), 4.33 - 4.30 (m, 1H, 5'H), 2.16 (s, 3H, C3'-O-(C=O)- CH_3), 2.10 (s, 3H, C5'-O-(C=O)- CH_3). **$^{13}\text{C-NMR}$ (151 MHz, CDCl_3 , ppm):** δ = 170.2 (C5'-O-(C=O)- CH_3), 169.9 (C3'-O-(C=O)- CH_3), 154.9 (C-4), 145.1 (C-2), 140.1 (C-6), 102.0 (C-5), 92.0 (d, $^2J_{\text{C-F}}$ = 37.3 Hz, C-1'), 90.8 (d, $^1J_{\text{C-F}}$ = 193.1 Hz, C-2'), 79.3 (C-4'), 69.3 (d, $^2J_{\text{C-F}}$ = 16.0 Hz, C-3'), 61.8 (C-5'), 20.8 (C5'-O-(C=O)- CH_3), 20.5 (C3'-O-(C=O)- CH_3). **$^{19}\text{F-NMR}$ (377 MHz, CDCl_3 , ppm):** δ = -199.4 (dt, $^2J_{\text{F-H}}$ = 51.8 Hz, $^3J_{\text{F-H}}$ = 18.4 Hz). **HRMS (ESI⁺):** calc. for $\text{C}_{13}\text{H}_{18}\text{FN}_4\text{O}_9^+$ [$\text{M}+\text{NH}_4$]⁺: 393.1052, found: 393.1048. **IR (ATR):** ν (cm^{-1}) = 3096 (w), 1737 (s), 1693 (s), 1644 (s), 1617 (m), 1558 (w), 1540 (w), 1506 (w), 1456 (w), 1436 (m), 1377 (m), 1289 (m), 1252 (m), 1224 (s), 1119 (m), 1100 (s), 1055 (s), 990 (m), 960 (m), 901 (m), 884 (m), 852 (w), 828 (m), 804 (m), 748 (m), 738 (m), 705 (w), 677 (w). **Melting Range:** 132 °C - 134 °C.

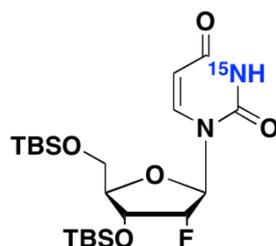
3',5'-di-O-acetyl-2'-deoxy-2'-(R)-fluoro-(N³-¹⁵N)-uridine (32)



First, ¹⁵NH₄Cl (400 mg, 7.34 mmol, 1.2 eq) was dissolved in H₂O (15 mL) and KOH (412 mg, 7.34 mmol, 1.2 eq) and NEt₃ (1.50 mL, 10.0 mmol, 1.6 eq) were added. **31** (2.30 g, 6.13 mmol, 1.0 eq) was dissolved in MeCN (15 mL) and also given to the clear solution, which was then vigorously stirred at room temperature overnight. Subsequently, MeCN was removed *in vacuo* and the resulting aqueous phase was neutralized through the addition of sat. NH₄Cl (100 mL). The aqueous layers were extracted with DCM (3 x 100 mL), organic layers were collected and dried over Na₂SO₄. The volatiles were removed under reduced pressure and the crude mixture was purified by column chromatography (DCM --> 2% MeOH/DCM). **32** (818 mg, 2.47 mmol, 40%) was yielded as a white foam.

¹H-NMR (800 MHz, DMSO-d₆, ppm): δ = 7.71 (d, ³J = 8.1 Hz, 1H, 6-H), 5.87 (dd, ³J_{H-F} = 22.4 Hz, ³J = 2.0 Hz, 1H, 1'-H), 5.69 - 5.68 (m, 1H, 5-H), 5.33 (ddd, ²J_{H-F} = 52.5 Hz, ³J = 5.3 Hz, ³J = 2.0 Hz, 1H, 2'-H), 5.26 (ddd, ³J_{H-F} = 25.2 Hz, ³J = 5.3 Hz, ³J = 2.8 Hz, 1H, 3'-H), 4.34 (dd, ²J = 12.4 Hz, ³J = 3.0 Hz, 1H, 5'-H), 4.27 (m, 1H, 4'-H), 4.16 (dd, ²J = 12.3 Hz, ³J = 5.8 Hz, 1H, 5'-H), 2.11 (s, 3H, C3'-O-(C=O)-CH₃), 2.04 (s, 3H, C5'-O-(C=O)-CH₃). **¹³C-NMR (202 MHz, DMSO-d₆, ppm):** δ = 170.2 (C5'-O-(C=O)-CH₃), 169.5 (C3'-O-(C=O)-CH₃), 163.2 (sd, ¹J_{C-N} = 8.7 Hz, C-4), 150.1 (sd, ¹J_{C-N} = 17.6 Hz, C-2), 142.9 (C-6), 102.1 (C-5), 90.8 (d, ¹J_{C-F} = 186.3 Hz, C-2'), 90.5 (d, ²J_{C-F} = 37.0 Hz, C-1'), 77.9 (C-4'), 69.8 (d, ²J_{C-F} = 14.3 Hz, C-3'), 62.7 (C-5'), 20.5 (C5'-O-(C=O)-CH₃), 20.3 (C3'-O-(C=O)-CH₃). **¹⁹F-NMR (377 MHz, DMSO-d₆, ppm):** δ = -198.0 (ddd, ²J_{H-F} = 52.5 Hz, ³J_{H-F} = 22.3, ³J_{H-F} = 17.2). **¹⁵N-NMR (41 MHz, DMSO-d₆, ppm):** δ = -222.0. **HRMS (ESI⁺):** calc. for C₁₃H₁₉FN₂¹⁵NO₇⁺ [M+NH₄]⁺: 349.1172, found: 349.1174. **HRMS (ESI⁻):** calc. for C₁₃H₁₄FN¹⁵NO₇⁻ [M-H]⁻: 330.0761, found: 330.0763. **IR (ATR):** ν (cm⁻¹) = 3092 (w), 3002 (w), 2929 (w), 2878 (w), 2817 (w), 1737 (s), 1716 (s), 1672 (s), 1464 (w), 1449 (w), 1429 (w), 1417 (m), 1380 (m), 1365 (m), 1335 (w), 1299 (w), 1266 (m), 1236 (s), 1120 (w), 1095 (m), 1073 (s), 1048 (s), 995 (m), 968 (m), 900 (s), 875 (m), 865 (m), 825 (s), 763 (m), 751 (m), 732 (w), 699 (w), 684 (w). **Melting Range:** 169 °C - 173 °C.

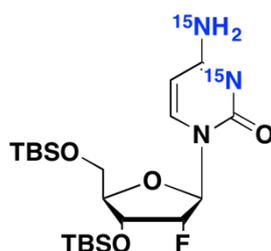
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)-2'-deoxy-2'-(*R*)-fluoro-(*N*³-¹⁵N)-uridine (33**)**



For the deprotection of **32** (800 mg, 2.42 mmol, 1.0 eq), the compound was dissolved in methanolic ammonia (7N, 5 mL) and stirred at room temperature overnight. The volatiles were then removed *in vacuo* and the mixture was dried on high vac. The resulting crude yellow oil was dissolved in pyridine (10 mL), imidazole (741 mg, 10.9 mmol, 4.5 eq) and TBSCl (1.09 g, 7.26 mmol, 3.0 eq) were added and the mixture was stirred at room temperature for 16 hours. After pyridine was removed by rotary evaporation, the residue was suspended in sat. NaHCO₃ (100 mL) and extracted with DCM (3 x 75 mL) and the combined organic layers were dried over Na₂SO₄. The obtained oily compound was then coevaporated with toluene (3 x 25 mL), whereas 1.11 g (2.33 mmol, 96%) of **33** were yielded as a colourless waxy solid.

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 7.92 (d, ³*J* = 8.2 Hz, 1H, 6-H), 6.06 (dd, ³*J*_{H-F} = 15.3 Hz, ³*J* = 2.0 Hz, 1H, 1'-H), 5.69 (dd, ³*J* = 8.2 Hz, ³*J*_{H-N} = 2.6 Hz, 1H, 5-H), 4.77 (ddd, ²*J*_{H-F} = 52.3 Hz, ³*J* = 4.3 Hz, ³*J* = 2.0 Hz, 1H, 2'-H), 4.29 (ddd, ³*J*_{H-F} = 18.9 Hz, ³*J* = 7.2 Hz, ³*J* = 4.2 Hz, 1H, 3'-H), 4.08 (dd, ³*J* = 7.2 Hz, ³*J* = 1.6 Hz, 1H, 4'-H), 4.05 (dd, ²*J* = 11.7 Hz, ³*J* = 2.1 Hz, 1H, 5'-H), 3.78 (dd, ²*J* = 11.7 Hz, ³*J* = 1.8 Hz, 1H, 5'-H), 0.93 - 0.90 (m, 18H, Si(CH₃)₂(^tBu)), 0.12 - 0.10 (m, 12H, Si(CH₃)₂(^tBu)). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 163.1 (d, ¹*J*_{C-N} = 8.7 Hz, 4-C), 150.0 (sd, ¹*J*_{C-N} = 17.8 Hz, 2-C), 139.9 (C-6), 102.5 (C-5), 93.2 (d, ¹*J*_{C-F} = 193.3 Hz, C-2'), 87.9 (d, ²*J*_{C-F} = 33.8 Hz, C-1'), 83.9 (C-4'), 68.7 (d, ²*J*_{C-F} = 16.2 Hz, C-3'), 60.9 (C-5'), 26.1 (Si(C(CH₃)₃)), 25.7 (Si(C(CH₃)₃)), 18.5 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.5 (Si(CH₃)₂^tBu), -4.9 (Si(CH₃)₂^tBu), -5.3 (Si(CH₃)₂^tBu), -5.4 (Si(CH₃)₂^tBu). **¹⁹F-NMR (377 MHz, CDCl₃, ppm):** δ = -202.5 (ddd, ²*J*_{H-F} = 52.4 Hz, ³*J*_{H-F} = 19.0, ³*J*_{H-F} = 15.3). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -224.4. **HRMS (ESI⁺):** calc. for C₂₁H₄₀FN¹⁵NO₅Si₂⁺ [M]⁺: 476.2425, found: 476.2429. **HRMS (ESI⁻):** calc. for C₂₁H₃₈FN¹⁵NO₅Si₂⁻ [M-H]⁻: 474.2279, found: 474.2291. **IR (ATR):** ν (cm⁻¹) = 3381 (w), 3254 (w), 2953 (w), 2927 (m), 2856 (m), 1749 (w), 1713 (s), 1707 (s), 1692 (s), 1681 (s), 1629 (w), 1471 (w), 1452 (m), 1398 (w), 1388 (w), 1362 (w), 1321 (w), 1362 (m), 1321 (w), 1276 (m), 1251 (m), 1230 (w), 1157 (m), 1120 (s), 1094 (m), 1072 (m), 1052 (m), 993 (m), 978 (m), 938 (w), 899 (w), 880 (m), 863 (m), 828 (s), 811 (m), 802 (s), 778 (s), 757 (s), 748 (m), 711 (m), 704 (m), 663 (m).

3',5'-bis-*O*-[*tert*-butyl(dimethyl)silyl]-2'-deoxy-2'-(*R*)-fluoro-(*N*³,*N*⁴-¹⁵N₂)-cytidine (34)

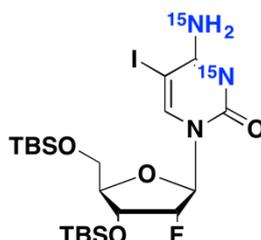


In a round bottom flask, 1,2,4-triazole (1.04 g, 15.1 mmol, 9.9 eq) was dissolved in MeCN (30 mL), cooled to 0 °C and stirred for ten minutes. To the resulting turbid solution, NEt₃ (2.00 mL, 14.4 mmol, 9.4 eq) was added dropwise and the mixture was stirred at 0 °C for another 20 minutes. Subsequently, a solution of **33** (730 mg, 1.53 mmol, 1.0 eq) in MeCN (10 mL) was given to the reaction and heated to 60 °C for 14 hours. After the starting material was consumed (as judged by TLC analysis), the mixture was quenched through addition of saturated NaHCO₃ (100 mL) and the resulting aqueous layer was extracted with DCM (3 x 100 mL). Combined organic layers were dried over Na₂SO₄ and the solvents were removed under reduced pressure. The resulting crude triazole derivative of **26** was then resuspended in dioxane (3.0 mL) and an aqueous solution of ¹⁵NH₃ (3N, 3.0 mL) was added. After the reaction was stirred at room temperature for 24 hours, starting material was still visible on the TLC. Thus, the heat was slightly increased to 30 °C and stirred for another 16 hours. The reaction was stopped through addition of saturated aqueous solution of NH₄Cl (50 mL), extracted with DCM (3 x 50 mL) and the combined organic layers were backwashed with brine (100 mL) and subsequently dried over Na₂SO₄. Purification through column chromatography (DCM --> 2% MeOH/DCM --> 3% MeOH/DCM --> 4% MeOH/DCM) yielded **34** (430 mg, 0.90 mmol, 59%) as a colourless foam.

¹H-NMR (800 MHz, CDCl₃, ppm): δ = 8.07 (d, ³J = 7.5 Hz, 1H, 6-H), 6.03 (dd, ³J_{H-F} = 15.9 Hz, ³J = 0.9 Hz, 1H, 1'-H), 5.80 (d, ³J = 7.5 Hz, 1H, 5-H), 4.79 (dd, ²J_{H-F} = 52.2 Hz, ³J = 3.9 Hz, 1H, 2'-H), 4.23 (ddd, ³J_{H-F} = 22.6 Hz, ³J = 8.6 Hz, ³J = 3.9 Hz, 1H, 3'-H), 4.12 - 4.07 (m, 2H, 4'-H + 5'-H), 3.79 (dd, ²J = 11.9 Hz, ³J = 1.8 Hz, 1H, 5'-H), 0.94 (s, 9 H, Si(CH₃)₂(^tBu)), 0.89 (s, 9 H, Si(CH₃)₂(^tBu)), 0.13 (s, 3H, Si(CH₃)₂(^tBu)), 0.12 (s, 3H, Si(CH₃)₂(^tBu)), 0.10 (s, 3H, Si(CH₃)₂(^tBu)), 0.08 (s, 3H, Si(CH₃)₂(^tBu)). **¹³C-NMR (201 MHz, CDCl₃, ppm):** δ = 165.0 (dd, ¹J_{C-N} = 21.8 Hz, ¹J_{C-N} = 8.54 Hz 4-C), 154.9 (sd, ¹J_{C-N} = 9.2 Hz, 2-C), 141.5 (C-6), 94.5 (C-5), 93.4 (d, ¹J_{C-F} = 192.3 Hz, C-2'), 88.9 (d, ²J_{C-F} = 33.4 Hz, C-1'), 83.1 (C-4'), 68.0 (d, ²J_{C-F} = 17.3 Hz, C-3'), 60.4 (C-5'), 26.1 (Si(C(CH₃)₃)), 25.8 (Si(C(CH₃)₃)), 18.6 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.5 (Si(CH₃)₂(^tBu)), -4.9 (Si(CH₃)₂(^tBu)), -5.3 (Si(CH₃)₂(^tBu)), -5.4 (Si(CH₃)₂(^tBu)). **¹⁹F-NMR (376 MHz, CDCl₃, ppm):** δ = -201.3 (ddd, ²J_{H-F} = 51.8 Hz, ³J_{H-F} = 20.8, ³J_{H-F} = 15.6). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -288.1, -288.3. **HRMS (ESI⁺):** calc. for C₂₁H₄₁FN¹⁵N₂O₄Si₂⁺ [M]⁺: 476.2555, found: 476.2558. **HRMS (ESI⁻):** calc. for C₂₁H₃₉FN¹⁵N₂O₄Si₂⁻ [M-H]⁻: 474.2409, found: 474.2411. **IR (ATR):** ν (cm⁻¹) = 3170 (bw), 2954 (w),

2929 (w), 2857 (w), 1717 (w), 1645 (m), 1519 (w), 1506 (m), 1394 (w), 1362 (w), 1278 (w), 1253 (m), 1157 (m), 1157 (w), 1120 (m), 1071 (m), 991 (w), 881 (m), 860 (m), 835 (s), 812 (m), 777 (s), 670 (m), 592 (m).

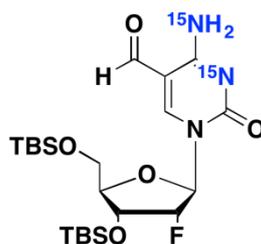
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)-5-iodo-2'-deoxy-2'-(*R*)-fluoro-(*N*³,*N*⁴-¹⁵N₂)-cytidine (35**)**



In a round bottom flask, **34** (30 mg, 0.06 mmol, 1.0 eq) was dissolved in MeCN (5 mL) and ceric ammonium nitrate (71 mg, 0.13 mmol, 2.2 eq) and iodine (32 mg, 0.13 mmol, 2.2 eq) were added. The dark brownish mixture was then stirred at 60 °C for a total time of three hours and after the reaction was cooled to room temperature, MeCN was removed *in vacuo*. Addition of DCM (20 mL) lead to the precipitation of cerium salts, which were filtered off and the DCM was finally removed under reduced pressure. Column chromatography (5% MeOH/DCM) of the crude mixture yielded **35** (37 mg, 60 μmol, quant.) as an orange oil.

¹H-NMR (400 MHz, CDCl₃, ppm): δ = 7.99 (s, 1H, 6-H), 5.98 (dd, ³J_{H-F} = 16.0 Hz, ³J = 1.8 Hz, 1H, 1'-H), 4.81 (ddd, ²J_{H-F} = 52.3 Hz, ³J = 4.4 Hz, ³J = 1.8 Hz, 1H, 2'-H), 4.19 (ddd, ³J_{H-F} = 18.9 Hz, ³J = 7.6 Hz, ³J = 4.4 Hz, 1H, 3'-H), 4.09 - 4.01 (m, 2H, 4'-H + 5'-H), 3.78 (dd, ²J = 11.9 Hz, ³J = 2.1 Hz, 1H, 5'-H), 0.95 (s, 9 H, C3'/C5'-O-Si(CH₃)₂(^tBu)), 0.89 (s, 9 H, Si(CH₃)₂(^tBu)), 0.16 (s, 3H, Si(CH₃)₂(^tBu)), 0.15 (s, 3H, Si(CH₃)₂(^tBu)), 0.10 (s, 3H, Si(CH₃)₂(^tBu)), 0.08 (s, 3H, Si(CH₃)₂(^tBu)). **¹³C-NMR (101 MHz, CDCl₃, ppm):** δ = 164.0 (dd, ¹J_{C-N} = 23.0 Hz, ¹J_{C-N} = 6.0 Hz, 4-C), 154.5 (sd, ¹J_{C-N} = 8.9 Hz, 2-C), 146.4 (C-6), 57.0 (C-5), 93.1 (d, ¹J_{C-F} = 193.3 Hz, C-2'), 89.2 (d, ²J_{C-F} = 33.0 Hz, C-1'), 83.9 (C-4'), 68.8 (d, ²J_{C-F} = 16.5 Hz, C-3'), 61.1 (C-5'), 57.4 (C-5), 26.5 (Si(C(CH₃)₃)), 25.8 (Si(C(CH₃)₃)), 18.9 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.4 (Si(CH₃)₂(^tBu)), -4.7 (Si(CH₃)₂(^tBu)), -4.8 (Si(CH₃)₂(^tBu)), -4.9 (Si(CH₃)₂(^tBu)). **¹⁹F-NMR (376 MHz, CDCl₃, ppm):** δ = -201.7 (ddd, ²J_{H-F} = 52.3 Hz, ³J_{H-F} = 19.0, ³J_{H-F} = 16.8). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -280.4, -280.5. **HRMS (ESI⁺):** calc. for C₂₁H₄₀FIN¹⁵N₂O₄Si₂⁺ [M]⁺: 602.1521, found: 602.1530. **HRMS (ESI⁻):** calc. for C₂₁H₃₈FIN¹⁵N₂O₄Si₂⁻ [M-H]⁻: 600.1376, found: 600.1378. **IR (ATR):** ν (cm⁻¹) = 3305 (w), 2955 (w), 2916 (s), 2850 (s), 1730 (s), 1635 (m), 1470 (s), 1418 (w), 1382 (w), 1361 (w), 1276 (m), 1253 (m), 1220 (m), 1178 (s), 1123 (m), 1105 (m), 1062 (m), 1047 (m), 992 (m), 942 (w), 889 (m), 864 (m), 836 (s), 777 (s), 740 (m), 719 (m), 670 (m), 643 (m), 622 (m), 593 (m).

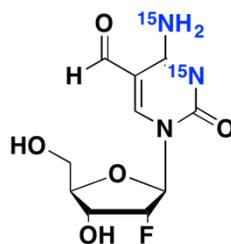
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)5-formyl-2'-deoxy-2'-(*R*)-fluoro-(*N*³,*N*⁴-¹⁵N₂)-cytidine (36)



The reaction vessel of a high pressure CO-autoclave was charged with **35** (220 mg, 0.37 mmol, 1.0 eq), PPh₃ (58 mg, 0.2 mmol, 0.6 eq), Pd₂(dba)₃CHCl₃ (38 mg, 0.04 mmol 0.1 eq) and toluene (5.5 mL). The mixture was purged with CO twice and finally set under a pressurized atmosphere of CO (50 psi). After heating to 65 °C, tributyltin hydride (120 μL, 0.44 mmol, 1.2 eq) was added to the reaction through a syringe pump (17 μL/h). After the addition was finished, the mixture was cooled to room temperature and the volatiles were removed under reduced pressure. The crude was subjected to purification by column chromatography (20% EtOAc/ ⁱHex --> 50% EtOAc/ ⁱHex) and **36** (158 mg, 0.31 mmol, 85%) was obtained as a yellow foam.

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 9.50 (s, 1H, CHO), 8.53 (s, 1H, 6-H), 8.26 (dd, ¹J_{H-N} = 92.9 Hz, ¹J = 6.0 Hz, 1H, ¹⁵NH₂), 7.29 (d, ¹J_{H-N} = 92.9 Hz, 1H, ¹⁵NH₂), 6.02 (dd, ³J_{H-F} = 16.3 Hz, ³J = 2.0 Hz, 1H, 1'-H), 4.92 (dd, ²J_{H-F} = 51.9 Hz, ³J = 4.1 Hz, 1H, 2'-H), 4.23 - 4.14 (m, 3H, 3'-H, 4'-H, 5'-H), 3.83 (dd, ²J = 12.2 Hz, ³J = 2.1 Hz, 1H, 5'-H), 0.94 (s, 9 H, Si(CH₃)₂(^tBu)), 0.89 (s, 9 H, Si(CH₃)₂(^tBu)), 0.14 (s, 3H, Si(CH₃)₂(^tBu)), 0.11 (s, 3H, Si(CH₃)₂(^tBu)), 0.10 (s, 3H, Si(CH₃)₂(^tBu)), 0.08 (s, 3H, Si(CH₃)₂(^tBu)). **¹³C-NMR (151 MHz, CDCl₃, ppm):** δ = 187.3 (CHO), 162.9 (dd, ¹J_{C-N} = 20.9 Hz, ¹J_{C-N} = 5.1 Hz, C-4), 153.2 (C-6), 153.0 (C-2), 105.7 (C-5), 93.0 (d, ¹J_{C-F} = 192.9 Hz, C-2'), 89.8 (d, ²J_{C-F} = 34.4 Hz, C-1'), 83.8 (C-4'), 68.2 (d, ²J_{C-F} = 16.5 Hz, C-3'), 60.7 (C-5'), 26.2 (Si(C(CH₃)₃)), 25.7 (Si(C(CH₃)₃)), 18.8 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.3 (Si(CH₃)₂(^tBu)), -4.8 (Si(CH₃)₂(C(CH₃)₃)), -4.8 (Si(CH₃)₂(^tBu)), -5.2 (Si(CH₃)₂(^tBu)). **¹⁹F-NMR (377 MHz, CDCl₃, ppm):** δ = -200.9 (ddd, ²J_{H-F} = 51.7 Hz, ³J_{H-F} = 19.9, ³J_{H-F} = 15.7). **¹⁵N-NMR (41 MHz, CDCl₃, ppm):** δ = -175.2, -286.0. **HRMS (ESI⁺):** calc. for C₂₂H₄₁FN¹⁵N₂O₅Si₂⁺ [M]⁺: 504.2504, found: 504.2513. **HRMS (ESI⁻):** calc. for C₂₂H₃₉FN¹⁵N₂O₅Si₂⁻ [M-H]⁻: 502.2358, found: 502.2357. **IR (ATR):** ν (cm⁻¹) = 3383 (bw), 2954 (w), 2929 (w), 2857 (w), 1663 (s), 1636 (m), 1506 (m), 1472 (w), 1418 (w), 1361 (w), 1252 (m), 1237 (m), 1167 (w), 1120 (m), 1070 (s), 993 (m), 979 (w), 835 (s), 813 (m), 777 (s), 699 (w), 667 (m), 598 (m), 575 (w).

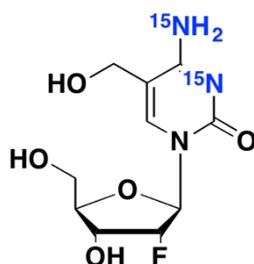
5-formyl-2'-deoxy-2'-(*R*)-fluoro-(N^3, N^4 - $^{15}\text{N}_2$)-cytidine (**20**)



For the TBS deprotection, **36** (100 mg, 0.20 mmol, 1.0 eq) was dissolved in a mixture of EtOAc (2.8 mL) and pyridine (0.8 mL). HF•pyridine (52 μL , 2.00 mmol, 10 eq) was added, the reaction was stirred for 16 hours at room temperature and a colourless precipitate formed. To quench the reaction, methoxytrimethyl silane (200 μL) was given to the mixture and stirred for another two hours at room temperature, whereas more of the colourless precipitate formed. The product was isolated through repeated washing with EtOAc after centrifuging (6000 rpm, three times in total) of the suspension. **20** (32 mg, 0.12 mmol, 58%) were yielded as a slightly yellowish powder, which was finally purified twice by semipreparative HPLC (Waters Nucleodur C18 ec, 0% MeCN/H₂O --> 13% MeCN/H₂O in 30 minutes, 5.0 mL/min) for cell-feeding purposes.

$^1\text{H-NMR}$ (800 MHz, D₂O, ppm): δ = 9.54 (s, 1H, CHO), 8.90 (s, 1H, 6-H), 6.10 (d, $^3J_{\text{H-F}} = 17.7$ Hz, 1H, 1'-H), 5.22 (dd, $^2J_{\text{H-F}} = 52.3$ Hz, $^3J = 4.3$ Hz, 1H, 2'-H), 4.39 (ddd, $^3J_{\text{H-F}} = 24.2$ Hz, $^3J = 9.3$ Hz, $^3J = 4.3$ Hz, 1H, 3'-H) 4.37 (dt, $^3J = 9.3$ Hz, $^3J = 3.0$ Hz, 1H, 4'-H), 4.15 (dd, $^2J = 13.1$ Hz, $^3J = 2.4$ Hz, 1H, 5'-H), 3.94 (dd, $^2J = 13.1$ Hz, $^3J = 3.5$ Hz, 1H, 5'-H). **$^{13}\text{C-NMR}$ (201 MHz, D₂O, ppm):** δ = 190.3 (CHO), 162.8 (dd, $^1J_{\text{C-N}} = 21.8$ Hz, $^1J_{\text{C-N}} = 6.5$ Hz, C-4), 155.0 (C-6), 154.9 (C-2), 105.9 (C-5), 93.7 (d, $^1J_{\text{C-F}} = 186.0$ Hz, C-2'), 89.9 (d, $^2J_{\text{C-F}} = 34.9$ Hz, C-1'), 82.5 (C-4'), 67.2 (d, $^2J_{\text{C-F}} = 16.7$ Hz, C-3'), 59.1 (C-5'). **$^{19}\text{F-NMR}$ (376 MHz, D₂O, ppm):** δ = -201.6 (ddd, $^2J_{\text{H-F}} = 52.3$ Hz, $^3J_{\text{H-F}} = 24.0$, $^3J_{\text{H-F}} = 17.7$). **$^{15}\text{N-NMR}$ (41 MHz, D₂O, ppm):** δ = -199.9, -200.4. **HRMS (ESI⁺):** calc. for C₁₀H₁₃FN¹⁵N₂O₅⁺ [M+H]⁺: 276.0774, found: 276.0778. **HRMS (ESI⁻):** calc. for C₁₀H₁₁FN¹⁵N₂O₅⁻ [M-H]⁻: 274.0629, found: 274.0629.

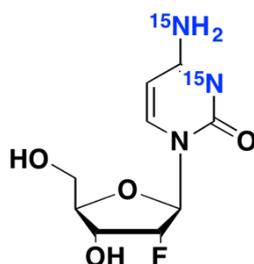
5-hydroxymethyl-2'-deoxy-2'-(R)-fluoro-(N^3,N^4 - $^{15}\text{N}_2$)-cytidine (37)



First, **20** (1.00 mg, 3.60 μmol , 1.0 eq) was dissolved in MeOH (1.0 mL), $\text{CeCl}_3 \cdot 7 \text{H}_2\text{O}$ (4.10 mg, 10.9 μmol , 3.0 eq) and NaBH_4 (0.20 mg, 3.60 μmol , 1.0 eq) were added subsequently, whereas an evolution of gas could be observed. After stirring at room temperature for one hour, complete conversion to the desired product could be observed (as judged by LC-MS). Then, a saturated aqueous solution of NH_4Cl (1.0 mL) was given to the crude and the solvent was removed by lyophilization. Purification of **37** was achieved through semipreparative HPLC (Waters Nucleodur C18 ec, 0% MeCN/ H_2O \rightarrow 10% MeCN/ H_2O in 45 minutes, 5.0 mL/min) and yielded 0.80 mg (2.80 μmol , 80%) of pure **37** as a white powder.

$^1\text{H-NMR}$ (800 MHz, D_2O , ppm): δ = 7.85 (s, 1H, 6-H), 5.92 (dd, $^3J_{\text{H-F}} = 19.4 \text{ Hz}$, $^3J = 1.1 \text{ Hz}$, 1H, 1'-H), 5.07 (dd, $^2J_{\text{H-F}} = 52.9 \text{ Hz}$, $^3J = 4.5 \text{ Hz}$, 1H, 2'-H), 4.35 (s, 2H, CH_2OH), 4.28 (ddd, $^3J_{\text{H-F}} = 22.0 \text{ Hz}$, $^3J = 8.9 \text{ Hz}$, $^3J = 4.5 \text{ Hz}$, 1H, 3'-H), 4.07 (dt, $^3J = 8.9 \text{ Hz}$, $^3J = 3.3 \text{ Hz}$, 1H, 4'-H), 3.97 (dd, $^2J = 13.0 \text{ Hz}$, $^3J = 2.4 \text{ Hz}$, 1H, 5'-H), 3.78 (dd, $^2J = 13.0 \text{ Hz}$, $^3J = 4.0 \text{ Hz}$, 1H, 5'-H). **$^{13}\text{C-NMR}$ (201 MHz, D_2O , ppm):** δ = 165.2 (dd, $^1J_{\text{C-N}} = 21.2 \text{ Hz}$, $^1J_{\text{C-N}} = 6.7 \text{ Hz}$, C-4), 156.9 (C-2), 140.9 (C-6), 106.5 (C-5), 93.7 (d, $^1J_{\text{C-F}} = 184.4 \text{ Hz}$, C-2'), 89.8 (d, $^2J_{\text{C-F}} = 35.0 \text{ Hz}$, C-1'), 82.1 (C-4'), 67.7 (d, $^2J_{\text{C-F}} = 16.1 \text{ Hz}$, C-3'), 59.6 (C-5'), 57.7 (CH_2OH). **$^{19}\text{F-NMR}$ (376 MHz, D_2O , ppm):** δ = -200.5 (ddd, $^2J_{\text{H-F}} = 52.9 \text{ Hz}$, $^3J_{\text{H-F}} = 22.0$, $^3J_{\text{H-F}} = 19.4$). **$^{15}\text{N-NMR}$ (41 MHz, D_2O , ppm):** δ = -201.9, -202.0. **HRMS (ESI $^+$):** calc. for $\text{C}_{10}\text{H}_{15}\text{FN}^{15}\text{N}_2\text{O}_5^+$ $[\text{M}+\text{H}]^+$: 278.0931, found: 278.0933. **HRMS (ESI $^-$):** calc. for $\text{C}_{10}\text{H}_{13}\text{FN}^{15}\text{N}_2\text{O}_5^-$ $[\text{M}-\text{H}]^-$: 276.0785, found: 276.0787.

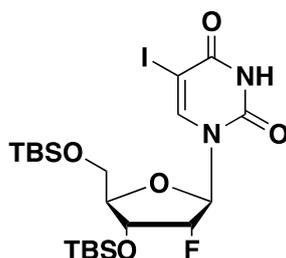
2'-deoxy-2'-(*R*)-fluoro-(N^3, N^4 - $^{15}\text{N}_2$)-cytidine (**19**)



34 (117 mg, 0.25 mmol, 1.0 eq) was dissolved in EtOAc (3.5 mL) and pyridine (98 μL). Subsequently HF•pyridine (70% solution, 123 μL , 4.92 mmol, 20 eq) was added and the mixture stirred at room temperature for 22 hours. TMSOMe (4.9 mL) was then added and after another hour of stirring at room temperature, the solvents were removed *in vacuo*. The residue was dissolved in H_2O and then purified by HPLC which yielded 44 mg (0.18 mmol, 72%) of **19** as a colourless solid.

$^1\text{H-NMR}$ (800 MHz, D_2O , ppm): δ = 7.81 (d, 3J = 7.6 Hz, 1H, 6-H), 6.02 (d, 3J = 7.7 Hz, 1H, 5-H), 5.97 (d, $^3J_{\text{H-F}}$ = 19.8 Hz, 1H, 1'-H), 5.15 (ddd, $^2J_{\text{H-F}}$ = 52.9 Hz, 3J = 4.9 Hz, 3J = 1.1 Hz, 1H, 2'-H), 4.34 (ddd, $^3J_{\text{H-F}}$ = 22.5 Hz, 3J = 9.0 Hz, 3J = 4.6 Hz, 1H, 3'-H), 4.16-4.08 (m, 1H, 4'-H), 4.01 (dd, 2J = 13.0 Hz, 3J = 2.4 Hz, 1H, 5'-H), 3.83 (dd, 2J = 13.0 Hz, 3J = 4.4 Hz, 1H, 5'-H). **$^{13}\text{C-NMR}$ (201 MHz, D_2O , ppm):** δ = 162.6 (C4), 149.9 (C2), 140.6 (C6), 96.0 (C5), 93.2 (d, $^1J_{\text{C-F}}$ = 193.8 Hz, C2'), 89.9 (d, $^2J_{\text{C-F}}$ = 35.7 Hz, C1'), 82.2 (C4'), 67.8 (d, $^2J_{\text{C-F}}$ = 15.5 Hz, C3'), 59.9 (C5'). **$^{19}\text{F-NMR}$ (376 MHz, D_2O , ppm):** δ = -200.3 (dt, $^2J_{\text{F-H}}$ = 53.1 Hz, $^3J_{\text{F-H}}$ = 19.0 Hz). **HRMS (ESI $^-$):** calc. for $\text{C}_{11}\text{H}_{15}\text{FN}^{15}\text{N}_2\text{O}_6^-$ [$\text{M}+\text{OAc}$] $^-$: 306.0891, found: 306.0894.

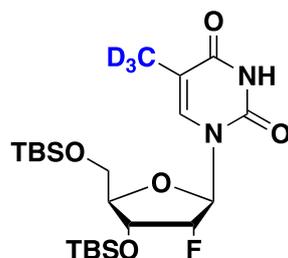
3',5'-bis-*O*-(*tert*-butyl(dimethyl)silyl)- 5-iodo-2'-deoxy-2'-(*R*)-fluoro -uridine (**38**)



TBS-protected 3',5'-bis-*O*-[*tert*-butyl(dimethyl)silyl]-2'-deoxy-2'-fluoro-uridine (0.50 g, 1.05 mmol, 1.0 eq) was dissolved in MeCN (100 mL) and CAN (2.30 g, 2.10 mmol, 2.0 eq) and I_2 (1.07 g, 2.10 mmol, 2.0 eq) were added to the mixture. The dark red brownish solution was heated to 65 $^\circ\text{C}$ for 1.5 hours and subsequently poured into saturated NaHCO_3 and a saturated solution of $\text{Na}_2\text{S}_2\text{O}_3$ was added until the solution became colourless. The aqueous phases were extracted with EtOAc (2x50 mL) and the combined organic fractions dried over Na_2SO_4 . Column chromatography (20% EtOAc/*Hex*) yielded the iodinated compound as a brownish foam (469 mg, 0.78 mmol, 74%).

¹H-NMR (400 MHz, CDCl₃, ppm): δ = 9.34 (s, 1H, NH), 7.98 (s, 1H, 6-H), 6.08 (dd, ³J_{H-F} = 14.3 Hz, ³J = 3.8 Hz, 1H, 1'-H), 4.82 (dt, ²J_{H-F} = 52.5 Hz, ³J = 4.2 Hz, 1H, 2'-H), 4.27 (dt, ³J_{H-F} = 12.9 Hz, ³J = 4.9 Hz, 1H, 3'-H), 4.08 - 4.06 (m, 1H, 4'-H), 3.98 (dd, ²J = 11.9 Hz, ³J = 1.8 Hz, 1H, 5'-H), 3.76 (dd, ²J = 11.9 Hz, ³J = 2.1 Hz, 1H, 5'-H), 0.95 (s, 9 H, C3'/C5'-O-Si(CH₃)₂(^tBu)), 0.90 (s, 9 H, Si(CH₃)₂(^tBu)), 0.16 (s, 3H, Si(CH₃)₂(^tBu)), 0.15 (s, 3H, Si(CH₃)₂(^tBu)), 0.10 (s, 3H, Si(CH₃)₂(^tBu)), 0.08 (s, 3H, Si(CH₃)₂(^tBu)). **¹³C-NMR (101 MHz, CDCl₃, ppm):** δ = 160.0 (4-C), 149.9 (2-C), 144.0 (C-6), 92.4 (d, ¹J_{C-F} = 195.3 Hz, C-2'), 87.4 (d, ²J_{C-F} = 33.7 Hz, C-1'), 85.1 (C-4'), 69.6 (d, ²J_{C-F} = 15.6 Hz, C-3'), 69.2 (C-5'), 61.7 (C-5), 26.3 (Si(C(CH₃)₃)), 25.7 (Si(C(CH₃)₃)), 18.7 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.6 (Si(CH₃)₂^tBu), -5.0 (Si(CH₃)₂^tBu), -5.1 (Si(CH₃)₂^tBu), -5.1 (Si(CH₃)₂^tBu). **HRMS (ESI⁺):** calc. for C₂₁H₃₉FIN₂O₅Si₂⁺ [M+H]⁺: 601.1421, found:601.1432.

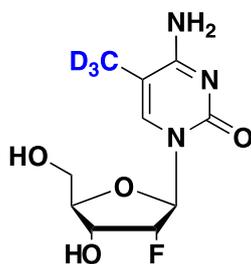
3',5'-bis-O-[tert-butyl(dimethyl)silyl]- 5-(D₃)-methyl 2'-deoxy-2'-(R)-fluoro-uridine (39)



38 (100 mg, 0.167 mmol, 1.0 eq) was dissolved in THF (4 mL) and Ni(dppp)Cl₂ (45 mg, 0.083 mmol, 0.5 eq) was added. The mixture was cooled to 0 °C and CD₃MgI (0.55 mL, 1M solution in ether, 3.3 eq) was added slowly. The clear red solution was slowly warmed to rt and stirred for one hour and subsequently quenched with sat. NH₄Cl (10 mL). The aqueous phase was extracted twice with EtOAc (15 mL), combined organic fractions dried over Na₂SO₄ and volatiles removed *in vacuo*. The crude mixture was finally purified through column chromatography (20% EtOAc/Hex) to yield 30 mg (0.061 mmol, 37%) of **39**.

¹H-NMR (400 MHz, CDCl₃, ppm): δ = 8.82 (s, 1H, NH), 7.38 (s, 1H, 6-H), 6.07 (dd, ³J_{H-F} = 15.5 Hz, ³J = 3.6 Hz, 1H, 1'-H), 4.83 (dt, ²J_{H-F} = 52.9 Hz, ³J = 4.2 Hz, 1H, 2'-H), 4.29 (dt, ³J_{H-F} = 13.4 Hz, ³J = 5.1 Hz, 1H, 3'-H), 4.08 - 4.00 (m, 1H, 4'-H), 3.97 (dd, ²J = 11.8 Hz, ³J = 2.0 Hz, 1H, 5'-H), 3.76 (dd, ²J = 11.8 Hz, ³J = 2.8 Hz, 1H, 5'-H), 0.93 (s, 9 H, C3'/C5'-O-Si(CH₃)₂(^tBu)), 0.91 (s, 9 H, Si(CH₃)₂(^tBu)), 0.12 - 0.18 (m, 12H, Si(CH₃)₂(^tBu)). **¹³C-NMR (101 MHz, CDCl₃, ppm):** δ = 163.7 (4-C), 150.2 (2-C), 135.5 (C-6), 111.6 (5-C), 92.3 (d, ¹J_{C-F} = 194.2 Hz, C-2'), 87.4 (d, ²J_{C-F} = 33.5 Hz, C-1'), 84.5 (C-4'), 69.5 (d, ²J_{C-F} = 15.8 Hz, C-3'), 61.7 (C-5'), 26.0 (Si(C(CH₃)₃)), 25.7 (Si(C(CH₃)₃)), 18.5 (Si(C(CH₃)₃)), 18.2 (Si(C(CH₃)₃)), -4.7 (Si(CH₃)₂^tBu), -5.1 (Si(CH₃)₂^tBu), -5.3 (Si(CH₃)₂^tBu), -5.3 (Si(CH₃)₂^tBu). **HRMS (ESI⁺):** calc. for C₂₂H₃₉D₃FN₂O₅Si₂⁺ [M+H]⁺: 492.2799, found: 492.2805.

5-(D₃)-methyl-2'-deoxy-2'-(R)-fluoro-5 -cytidine (18)



39 (30 mg, 0.06 mmol, 1.0 eq) was dissolved in pyridine (1.0 mL) and 1,2,4-triazole (34 mg, 0.49 mmol, 8.0 eq) and POCl₃ (12 μL, 0.12 mmol, 1.2 eq) were added at rt. The mixture was stirred for two hours and then poured into sat. NH₄Cl (10 mL). The resulting aqueous phase was quickly extracted with EtOAc (10 mL), dried over Na₂SO₄ and the solvents removed *in vacuo*. The residue was redissolved in dioxane (3.0 mL) and NH₄OH (25%, 1.0 mL) was added. After stirring at rt overnight, volatiles were removed *in vacuo* and the crude taken up in EtOAc (0.5 mL). Subsequently, HF•pyridine (70% solution, 8 μL, 0.32 mmol, 5.4 eq) and pyridine (12 μL) were added stirred at room temperature overnight. HF•pyridine was quenched through addition of TMSOMe (100 μL) and the residue was finally purified by HPLC (0 → 10% MeCN/H₂O in 45 mins) to give **18** as a white powder.

¹H-NMR (400 MHz, D₂O, ppm): δ = 7.54 (s, 1H, 6-H), 5.87 (dd, ³J_{H-F} = 19.8 Hz, ³J = 1.2 Hz, 1H, 1'-H), 5.00 (dd, ²J_{H-F} = 53.0 Hz, ³J = 4.6 Hz, 1H, 2'-H), 4.23 (ddd, ³J_{H-F} = 22.3 Hz, ³J = 8.9 Hz, ³J = 4.6 Hz, 1H, 3'-H), 4.04-3.97 (m, 1H, 4'-H), 3.91 (dd, ²J = 13.0 Hz, ³J = 2.4 Hz, 1H, 5'-H), 3.72 (dd, ²J = 13.0 Hz, ³J = 4.3 Hz, 1H, 5'-H). **¹³C-NMR (101 MHz, D₂O, ppm):** δ = 166.2 (C4), 157.0 (C2), 139.1 (C6), 104.4 (C5), 93.7 (d, ¹J_{C-F} = 184.5 Hz, C2'), 89.7 (d, ²J_{C-F} = 35.1 Hz, C1'), 82.1 (C4'), 67.8 (d, ²J_{C-F} = 16.6 Hz, C3'), 59.8 (C5').

HRMS (ESI⁺): calc. for C₁₀H₁₂D₃FN₂O₄⁺ [M+H]⁺: 263.1229, found:263.1230.

References

- 1 Schröder, A. S. *et al.* 2'-(R)-fluorinated mC, hmC, fC and caC triphosphates are excellent substrates for DNA polymerases and TET-enzymes. *Chem. Commun.* **52**, 14361-14364, (2016).
- 2 Brunner, K. *et al.* Cell-penetrating and neurotargeting dendritic siRNA nanostructures. *Angew. Chem. Int. Ed.* **54**, 1946-1949, (2015).
- 3 Saito, Y., Zevaco, T. A. & Agrofoglio, L. A. Chemical synthesis of ¹³C labeled anti-HIV nucleosides as mass-internal standards. *Tetrahedron* **58**, 9593-9603, (2002).
- 4 Schiesser, S. *et al.* Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed.* **51**, 6516-6520, (2012).

- 5 Robins, M. J. & Wilson, J. S. Smooth and efficient deoxygenation of secondary alcohols. A general procedure for the conversion of ribonucleosides to 2'-deoxynucleosides. *J. Am. Soc.* **103**, 932-933, (1981).

3.2 Anwendung eines neuartigen Reagenzes zur Quantifizierung von abasischen und β -eliminierten abasischen Stellen in genomischer DNA embryonaler Stammzellen im epigenetischen Kontext

„5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells“ René Rahimoff[#], Olesya Kosmatchev[#], Angie Kirchner[#], Toni Pfaffeneder, Fabio Spada, Victor Brantl, Markus Müller, Thomas Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359 – 10364.

[#]: Diese Autoren haben zu gleichen Teilen beigetragen.

Prolog

Die Entdeckung der von mdC ausgehend höher oxidierten Derivate hmdC, fdC und cadC legten die Grundlage für ein neues Verständnis der epigenetischen Regulation auf DNA-Ebene. Es wurde dabei schon früh festgestellt, dass fdC und cadC Substrate für die Basenexzisionsreparatur darstellen und unter Erzeugung einer abasischen Stelle aus dem Genom entfernt werden könnten. Letztendlich wird die dabei entstandene Lücke wieder mit unmodifiziertem dC aufgefüllt, wodurch eine aktive Demethylierung von mdC zu dC erreicht wäre. Um diesen Vorgang näher zu durchleuchten, beschreibt dieses Manuskript die Entwicklung und Anwendung eines neuartigen Hydroxylamin-Reagenz, welches für die spezifische Derivatisierung von abasischen Stellen in genomischer DNA entwickelt wurde. Das Molekül wurde nach einem struktur- und funktionsbasierten Design aufgebaut, um eine sensitive und zuverlässige Detektion und exakte Quantifizierung von abasischen Stellen zu ermöglichen. Zu diesen Strukturelementen gehört neben der reaktiven Hydroxylamin-Einheit ein Phenyl-Triazol-Kern, welcher durch sogenannte *stacking*-Effekte mit den flankierenden Basen durch Interkalation in die doppelsträngige DNA eine erhöhte Affinität besitzt. Außerdem kann die Triazol-Einheit unter Abspaltung von molekularem Stickstoff im verwendeten *Triple-Quadrupole*-Massenspektrometer unter milden Bedingungen fragmentiert werden, wodurch eine hohe Sensitivität bei einem sehr geringen *limit of detection* (LOD) erreicht wird (110 amol). Das Molekül weist außerdem eine quaternäre Ammoniumeinheit auf, welche zu intensiven Signalen im ESI-Experiment, einem definierten Ladungszustand von +1 und potentiell zu einer weiteren Erhöhung der Affinität zu doppelsträngiger DNA führt, da die einfach positiv geladene Spezies mit dem negativ geladenen Phosphatrückgrat interagieren kann. Die Koppelung dieser hochsensitiven Derivatisierungsmethode mit UHPL-Chromatographie und mit Hilfe der etablierten Isotopenverdünnungsmethode ist die genaue Identifizierung und Quantifizierung von abasischen Stellen möglich. Erstmals gelang es auch Fragmente wie die β -eliminierten abasischen Stellen eindeutig von abasischen Stellen zu unterscheiden. Es wurde zunächst gefunden, dass in einem Gleichgewichtszustand im Genom embryonaler Stammzellen von Mäusen mehr β -eliminierte als abasische Stellen vorhanden sind. Um den Ursprung dieser detektierten abasischen Stellen zu definieren, wurden im Zuge der vorgelegten Studie weiterhin

Isotopenverfolgungsexperimente mit isotopenmarkierten Nukleosiden durchgeführt. Diese waren an der Ribose mit ^{13}C -Atomen markiert, wodurch im Falle eines Basenverlusts ein abasisches Fragment entstehen würde, welches um fünf Masseeinheiten schwerer ist als eine unmarkierte abasische Stelle. Isotopenmarkierte Versionen von dG, dC und dT wurden daher sequentiell an Stammzellen von Mäusen verfüttert, wodurch eine genaue Rückverfolgung des Ursprungs der abasischen Stellen möglich wurde. Interessanterweise konnte festgestellt werden, dass im Falle von dC, von welchem auch epigenetisch relevante abasische Stellen erzeugt werden sollten, nur isotopenmarkierte β -Eliminierungsfragmente gefunden und quantifiziert wurden. Dies deutet im *steady-state* auf eine aktive BER dieser Basen durch bifunktionelle Glykosylasen hin. Außerdem waren die detektierten Mengen sehr gering und konnten auch im Falle einer Inhibition von Ape1 nicht erhöht werden. Dies könnte die Vermutung bestätigen, dass die Entfernung von fdC und cadC im Rahmen der aktiven Demethylierung von mdC zu dC in einem geschlossenen Enzymkomplex erfolgt, bei dem abasische Stellen nicht exponiert im Genom vorkommen.

Autorenbeitrag

Die Entwicklung der Synthese und Charakterisierung des isotopenmarkierten Standards der β -eliminierten abasischen Stelle sowie die Optimierung der Reaktionsbedingungen und der Aufreinigung der Derivatisierungsreaktion wurde von mir durchgeführt. Dieser Teil basierte auch zu Teilen auf Arbeiten meiner Masterarbeit. Die Isolation, Probenpräparation und Derivatisierung der genomischen DNA embryonaler Stammzellen wurde in enger Zusammenarbeit mit *Angie Kirchner*, die Quantifizierung der derivatisierten genomischen DNA wurde von *Olesea Kosmatchev* durchgeführt. Sämtliche Experimente für dieses Manuskript wurden in enger Zusammenarbeit mit *Angie Kirchner* und *Olesea Kosmatchev* entwickelt und geplant. Alle erhaltenen Daten wurden von *Olesea Kosmatchev*, *Angie Kirchner*, *Fabio Spada*, *Markus Müller*, *Thomas Carell* und mir diskutiert und interpretiert.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags; Copyright 2017 *American Chemical Society*.

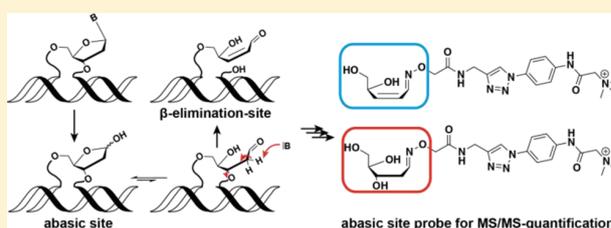
5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells

René Rahimoff,[†] Olesea Kosmatchev,[†] Angie Kirchner,[†] Toni Pfaffeneder, Fabio Spada, Victor Brantl, Markus Müller,[ⓑ] and Thomas Carell^{*ⓑ}

Center for Integrated Protein Science at the Department of Chemistry, LMU Munich, Butenandtstrasse 5-13, Munich 81377, Germany

Supporting Information

ABSTRACT: 5-Formyl-dC (fdC) and 5-carboxy-dC (cadC) are newly discovered bases in the mammalian genome that are supposed to be substrates for base excision repair (BER) in the framework of active demethylation. The bases are recognized by the monofunctional thymine DNA glycosylase (Tdg), which cleaves the glycosidic bond of the bases to give potentially harmful abasic sites (AP-sites). Because of the turnover of fdC and cadC during cell state transitions, it is an open question to what extent such harmful AP-sites may accumulate during these processes. Here, we report the development of a new reagent that in combination with mass spectrometry (MS) allows us to quantify the levels of AP-sites. This combination also allowed the quantification of β -elimination (β E) products, which are repair intermediates of bifunctional DNA glycosylases. In combination with feeding of isotopically labeled nucleosides, we were able to trace the intermediates back to their original nucleobases. We show that, while the steady-state levels of fdC and cadC are substantially increased in Tdg-deficient cells, those of both AP- and β E-sites are unaltered. The levels of the detected BER intermediates are 1 and 2 orders of magnitude lower than those of cadC and fdC, respectively. Thus, neither the presence of fdC nor that of cadC in stem cells leads to the accumulation of harmful AP- and β E-site intermediates.



INTRODUCTION

The loss of a base is a major damaging event in the genome that leads to the formation of an abasic site (AP-site). Abasic sites are formed spontaneously or during base excision repair, when cells remove damaged bases from the genome.¹ This repair process requires the action of specific DNA glycosylases, which recognize the damaged base (Figure 1A).^{2–4} Two different types of repair glycosylases are known. Monofunctional glycosylases produce AP-sites (Figure 1A, step a) that are subsequently removed by the action of endonucleases such as the apurinic/apyrimidinic endonuclease (Ape) 1 (Figure 1A, step b),^{5,6} which introduces a free 3'-OH-group and a 5'-phosphate ester (5'-dRP) that is further processed by the action of polymerase β (Pol β). Bifunctional glycosylases catalyze a β -elimination (β E) reaction (Figure 1A, step c), which gives transient β E-intermediates. These can then be converted into a single nucleotide gap by a subsequent δ -elimination reaction (Figure 1A, step d). Both steps (c and d) are carried out by the bifunctional glycosylases.⁷

Recently, it was discovered that the monofunctional glycosylase Tdg (Thymine DNA glycosylase)^{8,9} removes the newly discovered bases 5-formyl-cytosine (fdC)¹⁰ and 5-carboxy-cytosine (cadC).^{9,11} Pronounced amounts of these bases can be found particularly in stem cells, and fdC was recently also discovered in the genome of brain cells.^{12–14} Because fdC and cadC are supposed to be intermediates of an active demethylation process, the data suggest that Tdg-

mediated base excision repair contributes to epigenetic reprogramming. One of the main questions associated with the Tdg-mediated repair process is to what extent potentially harmful AP- and β E-site intermediates are generated in the genome in response to the presence of fdC and cadC.

Here, we report the development of reagent 1¹⁵ (Figure 1B) that in combination with highly sensitive UHPLC-triple quadrupole mass spectrometry (MS)¹⁶ enables quantification of both AP-sites and β E intermediates with unprecedented sensitivity (\sim 100 intermediates per genome). The method was used to quantify the steady-state levels of AP- and β E-sites during BER of fdC and cadC.

RESULTS AND DISCUSSION

Design and Synthesis of a Probe for MS-Based Abasic Site Quantification. Reagent 1 contains a reactive hydroxylamine unit, able to form stable and defined reaction products with the aldehydic form of both AP-sites and β E-intermediates.^{17,18} We discovered that the formed adducts (Figure 1B) do not disturb the action of hydrolytic enzymes (Supporting Information, section 3.8), so that the formed reaction products can be excised quantitatively from the genome. To achieve ultrahigh sensitivity, we developed the reagent for triple-quadrupole MS detection. In this method, not only is the

Received: April 27, 2017

Published: July 17, 2017

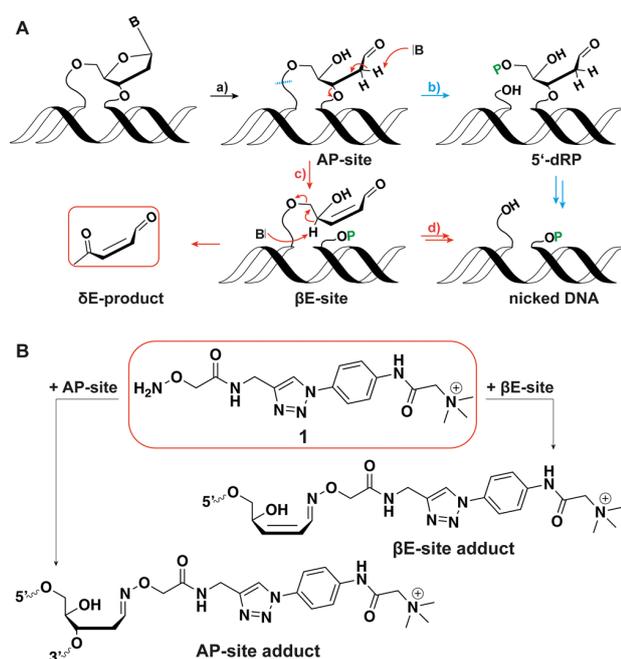


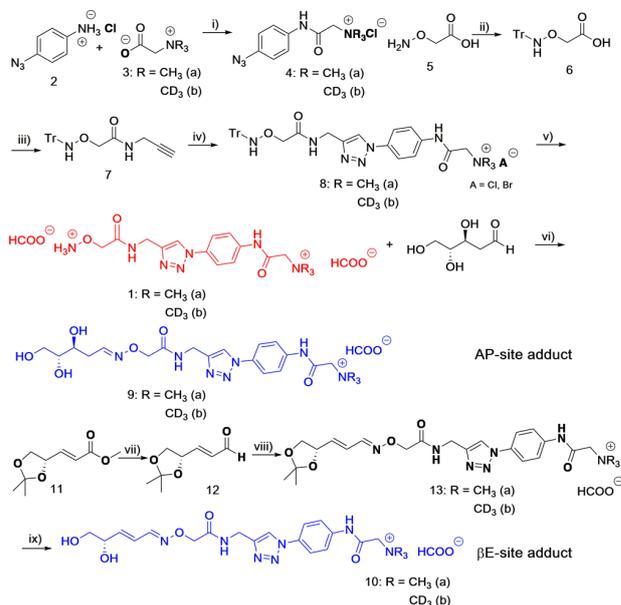
Figure 1. Mechanism of the base excision repair process. (A) Chemistry of BER with formation of AP- and β E-sites. Blue pathway: Monofunctional glycosylases act in concert with Ape1 and Pol β . Red pathway: Bifunctional glycosylases. (B) Depiction of reagent **1** and of the reaction products that are formed when **1** reacts with AP- and β E-sites.

molecular mass of the analyte observed, but also a specific mass transition that arises when the parent ion fragments. **1** therefore contains a triazole heterocycle that fragments in the triple quadrupole MS via collision-induced dissociation (CID) to give a loss of N_2 . The permanent positive charge that was designed into **1** ensures highest possible sensitivity in the mass spectrometer and was found to be beneficial, assuring fast reaction with the AP- and β E-sites that are embedded in the negatively charged DNA duplex. The coupling to mass spectrometry is essential to distinguish AP- and β E-intermediates, which has not been possible to date with previously established methods.^{17,19,20}

The synthesis of reagent **1** is shown in Scheme 1. It starts with *p*-azidoaniline **2**, which is reacted with trimethylamino glycine **3** using TBTU as the coupling reagent to give the azidoamide **4**. A stable isotopologue of reagent **1a**, needed for exact MS-based quantification (vide infra), is at this step accessed via a $(CD_3)_3$ -glycine derivative, which introduces nine D atoms making **1b** 9 au heavier (**1b**). *O*-(Carboxymethyl)-hydroxylamine **5** was in parallel Trt-protected to **6**,²¹ and **6** was reacted with propargyl amine using again TBTU as the coupling reagent to give the alkyne **7**. Reaction of **4** with **7** via a Cu(I)-catalyzed azide-alkyne click reaction furnished the triazole **8**. Cleavage of the Trt-group under harsh acidic conditions provided the reagents **1** in a light (CH_3 , **a**) and a heavy version (CD_3 , **b**), with a $\Delta m/z = 9$ in just five steps with a total yield of 47%.

The reagent was purified twice by reversed phase HPLC to establish a purity of >99.9%. For exact mass spectrometry-based quantification, we next needed to prepare the AP- and β E-site reaction products with **1** in unlabeled (**a**) and isotopically modified (**b**) versions to be used as internal standards. We

Scheme 1. Synthesis of **1** (Red) in a Light (**1a**) and Heavy (**1b**) Form and of the Light and Heavy Internal Standards **9a/b** and **10a/b** (Blue)^a



^a(i) TBTU, DIPEA, DMF, rt, 16 h, 90%; (ii) Trt-Cl, NET_3 , pyridine, rt, 22 h, 74%; (iii) propargylamine, TBTU, DIPEA, DCM, rt, 15 h, 92%; (iv) **7** + **4a/b**, CuBr·SMe₂, H₂O/DCM (1:1), rt, 16 h, 77%; (v) 6 M HCl, DCM/H₂O (1:1), rt, 1 h, quant.; (vi) H₂O, 30 °C, 16 h, HPLC, 15%; (vii) **1** DIBAL-H, DCM, -78 °C to rt, (2) DMP, DCM, 0 °C to rt, o/n, 47% over two steps; (viii) **1** **1a/b**, CHCl₃/H₂O (1:1), rt, 16 h, 68%, (2) pTSA-H₂O, H₂O, 25 °C, o/n, HPLC (2×), 14%.

reacted both isotopologues **1a/b** with 2'-deoxyribose, which provided the AP-site reaction products **9a/b**. To prepare the β E products **10a/b**, we reduced the acetonide-protected methyl ester **11** with DIBAL-H to the allyl alcohol, which was subsequently oxidized selectively to the aldehyde **12** using the Dess–Martin reagent. Reaction of **12** with reagent **1a/b** and final cleavage of the acetonide protecting group furnished the desired compounds **10a/b**, in a light and heavy version, respectively. Compounds **9a/b** and **10a/b** were finally purified by reversed phase HPLC to purities >99.9%.

We next developed the MS method (Figure 2A). Analysis of the AP-site reaction product **9a** provided a symmetric signal (at $t = 9.5$ min, for gradient see Supporting Information, section 3.3) for the MS transitions $m/z = 478.2 \rightarrow 450.2$ (quantifier) and $m/z = 478.2 \rightarrow 192.1$ (qualifier) caused by the two molecular fragments formed after CID fragmentation. One fragmentation is caused by the loss of N_2 , and the second fragmentation involves the formation of an aryl radical.

The first MS transition (quantifier) was used for exact quantification, while the second MS transition (qualifier) is needed for additional structure validation. The isotopologue **9b** showed the expected mass transitions $m/z = 487.3 \rightarrow 459.3$ and $m/z = 487.3 \rightarrow 201.2$. Similar data were obtained for the β E-reaction product **10a** and its isotopologue **10b** (Figure 2A). We next performed a dilution experiment with **9a** and **10a** and found that we could detect the compounds within a limit of detection of only 110 amol. This is due to the fact that **1** was designed so that its AP- and β E-adducts **9** and **10** have advantageous ionization (permanent positive charge) and fragmentation (triazole that leads to a neutral loss of N_2)

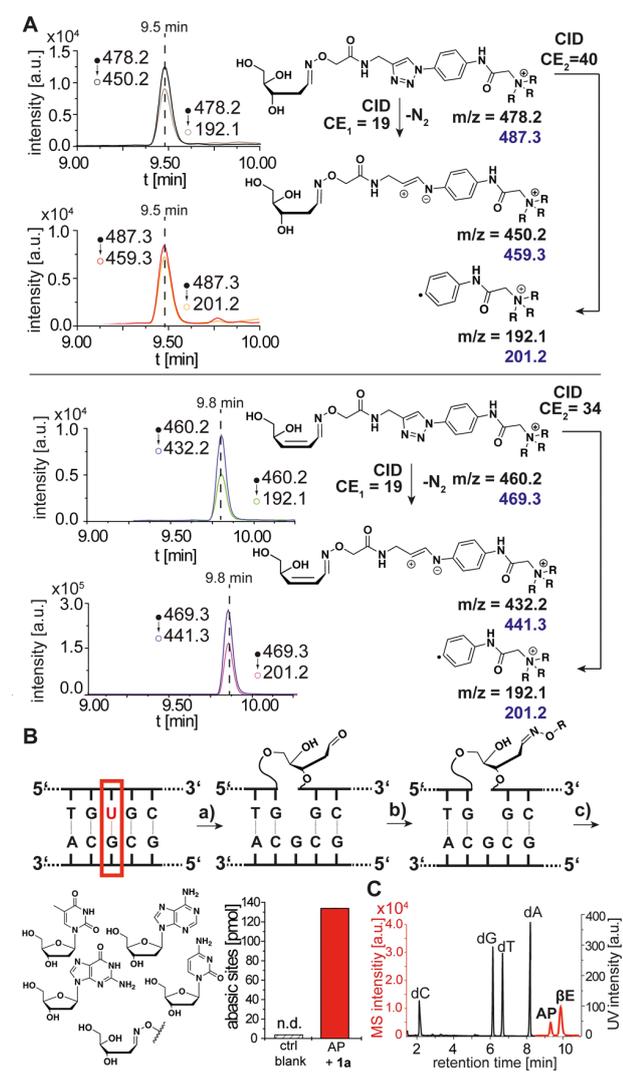


Figure 2. (A) Fragmentation pattern of **9a/b** and **10a/b**, which give defined daughter ions after loss of N_2 or an aryl radical. (B) Reaction of reagent **1** with a defined AP-site created inside a DNA duplex. Quantification of the reaction product **9a** after the treatment of the oligodeoxynucleotide with **1a**. (C) UHPLC-chromatogram of the canonical bases plus the AP (**9a**)- and β E-site (**10a**) reaction products formed with reagent **1a**.

properties that allow for ultrasensitive detection in a triple quadrupole mass spectrometer.

To investigate if the reagent is able to detect AP-sites directly in DNA, we prepared oligodeoxynucleotides with a single dU base (Figure 2B). To this DNA strand we added the U-cleaving monofunctional glycosylase UDG (uracil DNA glycosylase) to generate a defined AP-site (Supporting Information, section 3.6). We added **1a**, digested the DNA with a mixture of the enzymes nuclease S1, Antarctic phosphatase, and snake venom phosphodiesterase, added a defined amount of the internal standard **9b**, and quantified generated AP-sites using our developed method (Figure 2C; Supporting Information, section 3.8). Indeed, we clearly detected the AP-site adduct **9a** at the expected level, showing that **1a** reacts quantitatively under mild conditions and that the reaction product is excised quantitatively. Furthermore, we treated an 8-oxo-dG containing oligodeoxynucleotide with human 8-oxoguanine DNA glyco-

sylase 1 (OGG1) and observed the formation of both an AP- and a β E-site (Supporting Information, section 3.7).

Quantification of Abasic- and β E-Sites in Mouse Embryonic Stem Cells. To assess the utility of the method (derivatization plus mass spectrometry), we next measured the steady-state levels of AP- and β E-sites in mouse embryonic stem cells (mESC). Here, the removal of fdC and cadC by BER may lead to a particular increase of the levels of BER intermediates.^{22–24} Alternatively, it was shown that Tdg forms a complex with the ten-eleven translocation dioxygenase (Tet) 1 and Ape1 and possibly also with the nei endonuclease VIII-like (Neil) 1 and 2 enzymes. In these scenarios, BER takes place in a tight complex that does not allow accumulation of AP-sites.^{25,26} Naïve cultures of mESCs were primed for 5 days.²⁷ To measure the steady-state level of AP- and β E-sites, we treated the isolated genomic DNA before digestion with **1a**. To the digestion mixture were added the internal standards **9b** and **10b** in defined amounts, and the mixture was analyzed and quantified by UHPLC-MS.

It should be noted that this method does not allow the distinction between AP- and 5'-dRP-sites. As shown in Figure 3A, we measured a total of 8.8×10^{-7} AP- and 1.7×10^{-6} β E-

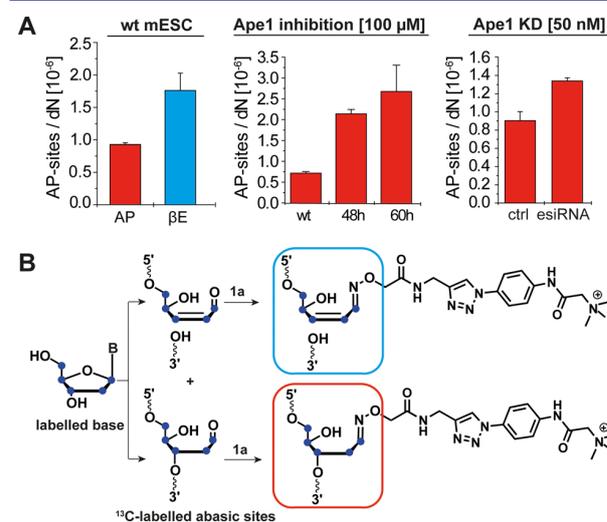


Figure 3. Quantification of BER intermediates and isotope tracing studies. (A) Quantitative data of global levels of AP- and β E-sites from mESCs. Inhibition and knockdown of Ape1 yields an increased amount of AP-sites. (B) Feeding of mESC cultures with labeled nucleosides results in the formation of ribose-labeled AP- and β E-site products **9*** and **10***, which are five mass units heavier than unlabeled products.

sites per dN, which is comparable to the levels in other cell types (Figure S4). These quantitative data show that in mESCs the steady-state levels of total AP- and β E-sites are quite substantial. Our data are slightly lower as compared to data obtained previously.^{17,19,20} While AP-sites also form by spontaneous depurination, this is not possible for β E-sites, showing that DNA glycosylases generate quite a number of harmful intermediates that need to be further processed. We measure a steady-state level of around ~ 9200 β E-sites per genome in mESCs. Several control experiments show that these intermediates are not generated during DNA isolation, handling, or derivatization (Supporting Information, sections 2.5 and 3.9). For example, to exclude that the measured AP-

and β E-intermediates are formed during the treatment of genomic DNA with **1a**, we incubated genomic DNA over different time points with **1a** and quantified the levels of AP- and β E-sites (Figure S9). Increasing levels of AP- and β E-sites were not detected even after a prolonged reaction time (up to 60 min), excluding this possibility. To exclude that the AP-sites are generated during the DNA isolation procedure, we studied how the major AP-endonuclease Ape1²⁸ influences the levels of AP-sites (Figure 3A).

This enzyme is known to operate together with monofunctional glycosylases to convert AP-sites into single nucleotide gaps. It is known that Ape1 and Tdg interact and that the release of Tdg from the AP-site depends on the presence of Ape1.²⁹ We therefore added the Ape1 inhibitor CRT0044867³⁰ to mESC cultures and measured the levels of AP-sites after treatment for 48 and 60 h. As shown in Figure 3A, we see a clear increase of AP-sites by a factor of 3–4, confirming that AP-sites accumulated when the activity of Ape1 is inhibited. On the basis of similar signal intensities, we concluded that the amount of β E-sites stayed unchanged in this experiment. We also down-regulated Ape1 expression using esiRNA in the last 2 days of priming. qPCR studies confirmed that the esiRNA approach reduces the expression level of Ape1 by 50%. This led to a 30% increased AP-site level (Figure 3A, right panel). These data show that our method is able to report the differential formation of BER intermediates *in vivo*.

Tracing Studies with Isotopically Labeled Nucleosides. We next analyzed the formation of AP- and β E-sites at the different types of DNA bases with a particular emphasis on dC and its derivatives mdC, hmdC, fdC, and cadC. To this end, we prepared four different mESC cultures and added either isotopically labeled dA*, dC*, dG*, or dT* nucleosides, in which all of the ¹²C atoms were exchanged against ¹³C. Correspondingly, the AP- and β E-sites formed from these bases are five atom units heavier than those obtained from unlabeled material (Figure 3B).

We first determined the steady-state levels of mdC, hmdC, fdC, and cadC. For this we isolated and digested the genomic DNA using a standard protocol (Supporting Information, sections 2.3 and 3.2). The obtained levels are as expected with fdC = 1.5×10^{-6} /dN and cadC = 8.9×10^{-7} /dN (Figure 4). Next, we monitored the efficiency of incorporation. We saw for

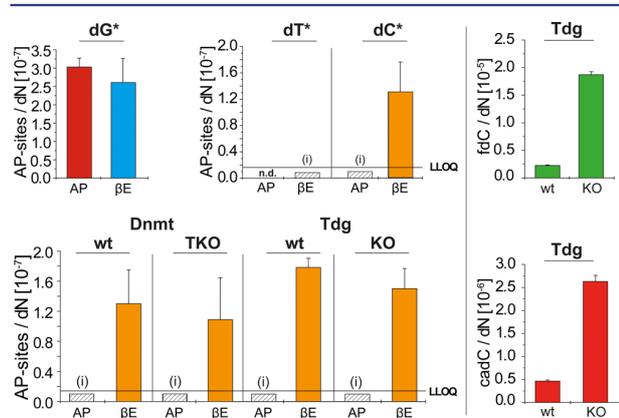


Figure 4. Quantification of labeled AP- and β E-intermediates of the isotope tracing studies and quantification of the fdC and cadC content in Tdg wt and KO cells. n.d., not detectable. (i) Levels were over the LOD, but under LLOQ.

¹³C₁₀-dG an incorporation of 93%. ¹³C₉-dC was incorporated to 40%, and the ¹³C₉-dT almost fully replaced dT (97% incorporation). Interestingly, no significant incorporation of dA was observed, making any study of this base impossible. We next performed exact quantification of the AP- and β E-sites at dG*, dT*, and dC* and normalized the obtained values for direct comparison. For BER at dG, which is known to be prone to oxidative damage and spontaneous depurination,³¹ we saw comparable amounts of AP- and β E-sites. We measured 2.7×10^{-7} sites per dN, which corresponds to a steady-state level of around ~ 1500 AP- and β E-sites derived from dG (Figure 4). The high level of β E-sites is consistent with the known BER repair pathway of the main oxidative lesion 8-oxoG that involves the action of the bifunctional DNA glycosylase Ogg1.³²

At dT, the levels of the β E-sites are below 10^{-8} per dN, which amounts to less than ~ 100 β E-intermediates per genome. This level is below the lower limit of quantification (LLOQ) of our reagent and mass spectrometry-based method (Figure 4). AP-sites deriving from dT were not detectable at all. The data show that spontaneous cleavage of the glycosidic bond of dT and BER at this base is a minor process.

Most interesting are the very low levels of AP-sites detected at dC* (below LLOQ). β E-sites are in contrast clearly measurable, and with 1.2×10^{-7} per dN they are about one-half of those measured for dG*. dC and its derivative mdC are prone to both spontaneous deamination^{33,34} and oxidation.³⁵ It has been recently reported that deamination preponderantly contributes to mutational load in tissues. Under cell culture conditions, however, oxidation seems to be the major source of DNA damage.³⁶ All known DNA glycosylases that process mismatches generated by deamination of dC-bases are monofunctional (uracil-N glycosylase (Ung) 2, single-strand specific monofunctional uracil DNA glycosylase (Smug) 1, Tdg, and methyl-binding domain glycosylase (Mbd) 4). Oxidized dC-bases are in contrast repaired by bifunctional Neil family DNA glycosylases.³⁷ The occurrence of oxidative damage in cultured cells is therefore consistent with higher levels of β E-products relative to AP-sites stemming from dC-type bases.

To study how mdC and its oxidized derivatives hmdC, fdC, and cadC contribute to the level of β E-sites, we repeated the dC* feeding study with mESCs lacking all three DNA methyltransferases (Dnmts). Indeed, these cells lack any genomic mdC, and, consequently, the oxidized bases hmdC, fdC, and cadC are not detected anymore. Despite this, we saw no AP- and β E-site changes, arguing that at the steady state the overwhelming majority of β E-sites are generated from canonical dC, for example, by oxidative damage. Obviously, if BER is low, the small amounts of fdC and cadC that are present in the wild type may provide only small amounts of AP- and β E-sites that are negligible. If BER is however a major process, the fact that we do not see changing AP- and β E-site levels could alternatively imply that the formed AP-sites are immediately processed.

To study this phenomenon in more detail, we next fed dC* to mESCs lacking the Tdg enzyme to interrogate the putative BER complex. We first measured the levels of fdC and cadC and saw the expected increase of fdC and cadC levels (Figure 4). Despite this, we again see unchanged levels of AP- and β E-sites. This inability to detect any significant change despite higher levels of fdC (2×10^{-5} per dN) and cadC (2.5×10^{-6} per dN) is puzzling. It may suggest that single nucleotide BER is no longer occurring or that fdC or cadC is removed by alternative pathways such as long patch BER or noncanonical

mismatch repair.^{38,39} Interestingly, in an experiment where we used the Ape1 inhibitor CRT0044867 and fed dC*, analysis of the labeled AP-sites did not show an increase as well. The measured levels stayed below the LLOQ (Figure S1A). We did detect, however, a small increase of the global AP-sites (Figure S1B). Because Ape1 is supposed to be in complex with Tdg, to quickly convert AP-sites, this lack of increase is surprising.

Our data are consistent with either a scenario in which the Tdg generated BER intermediates at fdC and cadC are quickly turned over without Ape1 or in which other processes remove fdC and cadC.

CONCLUSION

In summary, we report the development of reagent **1** that reacts quickly with aldehydes such as those formed during BER in the genome. The reagent **1** undergoes an efficient CID derived neutral molecule loss in the MS, which allows the ultrasensitive detection of AP- and β E-BER intermediates with 100 amol sensitivity. This, together with the feeding of isotopically labeled nucleosides, allowed us to show that removal of the epigenetically relevant bases fdC and cadC proceeds without the accumulation of significant amounts of these potentially harmful BER intermediates in the genome of stem cells. Alternatively, it may be that the BER-based removal of fdC and cadC is less important than thought. Further studies are now needed to fully decipher the BER processes that operate on fdC and cadC during the enigmatic process of active epigenetic demethylation.^{40,41}

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b04131.

Chemical synthesis, cell culture and transfection procedures, quantification of abasic sites, and additional references (PDF)

AUTHOR INFORMATION

Corresponding Author

*thomas.carell@cup.uni-muenchen.de

ORCID

Markus Müller: 0000-0002-3579-3317

Thomas Carell: 0000-0001-7898-2831

Author Contributions

†R.R., O.K., and A.K. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The Tdg KO mESC line was kindly provided by Primo Schär (University of Basel, Switzerland). We are also grateful to Masaki Okano (Kumamoto University, Japan) for providing the Dnmt TKO mESC line. We thank the Deutsche Forschungsgemeinschaft for financial support via SFB1032, SFB749, SPP1784, CA275, and the Excellence Cluster CiPSM.

REFERENCES

- (1) Lindahl, T. *Angew. Chem., Int. Ed.* **2016**, *55*, 8528–8534.
- (2) Fromme, J. C.; Banerjee, A.; Verdine, G. L. *Curr. Opin. Struct. Biol.* **2004**, *14*, 43–49.
- (3) Serre, L.; Pereira de Jesus, K.; Boiteux, S.; Zelwer, C.; Castaing, B. *EMBO J.* **2002**, *21*, 2854–2865.
- (4) Schneider, S.; Schorr, S.; Carell, T. *Curr. Opin. Struct. Biol.* **2009**, *19*, 87–95.
- (5) Stivers, J. T.; Jiang, Y. L. *Chem. Rev.* **2003**, *103*, 2729–2760.
- (6) Drohat, A. C.; Maiti, A. *Org. Biomol. Chem.* **2014**, *12*, 8367–8378.
- (7) Drohat, A. C.; Coey, C. T. *Chem. Rev.* **2016**, *116*, 12711–12729.
- (8) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334–35338.
- (9) 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*, 1303–1307.
- (10) Pfaffeneder, T.; Hackner, B.; Truss, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeyer, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008–7012.
- (11) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300–1303.
- (12) Globisch, D.; Münzel, M.; Müller, M.; Michalakakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. *PLoS One* **2010**, *5*, e15367.
- (13) Münzel, M.; Globisch, D.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 6460–6468.
- (14) Wagner, M.; Steinbacher, J.; Kraus, T. F. J.; Michalakakis, S.; Hackner, B.; Pfaffeneder, T.; Perera, A.; Müller, M.; Giese, A.; Kretzschmar, H. A.; Carell, T. *Angew. Chem., Int. Ed.* **2015**, *54*, 12511–12514.
- (15) Carell, T. LMU Munich, Patent Application EP17153895, 2017.
- (16) Tretyakova, N.; Villalta, P. W.; Kotapati, S. *Chem. Rev.* **2013**, *113*, 2395–2436.
- (17) Kubo, K.; Ide, H.; Wallace, S. S.; Kow, Y. W. *Biochemistry* **1992**, *31*, 3703–8.
- (18) Ide, H.; Akamatsu, K.; Kimura, Y.; Michiue, K.; Makino, K.; Asaeda, A.; Takamori, Y.; Kubo, K. *Biochemistry* **1993**, *32*, 8276–83.
- (19) Wei, S.; Shalhout, S.; Ahn, Y. H.; Bhagwat, A. S. *DNA Repair* **2015**, *27*, 9–18.
- (20) Roberts, K. P.; Sobrino, J. A.; Payton, J.; Mason, L. B.; Turesky, R. J. *Chem. Res. Toxicol.* **2006**, *19*, 300–309.
- (21) Kojima, N.; Takebayashi, T.; Mikami, A.; Ohtsuka, E.; Komatsu, Y. *J. Am. Chem. Soc.* **2009**, *131*, 13208–13209.
- (22) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334–35338.
- (23) Cortazar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; Steinacher, R.; Jiricny, J.; Bird, A.; Schär, P. *Nature* **2011**, *470*, 419–423.
- (24) 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*, 67–79.
- (25) Weber, A. R.; Krawczyk, C.; Robertson, A. B.; Kusnierczyk, A.; Vagbo, C. B.; Schuermann, D.; Klungland, A.; Schär, P. *Nat. Commun.* **2016**, *7*, 10806.
- (26) Schomacher, L.; Han, D.; Musheev, M. U.; Arab, K.; Kienhofer, S.; von Seggern, A.; Niehrs, C. *Nat. Struct. Mol. Biol.* **2016**, *23*, 116–24.
- (27) Nichols, J.; Smith, A. *Cell Stem Cell* **2009**, *4*, 487–492.
- (28) Marenstein, D. R.; Wilson Iii, D. M.; Teebor, G. W. *DNA Repair* **2004**, *3*, 527–533.
- (29) Fitzgerald, M. E.; Drohat, A. C. *J. Biol. Chem.* **2008**, *283*, 32680–90.
- (30) Madhusudan, S.; Smart, F.; Shrimpton, P.; Parsons, J. L.; Gardiner, L.; Houlbrook, S.; Talbot, D. C.; Hammonds, T.; Freemont, P. A.; Sternberg, M. J.; Dianov, G. L.; Hickson, I. D. *Nucleic Acids Res.* **2005**, *33*, 4711–4724.
- (31) Lindahl, T. *Nature* **1993**, *362*, 709–15.
- (32) Hill, J. W.; Hazra, T. K.; Izumi, T.; Mitra, S. *Nucleic Acids Res.* **2001**, *29*, 430–438.
- (33) Duncan, B. K.; Miller, J. H. *Nature* **1980**, *287*, 560–561.
- (34) Shen, J. C.; Rideout, W. M., 3rd; Jones, P. A. *Nucleic Acids Res.* **1994**, *22*, 972–976.

- (35) Madugundu, G. S.; Cadet, J.; Wagner, J. R. *Nucleic Acids Res.* **2014**, *42*, 7450–7460.
- (36) Rouhani, F. J.; Nik-Zainal, S.; Wuster, A.; Li, Y.; Conte, N.; Koike-Yusa, H.; Kumasaka, N.; Vallier, L.; Yusa, K.; Bradley, A. *PLoS Genet.* **2016**, *12*, e1005932.
- (37) Jacobs, A. L.; Schär, P. *Chromosoma* **2012**, *121*, 1–20.
- (38) Bochtler, M.; Kolano, A.; Xu, G. L. *BioEssays* **2017**, *39*, 1–13.
- (39) Santos, F.; Peat, J.; Burgess, H.; Rada, C.; Reik, W.; Dean, W. *Epigenet. Chromatin* **2013**, *6*, 39.
- (40) Ooi, S. K. T.; Bestor, T. H. *Cell* **2008**, *133*, 1145–1148.
- (41) Kohli, R. M.; Zhang, Y. *Nature* **2013**, *502*, 472–9.

Supporting Information

5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells

René Rahimoff[#], Olesea Kosmatchev[#], Angie Kirchner[#], Toni Pfaffeneder, Fabio Spada, Victor Brantl, Markus Müller, Thomas Carell*

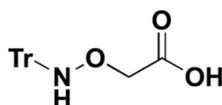
Center for Integrated Protein Science at the Department of Chemistry, LMU Munich,
Butenandtstrasse 5-13, Munich 81377, Germany

1. Chemical Synthesis

Unless noted otherwise, all reactions were performed using oven dried glassware under an atmosphere of nitrogen. Molsieve-dried solvents were used from *Sigma Aldrich* and chemicals were bought from *Sigma Aldrich*, *TCl*, *Carbolution* and *Carbosynth*. Isotopically labeled trimethylamino glycine was obtained from *Eurisotop*. For extraction and chromatography purposes, technical grade solvents were distilled prior to their usage. Reaction controls were performed using TLC-Plates from *Merck* (*Merck 60 F₂₅₄*), flash column chromatography purifications were performed on *Merck Geduran Si 60* (40-63 μ M). Visualization of the TLC plates was achieved through UV-absorption or through staining with *Hanessian's stain*. NMR spectra were recorded in deuterated solvents on *Varian VXR400S*, *Varian Inova 400*, *Bruker AMX 600*, *Bruker Ascend 400* and *Bruker Avance III HD*. HR-ESI-MS spectra were obtained from a *Thermo Finnigan LTQ FT-ICR*. IR-measurements were performed on a *Perkin Elmer Spectrum BX FT-IR* spectrometer with a diamond-ATR (*Attenuated Total Reflection*) unit. HPLC purifications were performed on a *Waters Breeze* system (2487 dual array detector, 1525 binary HPLC pump) using a *Nucleosil VP 250/10 C18* column from *Macherey Nagel*. HPLC-grade MeCN was purchased from *VWR*. For HPLC purifications of compounds **1a/b**, **9a/b** and **10a/b** a buffer system of 0.25 mM ammonium formate in H₂O, pH = 4.3 (referred to as buffer A) and 0.25 mM ammonium formate in 80% MeCN/H₂O (referred to as buffer B) was used.

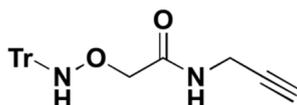
Synthesis of Hydroxylamine **1** and internal standards **9a/b** and **10a/b**

(*N*-Tritylaminoxy)acetic acid (**6**)



(*N*-Tritylaminoxy)acetic acid was synthesized according to Kojima *et al.*¹

N-(prop-2-ene-1-yl)-2-((tritylamino)oxy)acetamide (**7**)

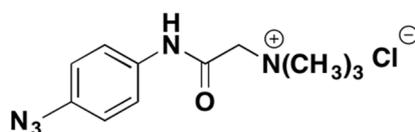


N-trityl protected aminoxyacetic acid **6** (2.50 g, 7.50 mmol, 1.0 eq) was suspended in DCM (40 mL) and was subsequently charged with TBTU (2.89 g, 9.00 mmol, 1.2 eq), DIPEA (1.60 mL, 9.00 mmol, 1.2 eq) and propargylamine (1.40 mL, 22.6 mmol, 3.0 eq). The suspension was stirred at rt, whereupon

after 15 hours a clear yellowish solution was formed. The mixture was diluted with EtOAc (300 mL), the organic phase was washed with NH₄Cl (300 mL) and NaHCO₃ (300 mL) and then dried over Na₂SO₄. Volatiles were finally removed *in vacuo* and the crude mixture was purified via column chromatography (10% EtOAc --> 40% EtOAc/Hex). **7** (2.57g, 6.93 mmol, 92%) was yielded as a colorless solid.

¹H-NMR (300 MHz, CDCl₃): δ/ppm = 7.37–7.22 (m, 15H, (C₆H₅)₃C), 6.59 (s, 1H, (C₆H₅)₃C-NH-O), 5.81 (bs, 1H, O=C-NH), 4.25 (s, 2H, O-CH₂C=O), 3.85 (dd, ³J = 5.5 Hz, ⁴J = 2.6 Hz, 2H, HN-CH₂), 2.15 (t, ⁴J = 2.6 Hz, 1H, C≡C-H). **¹³C-NMR (75 MHz, CDCl₃):** δ/ppm = 169.1 (C=O), 143.9 (3C, 3 × O-NH-C-C), 129.0 (6C, C_{Ar}-H), 128.2 (6C, C_{Ar}-H), 127.4 (3C, C_{tert}-H), 79.3 (C≡C-H), 74.6 (C(C₆H₅)₃), 73.4 (O-CH₂), 71.7 (C≡C-H), 28.8 (NH-CH₂). **HRMS (ESI⁺):** calc. for C₂₄H₂₂N₂NaO₂ [M+Na]⁺: 393.1573; found: 393.1571. **IR (ATR):** $\tilde{\nu}$ (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). **Melting Range:** 157 - 158 °C.

2-((4-Azidophenyl)amino)-N,N,N-trimethyl-2-oxoethaneaminium chloride (4a)

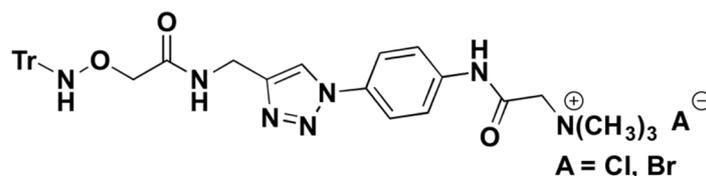


Betaine **3a** (0.30 g, 2.56 mmol, 1.0 eq) was first dried on high vac at 180 °C for 20 minutes. After cooling to rt, the colorless solid was suspended in DMF (25 mL). 4-Azidoanilin hydrochloride **2** (0.54g, 3.17 mmol, 1.2 eq), TBTU (0.99 g, 3.07 mmol, 1.2 eq) and DIPEA (1.10 mL, 6.32 mmol, 2.4 eq) were added whereupon a yellow brownish solution formed gradually. After stirring for one hour at rt all solids were dissolved and the reaction was further stirred at rt over night. DMF was then removed *in vacuo* and the crude mixture was purified by column chromatography (DCM/MeOH/H₂O/7N NH₃ in methanol = 90:10:0.6:0.6) and **4a** was yielded as the corresponding phenyl triazolate salt. The salt was then redissolved in H₂O (50 mL) and was acidified to pH = 1. The aqueous phase was extracted with Et₂O until TLC analysis of the organic phase fractions showed no UV absorption. The aqueous layer was subsequently neutralized with conc. NH₃ and the chloride salt of **4a** (0.62 g, 2.30 mmol, 90%) was yielded as a brownish powder.

¹H-NMR (300 MHz, dmsO d⁶): δ/ppm = 11.22 (s, 1H, NH), 7.69 (d, ³J = 8.9 Hz, 2H, CH=C-NH), 7.12 (d, ³J = 8.4 Hz, 2H, CH=C-N₃), 4.42 (s, 2H, CH₂), 3.30 (s, 9H, N(CH₃)₃). **¹³C-NMR (101 MHz, dmsO d⁶):** δ/ppm = 162.0 (C=O), 135.0 (NH-C=CH), 134.9 (N₃-C=CH), 121.2 (2C, NH-C=CH), 119.5 (2C, N₃-C=CH), 64.3 (CH₂),

53.4 (3C, N(CH₃)₃). **HRMS (ESI⁺)**: calc. for C₁₁H₁₆N₅O⁺ [M⁺]: 234.1349; found: 234.1348. **IR (ATR)**: $\tilde{\nu}$ (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). **Melting Range**: 144 - 146 °C.

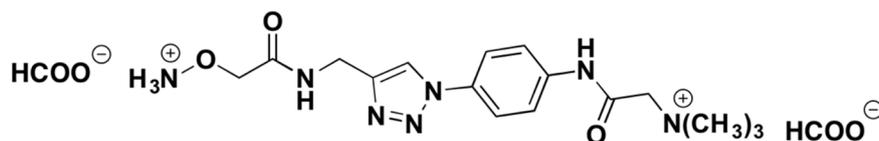
***N,N,N*-trimethyl-2-oxo-2-((4-(4-((2-((tritylamino)oxy)acetamido)methyl)-1*H*-1,2,3-triazole-1-yl)phenyl)amino)ethanaminium chloride/bromide (8a)**



First, a mixture of DCM and H₂O (à 5 mL) was freeze-pump-thaw degassed (3x) and then azide **4a** (0.18 g, 0.65 mmol, 1.0 eq), alkyne **7** (0.24 g, 0.65 mmol, 1.0 eq) and CuBr•SMe₂ (40 mg, 0.20 mmol, 0.3 eq) were added. The suspension was stirred vigorously over night at rt whereupon a colorless emulsion formed. The mixture was then concentrated under reduced pressure and purified by column chromatography using a short plug of silica (DCM/MeOH/H₂O/7N NH₃ in methanol = 80:20:0.6:0.6). **8a** was yielded as a slightly yellow brownish solid (0.32 g, 0.50 mmol, 77%).

¹H-NMR (300 MHz, dmso d⁶): δ /ppm = 11.67 (s, 1H, NH-C₆H₄), 8.55 (s, 1H, CH₂-C=CH-N), 8.34 (s, 1H, Ph₃C-NH), 8.32 (t, ³J = 5.8 Hz, 1H, O=C-NH-CH₂), 7.91–7.84 (m, 4H, C₆H₄), 7.34–7.19 (m, 15H, C(C₆H₅)₃), 4.53 (s, 2H, (CH₂-N(CH₃)₃), 4.45 (d, ³J = 5.8, 2H, NH-CH₂), 3.85 (s, 2H, N-O-CH₂), 3.33 (s, 9H, N(CH₃)₃). **¹³C-NMR (101 MHz, dmso d⁶)**: δ /ppm = 169.7 (O=C-NH-CH₂), 162.4 (O=C-CH₂-N), 146.0 (CH₂-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-H), 121.0 (CH₂-C=CH-N), 120.5 (4C, N-C=CH-CH=C-N), 73.7 (C(C₆H₆)₃), 73.2 (O-CH₂), 64.4 (CH₂-N(CH₃)₃), 53.4 (N(CH₃)₃), 33.8 (NH-CH₂). **HRMS (ESI⁺)**: calc. for C₃₅H₃₈N₇O₃⁺ [M⁺]: 604.3031; found: 604.3026. **IR (ATR)**: $\tilde{\nu}$ (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). **Melting Range**: 142 - 152 °C.

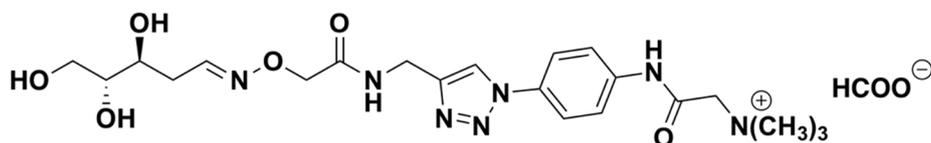
2-((4-(4-((2-(Aminoxy)acetoamido)methyl)-1H-1,2,3-triazole-1-yl)phenyl-amino)-N,N,N-trimethyl-2-oxoethanaminium formate (1a)



Trityl protected compound **8a** (0.24g, 0.38 mmol) was dissolved in DCM (6 mL) and then 6M HCl (6 mL) was added. The mixture was rigorously stirred at rt for one hour until a phase separation was visible. The aqueous phase was then extracted with DCM (5 x 5mL) until TLC analysis of the organic phase fractions showed no UV absorption. The pH was adjusted to 9-10 using 2M NH₃ and the aqueous phase was removed *in vacuo*. 75 mg of **1a** were then further purified by preparative HPLC (0% --> 20% buffer B) and yielded 23 mg (0.05 mmol, 26%) of **1a** as the colorless formate salt.

¹H-NMR (400 MHz, D₂O): δ/ppm = 8.40 (s, 1H, CH₂-C=CH-N), 8.21 (s, 1H, HCOO), 7.58 (d, ³J = 9.2 Hz, 2H, CH-CH=C-N₃), 7.52 (d, ³J = 9.2 Hz, 2H, CH-CH=C-NH), 4.53 (s, 2H, N-O-CH₂), 4.27 (s, 2H, NH-CH₂), 4.22 (s, 2H, CH₂-N(CH₃)₃), 3.36 (s, 9H, N(CH₃)₃). **¹³C-NMR (101 MHz, D₂O):** δ/ppm = 169.7 (O=C-NH-CH₂), 162.4 (O=C-CH₂-N), 146.0 (CH₂-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-H), 121.0 (CH₂-C=CH-N), 120.5 (4C, N-C=CH-CH=C-N), 73.7 (C(C₆H₆)₃), 73.2 (O-CH₂), 64.4 (CH₂-N(CH₃)₃), 53.4 (N(CH₃)₃), 33.8 (NH-CH₂). **HRMS (ESI⁺):** calc. for C₁₆H₂₄N₇O₃⁺ [M⁺]: 362.1935; found: 362.1935. **IR (ATR):** $\tilde{\nu}$ (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).

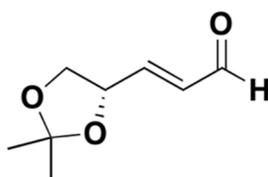
N,N,N-Trimethyl-2-oxo-2-((4-(4-(((3S,4R)-3,4,5-trihydroxypentylidene)amino)oxy)acet-amido)-methyl)-1H-1,2,3-triazol-1-yl)phenyl)amino)ethanaminium formate (9a)



1a (50.0 mg, 0.12 mmol, 1.0 eq) and 2'-desoxyribose (182 mg, 1.36 mmol, 11.8 eq) were dissolved in H₂O (2.7 mL) and incubated over night at 30 °C and 1400 rpm in a *Eppendorf Comfort* thermomixer. The mixture was filtered over a 0.2 μm syringe filter and was subsequently purified by HPLC twice (0 → 15% buffer B). Pure product **9a** (9.1 mg, 17 μmol, 15%) was obtained as a colorless foam. The compound was present as a mixture of *E/Z* isomers in aqueous solution that were not assigned.

¹H-NMR (600 MHz, D₂O): δ/ppm = 8.46 (s, 1H, HCOO), 8.34 (s, 1H, CH₂-C=CH-N), 7.79 (d, J=9.0 Hz, 2H, CH-CH=C-N₃), 7.74 - 7.71 (m, 8H, CH-CH=C-NH, C1'-H^A), 7.08 (t, ³J = 5.4 Hz, 1H, C1'-H^B), 4.67 (s, 2H, N-O-CH₂^B), 4.63 (s, 2H, NH-CH₂), 4.62 (s, 2H, N-O-CH₂^A), 4.35 (s, 2H, CH₂-N(CH₃)₃), 3.92–3.87 (m, 1H, C3'-H^B), 3.85 - 3.80 (m, 1H, C3'-H^A), 3.78–3.69 (m, 1H, C5'-H), 3.66–3.53 (m, 2H, C5'-H, C4'-H), 3.42 (s, 9H, N(CH₃)₃), 2.79–2.69 (m, 2H, C2'-H^B), 2.58 - 2.54 (m, 1H, C2'-H^A), 2.41 - 2.35 (m, 1H, C2'-H^A). **¹³C-NMR (150 MHz, D₂O):** δ/ppm = 172.4 (O=C-NH-CH₂), 170.9 (HCOO), 162.7 (O=C-CH₂-N), 153.5 (C1'^A), 153.1 (C1'^B), 145.1 (CH₂-C=C), 136.8 (N-C=CH-CH), 133.5 (N-C=CH-CH), 122.5 (2C, CH=C-NH), 122.3 (CH₂-C=CH-N), 121.9 (2C, CH=C-N₃), 74.2 (C4'), 74.0 (C4'), 71.7 (N-O-CH₂^B), 71.5 (N-O-CH₂^A), 69.0 (C3'^A), 68.8 (C3'^B), 65.1 (CH₂-N(CH₃)₃), 62.3 (C5'), 54.3 (N(CH₃)₃), 34.1 (NH-CH₂), 32.4 (C2'^A), 29.2 (C2'^B). **HRMS (ESI⁺):** calc. for C₂₁H₃₂N₇O₆⁺ [M]⁺: 478.2409; found: 478.2404.

(S,E)-3-(2,2-Dimethyl-1,3-dioxolan-4-yl)acrylaldehyde (12)



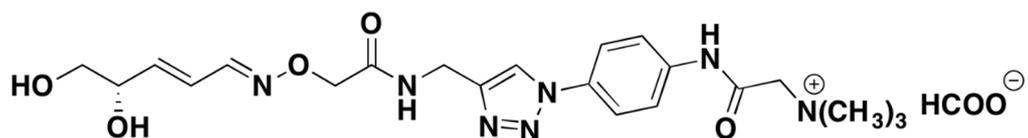
Methyl (2E)-3-[(4S)-2,2-dimethyl-1,3-dioxolan-4-yl]prop-2-enoate **11** (0.20 g, 1.08 mmol, 1.0 eq) was dissolved in DCM (2.0 mL) and cooled to -78 °C. DIBAL-H (2.20 mL, 2M in toluene, 2.1 eq) was added and the yellowish mixture was slowly warmed to rt. After 90 minutes, DCM (5.0 mL) and H₂O (4.0 mL) and NaOH (2M, 2.0 mL) were added. After stirring for an additional hour at rt, the organic phase was separated from the aqueous and dried over Na₂SO₄. Volatiles were removed under reduced pressure and the allylic alcohol was isolated in quantitative yield and used without further purification.

¹H-NMR (400 MHz, CDCl₃): δ/ppm = 5.88 (dt, ³J = 15.4 Hz, ⁴J = 5.0 Hz, 1H, 5'-H) 5.65 (dd, ³J = 15.6 Hz, ⁴J = 7.6 Hz, 1H, 5'-H), 4.47 (q, ³J = 7.3 Hz, 1H, 4'-H), 4.08 (d, ³J = 5.1 Hz, 2H, 1'-H), 4.30 (dd, ³J = 8.2 Hz, ⁴J = 6.1 Hz, 1H, 3'-H), 3.53 (t, ³J = 7.9 Hz, 1H, 2'-H), 2.34 (br s, 1H, CH₂-OH), 1.36 (s, 3H, O-C(CH₃)(CH₃)-O), 1.32 (s, 3H, O-C(CH₃)(CH₃)-O).

The allylic alcohol was dissolved in DCM (2.0 mL), cooled to 0 °C and was charged with *Dess-Martin-periodinan* (0.45 g, 1.08 mmol, 1.0 eq). The milky suspension was slowly warmed to rt and stirred over night. After the addition of saturated Na₂SO₄ (10 mL) and a solution of Na₂S₂O₃ (171 mg, dissolved in 10 mL H₂O), the mixture was extracted with DCM (3 x 15 mL) and dried over Na₂SO₄. Organic solvents were removed *in vacuo* and the crude mixture was purified via column chromatography (2.5% MeOH/DCM). Aldehyde **12** (80 mg, 0.51 mmol, 47%) was isolated as a colorless oil.

¹H-NMR (400 MHz, CD₂Cl₂): δ/ppm = 9.50 (d, ³J = 7.6 Hz, 1H, 1'-CHO), 6.70 (dd, ³J = 15.6 Hz, ⁴J = 5.3 Hz, 1H, 3'-H), 6.23 (dt, ³J = 15.6 Hz, ⁴J = 5.8 Hz, 1H, 2'-H), 4.73 (q, ³J = 6.8 Hz, 1H, 4'-H), 4.18 (dd, ³J = 8.4 Hz, ⁴J = 6.8 Hz, 1H, 5'-H), 3.67 (dd, ³J = 8.4 Hz, ⁴J = 6.8 Hz, 1H, 5'-H), 1.39 (s, 3H, O-C(CH₃)(CH₃)-O), 1.35 O-C(CH₃)(CH₃)-O). **¹³C-NMR (101 MHz, CD₂Cl₂):** δ/ppm = 193.0 (-CHO), 153.4 (3'-C), 132.1 (2'-C), 110.3 (C_{quart}), 74.9 (4'-C), 68.7 (5'-C), 26.2 (O-C(CH₃)(CH₃)-O), 25.4 (O-C(CH₃)(CH₃)-O). **HRMS (EI):** calc. for C₈H₁₁O₃⁺ [M - H]⁺: 155.0708; found: 155.0707.

2-((4-(4-(((1E,2E)-3-((S)-2,2-Dimethyl-1,3-dioxolane-4-yl)allylidene)-amino)oxy)acetamido)methyl)-1H-1,2,3-triazole-1-yl)phenyl)amino)-N,N,N-trimethyl-2-oxoethane-1-aminium formate (10a)



Aldehyde **12** (20 mg, 0.13 mmol, 9.0 eq) and hydroxylamine **1a** were dissolved in a 1:1 mixture of H₂O and CHCl₃ (à 2.5 mL) and stirred at rt. The course of the reaction was monitored by HPLC (0 --> 30% buffer B) whereas it was determined that after one hour the reaction was complete. The aqueous phase was then washed with DCM (3 x 10 mL) and concentrated *in vacuo*. **13a** (5.10 mg, 9.50 μmol, 68%) was yielded as a brownish viscous oil that was used without further purification.

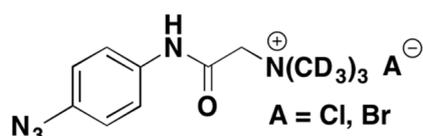
¹H-NMR (400 MHz, D₂O): δ/ppm = 8.31 (s, 1H, HCOO), 8.12 (s, 1H, CH₂-C=CH-N), 7.84 (d, 1H, 1'-H), 7.59 – 7.53 (m, 4H, CH-CH=C-N₃, CH-CH=C-NH), 6.20 – 6.03 (m, 2H, 2' + 3'-H's), 4.54 – 4.52 (m, 1H, 4'-H), 4.50 (s, 2H, N-O-CH₂), 4.47 (s, 2H, NH-CH₂), 4.20 (s, 2H, CH₂-N(CH₃)₃), 4.02 – 3.98 (m, 1H, 5'-H), 3.45 – 3.50 (m, 1H, 5'-H), 3.27 (s, 9H, CH₂-N(CH₃)₃), 1.26 (s, 3H, O-C(CH₃)(CH₃)-O), 1.24 (s, 3H, O-C(CH₃)(CH₃)-O). **HRMS (ESI⁺):** calc. for C₂₄H₃₄N₇O₅⁺ [M⁺]: 500.2616; found: 500.2617.

For the deprotection of the acetone, **13a** (4.00 mg, 7.50 μmol, 1.0 eq) was dissolved in MeOH and PTSA • H₂O (1.40 mg, 7.50 μmol, 1.0 eq) was added. The mixture was incubated in a *Eppendorf comfort* thermomixer (1300 rpm, 25 °C) over night and the solvent was removed *in vacuo* by lyophilization. The crude product was finally purified by semi preparative HPLC (0 --> 35% buffer B in 45 minutes) and pure **10a** was yielded as a colorless foam.

¹H-NMR (400 MHz, D₂O): δ/ppm = 8.53 (s, 1H, HCOO), 8.30 (s, 1H, CH₂-C=CH-N), 8.01 (s, d, ³J = 8.9 Hz, 1H, 1'-H), 7.78 – 7.70 (m, 4H, CH-CH=C-N₃, CH-CH=C-NH), 6.31 – 6.32 (m, 2H, 2' + 3'-H's), 4.63 (s, 2H, N-O-CH₂), 4.62 (s, 2H, NH-CH₂), 4.35 – 4.32 (m, 3H, CH₂-N(CH₃)₃ + 4'-H), 3.61 (dd, ¹J = 11.7 Hz, ³J = 4.4, 1H,

5'-H), 3.51 (dd, $^1J = 11.7$ Hz, $^3J = 6.5$, 1H, 5'-H), 3.22 (s, 9H, $\text{CH}_2\text{-N}(\text{CH}_3)_3$). $^{13}\text{C-NMR}$ (101 MHz, D_2O): $\delta/\text{ppm} = 172.3$ ($\text{O}=\text{C-NH-CH}_2$), 170.9 (HCOO), 162.7 ($\text{O}=\text{C-CH}_2\text{-N}$), 153.7 ($1'\text{-C}$), 144.5 ($\text{CH}_2\text{-C}=\text{C}$), 142.8 ($3'\text{-C}$), 136.9 ($\text{N-C}=\text{CH-CH}$), 133.6 ($\text{N-C}=\text{CH-CH}$), 123.0 ($2'\text{-C}$), 122.5 (2C, $\text{CH}=\text{C-NH}$), 122.3 ($\text{CH}_2\text{-C}=\text{CH-N}$), 122.0 (2C, $\text{CH}=\text{C-N}_3$), 72.0 (N-O-CH_2) 71.6 ($4'\text{-C}$), 64.4 ($5'\text{-C}$), 65.1 ($\text{CH}_2\text{-N}(\text{CH}_3)_3$), 54.3 ($\text{N}(\text{CH}_3)_3$), 34.1 (NH-CH_2). **HRMS (ESI⁺)**: calc. for $\text{C}_{21}\text{H}_{30}\text{N}_7\text{O}_5^+$ [M]⁺: 460.2303; found: 460.2305.

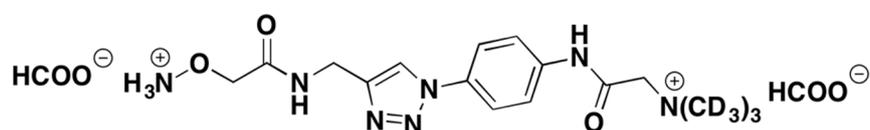
2-((4-Azidophenyl)amino)-*N,N,N*-tri(methyl-*d*₃)-2-oxoethanaminium chloride (4b)



4b was synthesized analogously to **4a** except that [*d*₁₁]-betaine (98% deuterium, *Euriso-Top GmbH*) was used to introduce isotopic labels. Deuterium labels from the methylene group were not stable under the reaction conditions and a complete D/H exchange was observed. Thus, a [*d*₉]-labeled product was obtained.

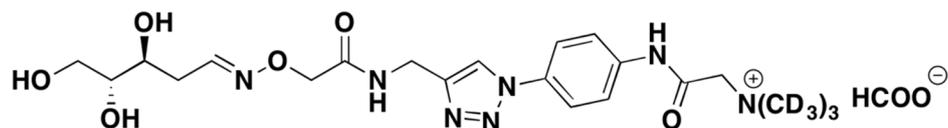
$^1\text{H-NMR}$ (600 MHz, D_2O): $\delta/\text{ppm} = 7.39$ (d, $^3J=8.6$, 2H, $\text{CH-CH}=\text{C-NH}$), 7.04 (d, $^3J=8.5$, 2H, $\text{CH-CH}=\text{C-N}_3$), 4.18 (s, 2H, CH_2). $^{13}\text{C-NMR}$ (150 MHz, D_2O , ppm): $\delta/\text{ppm} = 162.7$ (C=O), 137.5 ($\text{NH-C}=\text{CH}$), 132.5 ($\text{N}_3\text{-C}=\text{CH}$), 123.5 (2C, $\text{NH-C}=\text{CH}$), 119.6 (2C, $\text{N}_3\text{-C}=\text{CH}$), 65.0 (CH_2). **HRMS (ESI⁺)**: calc. for $\text{C}_{11}\text{H}_7\text{D}_9\text{N}_5\text{O}^+$ [M]⁺: 243.1914; found: 243.1916.

2-((4-(4-((2-(Aminoxy)acetoamido)methyl)-1*H*-1,2,3-triazole-1-yl)phenyl-amino)-*N,N,N*-tri(methyl-*d*₃)-2-oxoethanaminium formate (1b)



Isotopologue **1b** was synthesized according to **1a** with the slight modification that the trityl protected intermediate **8b** was not isolated and deprotected without further purification.

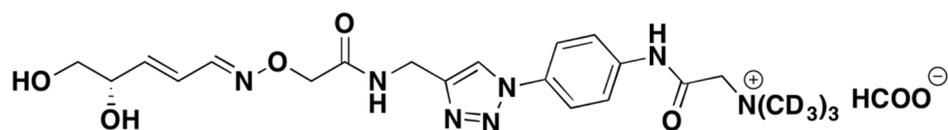
***N,N,N*-Tri(methyl-*d*₃)-2-oxo-2-((4-(4-((2-(((3*S*,4*R*)-3,4,5-trihydroxypentylidene)amino)oxy)acetamido)-methyl)-1*H*-1,2,3-triazol-1-yl)phenyl)amino)ethanaminium formate (9b)**



Internal standard **9b** was synthesized analogously to **9a**, whereupon a mixture of (*E*)/(*Z*)-isomers was obtained (depicted as **A** and **B**).

¹H-NMR (600 MHz, D₂O): δ/ppm = 8.46 (s, 2H, HCOO), 8.35 (s, 1H, CH₂-C=CH-N), 7.81 (d, ³J = 7.5 Hz, 2H, CH-CH=C-N₃), 7.75 - 7.73 (m, 8H, CH-CH=C-NH, C1'-H^A), 7.08 (t, ³J = 5.4 Hz, 0.1H, C1'-H^B), 4.67 (s, 2H, N-O-CH₂^B), 4.64 (s, 2H, NH-CH₂), 4.62 (s, 2H, N-O-CH₂^A), 4.34 (s, 2H, CH₂-N(CH₃)₃), 3.93–3.87 (m, 1H, C3'-H^B), 3.86–3.79 (m, 1H, 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, 2H, C2'-H₂^B), 2.59 - 2.54 (m, 1H, C2'-H^A), 2.43–2.34 (m, 1H, C2'-H^A). **¹³C-NMR (150 MHz, D₂O, ppm):** δ/ppm = 172.4 (O=C-NH-CH₂), 170.9 (HCOO), 162.8 (O=C-CH₂-N), 153.5 (C1'^A), 153.1 (C1'^B), 145.1 (CH₂-C=C), 136.8 (N-C=CH-CH), 133.6 (N-C=CH-CH), 122.6 (2C, CH=C-NH), 122.4 (CH₂-C=CH-N), 122.0 (2C, CH=C-N₃), 74.2 (C4'), 74.0 (C4'), 71.7 (N-O-CH₂^B), 71.5 (N-O-CH₂^A), 69.0 (C3'^A), 68.8 (C3'^B), 64.9 (CH₂-N(CD₃)₃), 62.3 (C5'), 53.3 (N(CD₃)₃), 34.1 (NH-CH₂), 32.4 (C2'^A), 29.2 (C2'^B). **HRMS (ESI⁺):** calc. for C₂₁H₂₃D₉N₇O₆⁺ [M]⁺: 487.2973; found: 487.2967.

2-((4-(4-((2-(((1*E*,2*E*)-3-((*S*)-2,2-Dimethyl-1,3-dioxolane-4-yl)allylidene)-amino)oxy)acetamido)-methyl)-1*H*-1,2,3-triazole-1-yl)phenyl)amino)-*N,N,N*-tri(methyl-*d*₃)-2-oxoethane-1-aminium formate (10b)



10b was synthesized according to **10a**.

¹H-NMR (800 MHz, D₂O): δ/ppm = 8.47 (s, 1H, HCOO), 8.32 (s, 2H, CH₂-C=CH-N), 8.02 (d, ³J = 9.4 Hz, 1H, 1'-H), 7.80 - 7.71 (m, 4H, CH-CH=C-N₃, CH-CH=C-NH), 6.34 - 6.26 (m, 2H, 2' + 3'-H's), 4.65 (s, 2H, N-O-CH₂), 4.63 (s, 2H, NH-CH₂), 4.35 - 4.32 (m, 3H, CH₂-N(CH₃)₃ + 4'-H), 3.63 (dd, ¹J = 11.7 Hz, ³J = 4.4, 1H, 5'-H), 3.53 (dd, ¹J = 11.7 Hz, ³J = 6.5, 1H, 5'-H). **¹³C-NMR (150 MHz, D₂O):** δ/ppm = 172.3 (O=C-NH-CH₂), 170.9 (HCOO), 162.8 (O=C-CH₂-N), 153.7 (1'-C), 145.2 (CH₂-C=C), 142.8 (3'-C), 136.9 (N-C=CH-CH), 133.6 (N-C=CH-CH), 123.0 (2'-C), 122.5 (2C, CH=C-NH), 122.3 (CH₂-C=CH-N), 122.0 (2C, CH=C-N₃), 72.1 (N-O-

CH₂) 71.6 (4'-C), 64.4 (5'-C), 65.1 (CH₂-N(CH₃)₃), 53.3 (N(CD₃)₃), 34.1 (NH-CH₂). HRMS (ESI⁺): calc. for C₂₁H₂₁D₉N₇O₅⁺ [M]⁺: 469.2868; found: 469.2874.

2. Cell Culture and Transfection Procedures

DMEM high glucose containing 10% FBS (*PAN Biotech*), 2 mM L-glutamine, 100 U/mL penicillin, 100 µg/mL streptomycin, 1x MEM Non-essential Amino Acid Solution and 0.1 mM β-mercaptoethanol (*Sigma Aldrich*) was used as basal medium for mESC cultures. The mESC lines were maintained in naïve state on gelatin coated plates by supplementing basal medium with 1000 U/mL LIF (*ORF Genetics*), 3.0 µM GSK3 inhibitor CHIR99021 and 1.0 µM Mek inhibitor PD0325901 (*2i*; *Selleckchem*). Metabolic labelling experiments with isotope-labeled nucleosides were performed by plating mESCs in priming conditions consisting of basal mESC medium supplemented with 1000 U/mL LIF. Labeled nucleosides (*B.A.C.H. UG*) were added to the culture medium at the following concentrations: dG* [¹³C₁₀¹⁵N₅] at a concentration of 100 µM for three days, followed by treatment with 200 µM labeled dG* for two days; dC* [¹³C₉¹⁵N₃] and dT* [¹³C₁₀¹⁵N₂] was used at a concentration of 100 µM for five days. J1 wild type mESCs were obtained from the 129S4/SvJae strain and Dnmt TKO J1 mESCs were described in *Li et al.* and *Tsumura et al.*²⁻³ The Tdg wild type and the Tdg KO cell lines were reported in *Cortazar et al.*⁴

HEK293T cells were cultured in RPMI-1640 medium containing 10% FBS, 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin. ENC1 neural stem cells were cultured as previously described.⁵

2.1 Ape1 Inhibition

a) For treatment with the Ape1 inhibitor CRT0044876 (*Sigma-Aldrich*), mESCs were cultured for three days in priming conditions, passaged and re-plated in priming medium containing 100 µM CRT0044876 for 48 or 60 h. After 24 h fresh medium containing the inhibitor was added and DMSO was used as vehicle control. **b)** Wild type and Dnmt TKO mESCs were cultured for three days in priming medium containing 100 µM dC*, passaged and re-plated under priming conditions with the addition of 100 µM CRT0044876 (*Sigma-Aldrich*) and 100 µM dC* for 48 h. After 24 h fresh medium containing the inhibitor and 100 µM dC* was added. DMSO was used as vehicle control.

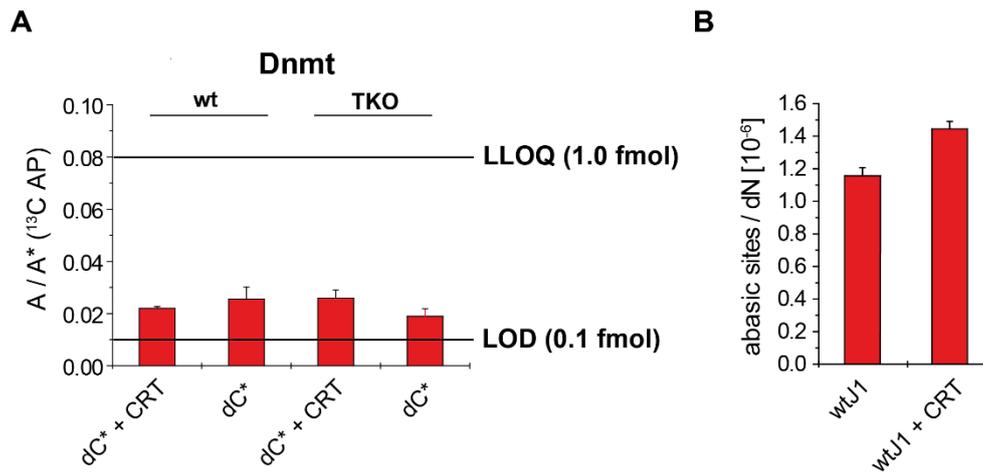


Figure S1: AP-sites after inhibition of Ape1. **(A)** A/A* ratios of labeled AP-sites in wild type and Dnmt TKO mESCs were over the LOD but below the LLOQ. **(B)** Quantification of unlabeled AP-sites of the dC* feeding experiment show an increase by treatment with the inhibitor CRT0044876.

2.2 Ape1 Knockdown

mESCs were cultured under priming conditions for three days, passaged and transfected with Apex1 Mission[®] esiRNA (purchased from *Sigma-Aldrich*) at 50 nM using Lipofectamine 2000 according to the manufacturer. The medium was changed after 24 h and cells were harvested 48 h post transfection. For qPCR analysis a second independent knockdown experiment was performed.

2.3 Cell lysis and DNA isolation

Isolation of genomic DNA was achieved using the *QIAamp DNA Mini Kit* from *Qiagen*. All mESC samples were washed with PBS (*Sigma*) and directly lysed in the plates by adding G2 buffer containing 400 μM of 2,6-di-tert-butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM). DNA was sheered by bead milling in a microfuge tube using one 5 mm diameter stainless steel bead per tube and MM400 bead mill (*Retsch*) at 30 Hz for one minute and subsequent centrifugation at 15000 rpm for ten minutes. Depending on the amount of genomic DNA to isolate, the cell lysate was treated with proteinase K (25 μL for *genomic tips 20G* or 100 μL for *genomic tips 100G*) and RNase A (2.0 $\mu\text{L}/20\text{G}$, 10 $\mu\text{L}/100\text{G}$) at 50 $^{\circ}\text{C}$ for one hour. After 30 minutes, additional RNase A (2.0 μL or 10 μL , respectively) was added to the mixture. *Genomic tip* columns were then equilibrated with QBT loading buffer (1.0 mL/20G or 4.0 mL/100G) and then the lysate, which was vortexed for one minute, was applied on the columns. After the entire liquid had entered the column, washing steps were carried out with QC buffer (2.0 mL/20G or 2x7.5 mL/100G) and the genomic DNA was finally eluted with QF buffer (2.0 mL/20G or 5.0 mL/100G) supplemented with 400 μM BHT. Precipitation was then achieved through the

addition of *i*-PrOH (1.4 mL/20G or 3.5 mL/100G, 70% Vol) and the resulting genomic DNA pellet was centrifuged (15 minutes, 6000 g, 4 °C). The supernatant was discarded, washing steps were carried out using 70% EtOH (5.0 mL, 15 minutes, 6000 g, 4 °C). Finally, the pure DNA pellet was resuspended in 1.0 mL 70% EtOH and centrifuged (10 minutes, 15000 rpm, 4 °C). Next, the supernatant was removed and the pellet was re-dissolved in ddH₂O (50 - 100 µL) with 20 µM BHT. The concentration was finally determined with a *NanoDrop* (ND 1000, Peqlab).

2.4 Artificial Generation of AP-Sites

Salmon sperm DNA (à 5.0 µg, *Sigma Aldrich*) was heated in a 10 mM sodium citrate buffer (pH = 5.0) for three and ten minutes at 70 °C, respectively. The mixture was subsequently neutralized to pH = 7.5 using 0.2 mM NaOH and then derivatized and quantified as described before. The data shows that harsh conditions need to be applied in order to generate AP-sites artificially.

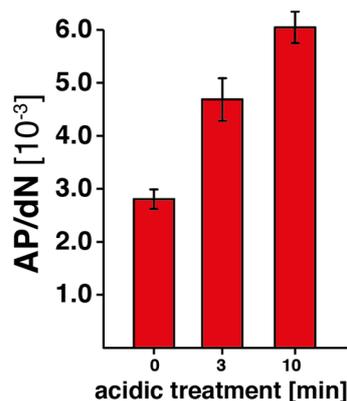


Figure S2: Artificial generation of AP-sites under harsh conditions.

2.5 Re-Isolation of wtJ1 mESC DNA

Reisolation of genomic DNA was performed by adding 30 µg of DNA to 3200 µL of G2 buffer. The DNA containing lysis buffer was applied to a pre-equilibrated 20G column and DNA isolation was performed as described (see cell lysis and DNA isolation). This experiment shows that the isolation procedure does not introduce β-elimination product artificially.

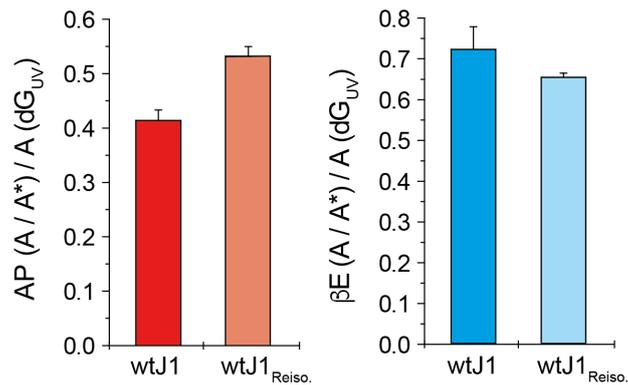


Figure S3: A/A* ratios of AP- and βE-sites normalized by dG. The DNA of wild type mESCs shows similar values of the BER intermediates before and after reisolation.

2.6 Quantification of AP- and βE-Sites in Different Cell Lines

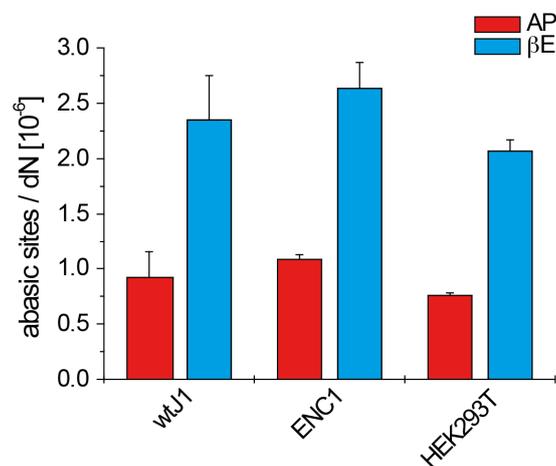


Figure S4: Quantification of AP- and βE-intermediates in wild type mESCs, ENC1 neural stem cells and somatic HEK293T cells shows similar values for the BER intermediates in all three cell lines. The quantification results of one biological replicate is depicted.

2.7 RNA Isolation and Real-Time Quantitative Polymerase Chain Reaction (qPCR)

At 48 h post-transfection, total RNA was isolated with the peqGOLD Total RNA Kit (*PeqLab*) according to the manufacturer's instructions. Prior to the qPCR, 1 μg of RNA was transcribed to 1 μg of cDNA using the iScript cDNA Synthesis Kit (*Bio-Rad*) according to the manufacturer. The final cDNA concentration was 50 ng/μL. To 10 μL of iTaq Universal SYBR Green Supermix (*Bio-Rad*) 2 μL of cDNA were added per reaction. For each primer pair, the forward and the reverse primer were mixed together in equal amounts and diluted with nuclease-free ddH₂O to a final concentration of 1.25 μM. 12 μL of cDNA/iTaq mastermix and 8 μL of primer mix were added per well. Each sample was measured in triplicates on a MasterCycler RealPlex (*Eppendorf*) with the following PCR conditions: Step 1 95 °C (2:00 minutes), Step 2 95 °C (0:15 minutes), Step 3 55 °C (0:15 minutes), Step 4 72 °C (0:20 minutes); Step 2 – 4 were repeated 40 times.

Ape1 forward primer: 5'-GGTCAGCTCCGTCAGACAAA-3'

Ape1 reverse primer: 5'-TCGGAAGGCTTCATCCCAAC-3'

Alpha-tubulin forward primer: 5'-TGTGGATTCTGTGGAAGGCG-3'

Alpha-tubulin reverse primer: 5'-AGCACACATTGCCACATACAAA-3'

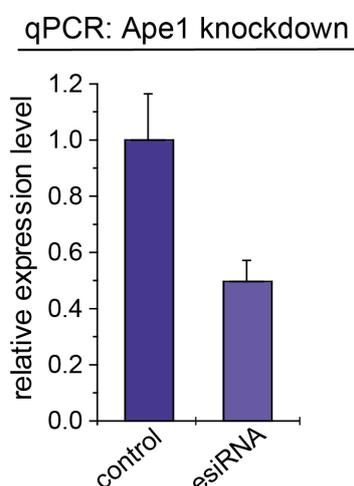


Figure S5: Relative mRNA expression levels show an estimated knockdown efficiency of about 50% using Ape1 (Apex1) esiRNA. As housekeeping reference gene alpha-tubulin was used.

3. Quantification of Abasic Sites

3.1 Derivatization of Genomic DNA with **1a**

Derivatization of abasic sites (5.0 μ g for unlabeled gDNA, 20 μ g for labeled gDNA) with **1** was carried out in a total volume of 20 μ L, whereas the solution was buffered with HEPES (20 mM, pH = 7.5) and Na₂EDTA (0.1 mM). A stock of **1a** in H₂O (23.8 mM) was added to the buffered solution (final concentration of **1a** = 1.5 mM) and the reaction was started by vortexing the mixture for 5 seconds. The gDNA was incubated for 40 minutes at 37 °C/1400 rpm in an *Eppendorf Comfort* thermomixer. The reaction was stopped via addition of 1-naphthylaldehyde (66.7 μ L, 2M in *i*-PrOH) to quench excess of **1a** and incubated again for 10 minutes at 37 °C/1400 rpm. Derivatized DNA was then precipitated via addition of NaOAc (3.3 μ L, 3M), vortexing and incubation at 37 °C/1400 rpm for another 5 minutes. After absolute *i*-PrOH (66.7 μ L) was added, the tubes were inverted several times and then centrifuged (60 minutes, 10 °C, 15000 rpm). The supernatant was removed and washing steps were carried out (1x75% *i*-PrOH, 10 °C, 15000 rpm, 30 minutes; 2x75% cold EtOH, 4 °C, 15000 rpm, 30 minutes), whereas after each washing step the supernatant was carefully removed. The resulting DNA pellet was finally re-dissolved in 35 μ L of ddH₂O and then enzymatically digested to the nucleoside level.

3.2 Enzymatic Digestion of Derivatized Genomic DNA

For enzymatic digestion, genomic DNA (5.0 µg for unlabeled gDNA or 20 µg for labeled gDNA in 35 µL H₂O) was incubated in an aqueous solution containing 480 µM ZnSO₄, 5 U Antarctic phosphatase (*New England BioLabs*) and 42 U nuclease S1 (*Aspergillus oryzae*, *Sigma-Aldrich*) at 37 °C for 3 h. The solution also contained specific amounts of labeled internal standards for accurate quantification of DNA-modifications and the derivatized abasic sites. In the second digestion round, 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, *USB corporation*) in 7.5 µL of a 520 µM [Na]₂-EDTA solution was added and the mixture was incubated for further 3h or overnight at 37 °C (when digesting 20 µg of gDNA). After digestion, the samples were stored at -20 °C and filtered using an *AcroPrep Advance* 96 filter plate 0.2 µm (0.20 µm *Supor*, *Pall Life Sciences*) prior to LC-MS/MS analysis (39 µL injection volume at 4 °C).

3.3 LC-ESI-MS/MS Analysis of DNA Samples

For the LC-MS/MS studies a triple quadrupole mass spectrometer (*Agilent 6490*) and an *Agilent 1290* UHPLC system with UV detector were used. Based on earlier published work⁶⁻¹⁰, a new method was developed that was coupled to the isotope dilution technique, allowing an exact quantification of derivatized abasic sites, all canonical nucleosides and cytosine modifications in one single analytical run. The chromatographical separation was performed on a *Poroshell 120 SB-C8* column (*Agilent*, 2.7 µm, 2.1 mm × 150 mm). Elution buffers were H₂O and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 ml/min at 30 °C. The gradient was: 0 → 5 min; 0 → 3.5% (v/v) MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 13.2 min; 5 → 80% MeCN; 13.2 → 14.8 min; 80% MeCN; 14.8 → 15.3 min; 80 → 0% MeCN; 15.3 → 17 min; 0% MeCN. The eluent up to 1.5 min and after 12.2 min was diverted to waste by a *Valco* valve.

The source-dependent parameters were as follow: gas temperature 80 °C, gas flow 15 l/min (N₂), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 l/min (N₂), capillary voltage 2500 V (positive mode) and -2250 V (negative ion mode), nozzle voltage 500 V, fragmentor voltage 380 V, Δ EMV 500 (positive mode) and 800 (negative mode). Compound-dependent parameters which gave highest intensities during method development are summarized in **Supplementary Table 1**.

Supplementary Table 1. 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 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 1.5–3.8 minutes								
$[^{15}N_2]5cadC$	274.08	wide	158.03	wide	170	5	5	Positive
5cadC	272.09	wide	156.04	wide	170	5	5	Positive
$[^{15}N_2,D_2]5hmdC$	262.12	enhanced	146.07	enhanced	40	27	1	Positive
5hmdC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
$[D_3]5mdC$	245.13	enhanced	129.09	enhanced	30	60	1	Positive
5mdC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
dC	228.12	enhanced	112.05	enhanced	25	5	5	Positive
$[^{13}C_9,^{15}N_3]dC$	240.12	enhanced	119.06	enhanced	25	5	5	Positive
Time segment 3.8–5.8 minutes								
$[D_2]5hmdU$	259.09	wide	216.08	wide	48	7	5	Negative
5hmdU	257.08	wide	214.07	wide	48	7	5	Negative
$[^{15}N_2]5fdU$	257.06	wide	213.05	wide	48	6	5	Negative
5fdU	255.06	wide	212.06	wide	48	6	5	Negative
Time segment 5.8–8.1 minutes								
$[^{15}N_5]8oxodG$	289.08	wide	173.04	wide	90	9	7	Positive
8oxodG	284.10	wide	168.05	wide	90	9	7	Positive
dG	268.10	wide	152.06	wide	75	45	3	Positive
$[^{13}C_{10},^{15}N_5]dG$	283.12	wide	162.06	wide	75	45	3	Positive
$[^{15}N_2]5fdC$	258.09	wide	142.04	wide	50	5	5	Positive
5fdC	256.09	wide	140.05	wide	50	5	5	Positive
dT	243.1	enhanced	127.05	enhanced	35	40	3	Positive
$[^{13}C_{10},^{15}N_2]dT$	255.12	wide	130.07	wide	50	8	5	Positive
Time segment 8.1–12.2 minutes								
1-Naphthyl-Oxim	500.24	wide	472.23	wide	5	19	5	Positive
9b_1	487.30	wide	459.29	wide	38	19	5	Positive
9b_2	487.30	wide	201.18	wide	38	40	5	Positive
$[^{13}C_5]9a_1$	483.26	wide	455.25	wide	38	19	5	Positive
$[^{13}C_5]9a_2$	483.26	wide	192.13	wide	38	40	5	Positive
9a_1	478.24	wide	450.23	wide	38	19	5	Positive
9a_2	478.24	wide	192.13	wide	38	40	5	Positive
10b_1	469.29	wide	441.28	wide	38	19	3	Positive
10b_2	469.29	wide	201.18	wide	38	33	3	Positive
$[^{13}C_5]10a_1$	465.23	wide	437.22	wide	38	19	3	Positive
$[^{13}C_5]10a_2$	465.23	wide	192.13	wide	38	34	3	Positive
10a_1	460.23	wide	432.22	wide	38	19	3	Positive
10a_2	460.23	wide	192.13	wide	38	34	3	Positive
1b	371.25	wide	343.24	wide	5	19	5	Positive
1a	362.19	wide	334.19	wide	5	19	5	Positive

For the chromatographical separation of enzymatically hydrolysed synthetic DNA containing the adduct **9a**, a slightly different gradient was used: 0 → 5 min; 0 → 3.5% (v/v) 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.

The source-dependent parameters were the same as for genomic DNA measurements except for Δ EMV which was set to 300 V. The time segments were 0 - 1.5 minutes; 1.5 - 4 minutes; 4 - 8 minutes and 8 - 12.2 minutes. Fragmentation parameters for **9a/b** and **1a/b** are summarized in **Supplementary Table 2**.

Supplementary Table 2. Compound-dependent LC-MS/MS-parameters for **9a/b** and **1a/b** used for the analysis of a synthetic DNA oligo containing **9a**. 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 the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

Time segment 8 –12.2 minutes								
9b_1	487.30	<i>wide</i>	459.29	<i>wide</i>	38	13	2	Positive
9b_2	487.30	<i>wide</i>	201.18	<i>wide</i>	38	25	2	Positive
9a_1	478.24	<i>wide</i>	450.23	<i>wide</i>	38	13	2	Positive
9a_2	478.24	<i>wide</i>	192.13	<i>wide</i>	38	25	2	Positive
1b	371.25	<i>wide</i>	343.24	<i>wide</i>	5	10	1	Positive
1a	362.19	<i>wide</i>	334.19	<i>wide</i>	5	10	1	Positive

3.4 Method Development

Method validation and data processing were performed as described in earlier published work.⁷ In order to obtain calibration curves, each standard (5–8 standard concentrations) was analysed as technical triplicates and linear regression was applied using Origin® 6.0 (Microcal™). Therefore, the ratio of the area under the curve (A/A^*) of the unlabeled derivatized abasic sites **9a** or **10a**, respectively, to the internal standard (*) was plotted against the ratio of the amount of substance (n/n^*) of the unlabeled derivatized abasic site **9a** and **10a**, respectively, to the internal standard (*) (Figure S3). Calibration functions were calculated without weighing. Acceptable precision (< 20% relative s.d.) and accuracy (80–124%) was achieved. The precision was obtained when A/A^* ratios measured in technical triplicates for each calibration standard had standard deviations <20%. The accuracy was the ratio of the used to the calculated amount of substance in percent for each concentration. To prove the accuracy, we used the respective calibration function for calculation of the substance amount n from A/A^* ration for each calibration standard.

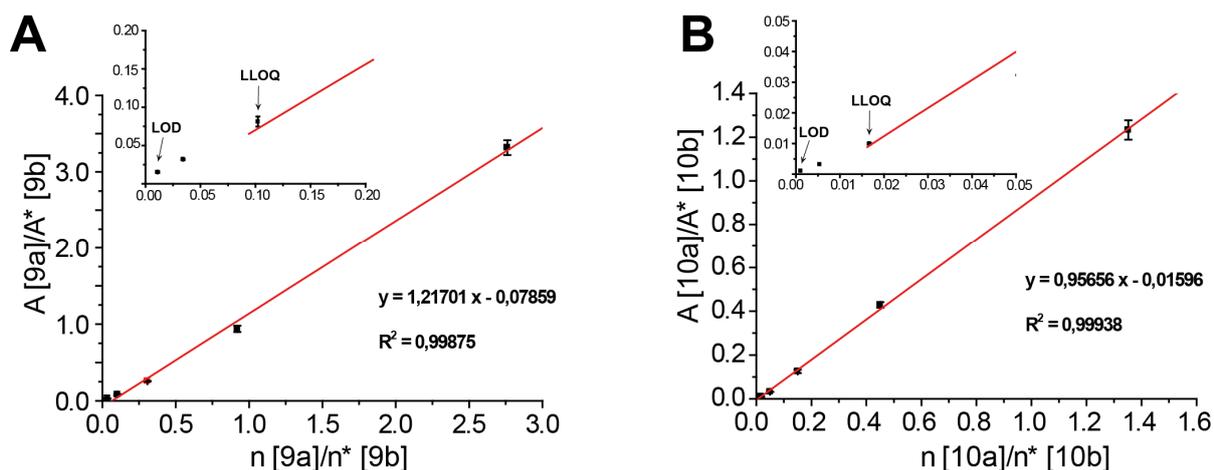


Figure S6: Calibration curves for **(A)** AP- (**9**) and **(B)** β E -site (**10**).

The lower limit of detection was defined as the detected amount that is three times higher compared to the blank response (LOD). The lower limit of quantification (LLOQ) was defined as the lowest concentration fulfilling the requirements of accuracy and precision and achieving an amount higher than LOD. A compilation of LLOQs and LODs is shown in **Supplementary Table 3**.

Supplementary Table 3. Compilation of absolute lower limits of quantification [fmol] (LLOQ and relative LLOQs [per dN] depending on the amount of digested DNA. The relative LLOQs were calculated by generating ratios of the absolute LLOQ [fmol] to the total amounts of nucleosides (N; [fmol]) in the respective amount of DNA [μ g]. The total amounts of nucleosides were obtained by using the average molar mass of 308.91 g mol⁻¹ for the monomeric DNA entity by taking the G-content (21% G) in mESC into account.

	Absolute LOD [fmol]	Absolute LLOQ [fmol]	Relative LLOQ [per dN]	Relative LLOQ [per dN]
DNA amount			5 μg	20 μg
9a	0.11	1.02	6.3E-08	1.57E-8
10a	0.11	1.01	6.3E-08	1.56E-8

3.5 Preparation of a Synthetic 13-mer with Defined Abasic Site

Oligonucleotides (5'-GTA ATG UGC TAG G-3' and 3'-CAT TAG ACG ATC C-5', à 15 nmol, *Metabion*) were incubated in *UDG*-buffer (150 µL, 20 mM Tris-HCl, pH = 8.0, 1 mM DTT, 1.0 mM EDTA, *New England Biolabs*) at 95 °C for 5 minutes and then slowly cooled to rt. *UDG* (5.0 µL, 25 units, *New England Biolabs*) was added, carefully mixed and the mixture was incubated for 2 hours at 37 °C. The oligonucleotide was then isolated by chloroform/phenol extraction as described in the following paragraph. A CHCl₃/phenol solution (200 µL, *Roti Phenol*) was added, vortexed for 30 seconds and centrifuged for 3 minutes at rt and 13400 rpm. The aqueous phase was carefully removed and CHCl₃/phenol treatment was repeated twice. After addition of NaOAc (20 µL, 3M), the oligonucleotide was precipitated with *i*-PrOH (600 µL). The resulting DNA pellet was centrifuged at rt for 30 minutes (15000 rpm), washed with cold EtOH (300 µL) and centrifuged at 4 °C and 15000 rpm for another 30 minutes. The washing step was repeated once more, the supernatant removed and the pellet was dried on air for five minutes before the oligonucleotide was re-dissolved in ddH₂O (150 µL). The identity was finally confirmed by MALDI-TOF analysis.

3.6 Reaction Kinetics on a Synthetic Oligo with a Defined Abasic Site

In a total reaction volume of 20 µL, the oligonucleotide (300 pmol) was buffered with a solution of HEPES (20 mM, pH = 7.5), Na₂EDTA (0.1 mM) and **1a** (1.26 µL of 23.8 mM stock). The reaction (37 °C, 800 rpm, *Eppendorf comfort* thermomixer) was started by vortexing the mixture for 5 seconds and after specific time points (t = 15 s, 30 s, 45 s, 90 s, 120 s, 150 s, 180 s, 4 min, 6 min, 8 min, 15 min, 20 min) stopped via addition of acetone (200 µL). Excess acetone was removed on a speed vac (RVC-2-33 IR, *Christ*) and was filtered on a *AcroPrep Advance 96* filter plate (0.20 µm *Supor*, *Pall Life Sciences*). 75 pmol of DNA were subsequently injected into a *Dionex micro HPLC* system and reaction products were separated using a *Zorbax SB-C₁₈* column (0.55 x 250 mm, 5.0 µm pore size) with a flow rate of 350 µL/min. The analysis was run at a column temperature of 60 °C and a gradient of 0% -> 20% buffer B in 45 minutes (whereas buffer A = 10 mM TEAB, pH = 7.5 in H₂O and buffer B = 10 mM TEAB, pH = 7.5 in 80% MeCN/H₂O). Integration of the obtained UV signals (Figure S4) finally showed that the reaction of **1a** with abasic sites on an ODN is complete after 20 minutes and that no other fragments were generated under physiological conditions.

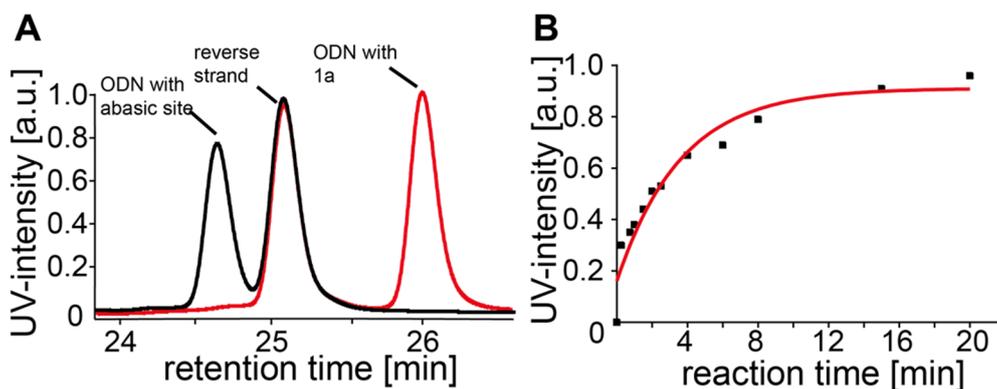


Figure S7: Reaction kinetics on oligo with defined abasic site. (A) Obtained UV-signals of ODN with abasic site and reverse strand before (black lines) and after derivatization with 1a. (B) Normalized UV signals of ODN + 1 after specific time points.

3.7 *In vitro* Assay of a Synthetic Oligonucleotide with hOGG1

An 8oxoG containing synthetic 13mer 5'-GTAATG8oxoGGCTAGG-3' and its counter strand 5'-CCTAGCCATTAC-3' were hybridized. hOGG1 activity was determined by treating 96 pmol of a single 8oxodG containing ds oligo in 1x buffer 2 (*New England BioLabs*), 100 $\mu\text{g}/\text{mL}$ nuclease free bovine serum albumin (BSA) and 4.8 U hOGG1 (*New England BioLabs*). The reaction was incubated for 20 h at 37°C and stopped by heat inactivation at 65°C for 5 minutes. The ds oligo was recovered using the *Oligo Clean & Concentrator Kit* (*Zymo Research*) according to the manufacturer. The assay was performed in triplicates and diluted for exact quantification.

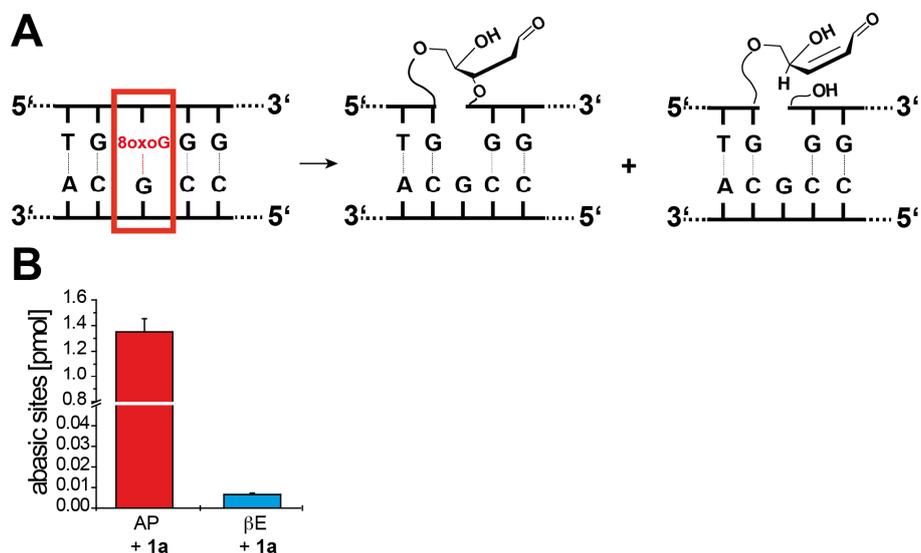


Figure S8: Reaction of an 8oxodG containing oligonucleotide with hOGG1. (A) Reaction of hOGG1 creating a defined AP- and βE -site inside a DNA duplex. (B) Quantification of the generated AP- and βE -sites after treatment of the oligonucleotide with 1a.

3.8 Efficiency of Enzymatic Digestion

In order to verify that the derivatized abasic site adducts can be completely enzymatically excised from the DNA, an aliquot of the fully reacted AP-site of the ODN (see section before) was diluted 1/4000 and the amount of AP-site **9a** was quantified as described in the sections before. In this sample an amount of 133 pmol of **9a** was found (for the undiluted sample). The content of dG in this oligo was quantified to be 895 pmol and in total dG was present 7 times in the ODN. Hence, if all the derivatized abasic sites were efficiently hydrolyzed, an amount of 128 pmol of **9a** was expected which is in good agreement with the 133 pmol that were quantified.

3.9 Reaction Kinetics on Abasic Sites in Genomic DNA

Reactions were carried out by derivatizing 5 μg of gDNA with **1a** using the same conditions as mentioned above (*Derivatization of genomic DNA with 1a*). The reaction was stopped through the addition of 1-naphthylaldehyde (66.7 μL , 2M in *i*-PrOH) at specific time points ($t = 1 \text{ min}$, 2.5 min, 5 min, 10 min, 20 min, 30 min, 60 min). Reaction aliquots were finally digested to the nucleoside level and quantified whereas it was shown that after one minute of reaction time, all abasic sites were derivatized and a prolonged incubation up to 60 minutes showed that no abasic sites are generated artificially under these conditions (Figure S5A). Since the reaction with genomic DNA was very fast, shorter time points were also investigated ($t = 0 \text{ s}$, 10 s, 15 s, 30 s, 45 s, 1 min, 2 min, 5 min, 10 min, 15 min) and again, after one minute incubation time, all abasic- and βE - sites were derivatized (Figure S5B and C).

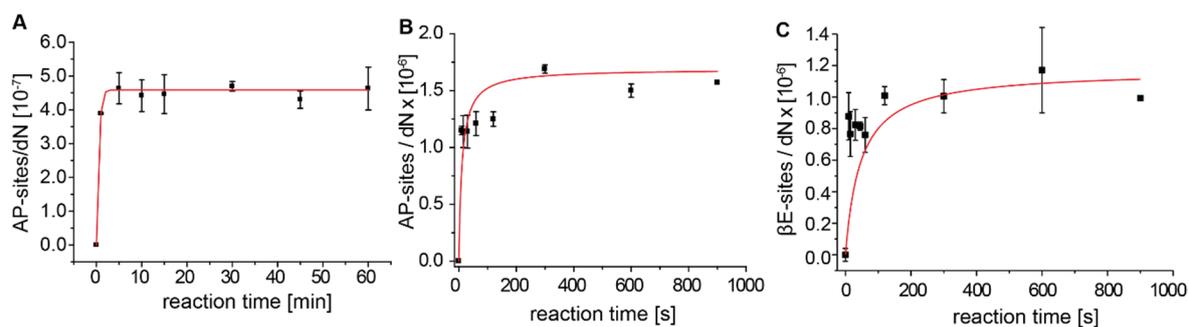


Figure S9: Derivatization of gDNA with **1** shows a fast derivatization reaction and does not artificially generate abasic sites.

Literature

1. Kojima, N.; Takebayashi, T.; Mikami, A.; Ohtsuka, E.; Komatsu, Y., *J. Am. Chem. Soc.* **2009**, *131*, 13208-13209.
2. Li, E.; Bestor, T. H.; Jaenisch, R., *Cell* **1992**, *69*, 915-26.
3. Tsumura, A.; Hayakawa, T.; Kumaki, Y.; Takebayashi, S.; Sakaue, M.; Matsuoka, C.; Shimotohno, K.; Ishikawa, F.; Li, E.; Ueda, H. R.; Nakayama, J.; Okano, M., *Genes Cells* **2006**, *11*, 805-14.
4. Cortazar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; Steinacher, R.; Jiricny, J.; Bird, A.; Schar, P., *Nature* **2011**, *470*, 419-423.
5. Liu, N.; Wang, M.; Deng, W.; Schmidt, C. S.; Qin, W.; Leonhardt, H.; Spada, F., *PLoS One* **2013**, *8*, e62755.
6. Cao, H.; Wang, Y., *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1335-41.
7. Pfaffeneder, T.; Spada, F.; Wagner, M.; Brandmayr, C.; Laube, S. K.; Eisen, D.; Truss, M.; Steinbacher, J.; Hackner, B.; Kotljarova, O.; Schuermann, D.; Michalakis, S.; Kosmatchev, O.; Schiesser, S.; Steigenberger, B.; Raddaoui, N.; Kashiwazaki, G.; Müller, U.; Spruijt, C. G.; Vermeulen, M., *Nature Chem. Biol.* **2014**, *10*, 574-581.
8. Schiesser, S.; Pfaffeneder, T.; Sadeghian, K.; Hackner, B.; Steigenberger, B.; Schröder, A. S.; Steinbacher, J.; Kashiwazaki, G.; Höfner, G.; Wanner, K. T.; Ochsenfeld, C.; Carell, T., *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599.
9. Spruijt, C. G.; Gnerlich, F.; Smits, A. H.; Pfaffeneder, T.; Jansen, P. W.; Bauer, C.; Munzel, M.; Wagner, M.; Müller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, A. B.; Lephikov, K.; Müller, U.; Walter, J.; Boelens, R.; van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M., *Cell* **2013**, *152*, 1146-59.
10. Wang, J.; Yuan, B.; Guerrero, C.; Bahde, R.; Gupta, S.; Wang, Y., *Anal. Chem.* **2011**, *83*, 2201-2209.

3.3 Entwicklung der Synthese eines Nukleosidstandards zur Untersuchung des m⁶dA und m⁴dC Gehalts in embryonalen Stammzellen von Mäusen

„Quantitative LC-MS Provides No Evidence for m⁶dA or m⁴dC in the Genome of Mouse Embryonic Stem Cells and Tissues“ Sarah Schiffers, Charlotte Ebert, René Rahimoff, Olesya Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, Fabio Spada, Stylianos Michalakis, Jörg Nickelsen, Markus Müller, Thomas Carell, *Angew. Chem. Int. Ed.* **2017**, *56*, 11268 – 11271.

Prolog

Bevor hmdC, fdC und cadC als aktiv erzeugte, epigenetisch relevante Modifikationen in genomischer DNA höherer Organismen entdeckt werden konnten, galt mdC als einzige Modifikation, die in der Kontrolle der Genexpression eine distinkte Rolle besitzt. Dies steht in starkem Kontrast zur RNA, bei der mittlerweile weit über 100 verschiedene Basenmodifikationen bekannt sind und denen mitunter sogar bestimmte Funktionen in der Genexpressionskontrolle nachgewiesen werden konnten. Vor allem für das am exozyklischen Amin von Adenosin monomethylierte m⁶A wurde eine Reihe spezifischer *reader* Proteine entdeckt und es konnte beschrieben werden, dass diese einfache Modifikation bei der Stabilität und Translationskontrolle der mRNA eine enorm wichtige Rolle spielt. Natürlich stellt sich die Frage, ob es solche an der exozyklischen Aminfunktion methylierte DNA Basen ebenfalls im Genom höherer Organismen geben kann und, wenn ja, ob ihnen eine spezifische Funktion zukommt. Tatsächlich häuften sich in letzter Zeit immer mehr Berichte über die Entdeckung von m⁶dA im Genom verschiedener Eukaryoten, das, zusammen mit m⁴dC, zunächst als Bestandteil des genetischen Materials von Bakterien bekannt war. In dem vorliegenden Manuskript wurde eine quantitative UHPLC-MS/MS Methode entwickelt, um mit Hilfe der Zugabe isotope markierter Nukleosidstandards den genauen Gehalt an m⁶dA und m⁴dC in genomischer DNA zu untersuchen. Als Positivkontrolle konnte die Präsenz von m⁶dA und m⁴dC im Cyanobakterium *Synechocystis* und m⁶dA in der Grünalge *C. reinhardtii* bestätigt werden. Interessanterweise wurden diese Modifikationen jedoch nicht in Stammzell-DNA und im Leber- oder Gehirngewebe von Mäusen nachgewiesen. Das steht im starken Widerspruch zu bereits veröffentlichten Studien. Vielmehr konnte gezeigt werden, dass im Falle dieser vorliegenden Studie der Gleichgewichtsgehalt an m⁶dA mit einer möglichen Verunreinigung aus dem verwendeten Enzymcocktail, mit dem die genomische DNA auf Nukleosid-Ebene verdaut wird, in Verbindung gebracht werden könnte. Letztendlich wurde geschlussfolgert, dass m⁶dA vermutlich keine epigenetisch relevante Funktion besitzt.

Autorenbeitrag

Für diese Arbeit entwickelte ich eine Syntheseroute für den m^4dC -Standard und war für dessen Synthese und HPLC-Aufreinigung verantwortlich. *Charlotte Ebert* synthetisierte auf dieser Grundlage den isotopenmarkierten D_3-m^4dC Standard und entwickelte die Synthese für m^6dA in isotopenmarkierter und unmarkierter Form. Die Hauptarbeit dieser Studie lag bei *Sarah Schiffers*.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags; Copyright 2016 *Wiley VCH Verlag GmbH und Co. KGaA, Weinheim*.

Epigenetics

International Edition: DOI: 10.1002/anie.201700424

German Edition: DOI: 10.1002/ange.201700424

Quantitative LC–MS Provides No Evidence for m⁶dA or m⁴dC in the Genome of Mouse Embryonic Stem Cells and Tissues

Sarah Schiffrers, Charlotte Ebert, René Rahimoff, Olesya Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, Fabio Spada, Stylianos Michalakis, Jörg Nickelsen, Markus Müller, and Thomas Carell*

Abstract: Until recently, it was believed that the genomes of higher organisms contain, in addition to the four canonical DNA bases, only 5-methyl-dC (m⁵dC) as a modified base to control epigenetic processes. In recent years, this view has changed dramatically with the discovery of 5-hydroxymethyl-dC (hmdC), 5-formyl-dC (fdC), and 5-carboxy-dC (cadC) in DNA from stem cells and brain tissue. N⁶-methyldeoxyadenosine (m⁶dA) is the most recent base reported to be present in the genome of various eukaryotic organisms. This base, together with N⁴-methyldeoxycytidine (m⁴dC), was first reported to be a component of bacterial genomes. In this work, we investigated the levels and distribution of these potentially epigenetically relevant DNA bases by using a novel ultrasensitive UHPLC–MS method. We further report quantitative data for m⁵dC, hmdC, fdC, and cadC, but we were unable to detect either m⁴dC or m⁶dA in DNA isolated from mouse embryonic stem cells or brain and liver tissue, which calls into question their epigenetic relevance.

The genetic material of living organisms is constructed from the four canonical nucleobases dA, dC, dG, and dT, which establish the sequence information that, in multicellular organisms, is stored in the nucleus of every cell (Figure 1). In addition to the canonical bases, the methylated dC base 5-methyldeoxycytidine (m⁵dC) is frequently found.^[1] The presence or absence of this base in specific promoter segments determines whether the gene is actively transcribed or silenced.^[1] The cell-type-specific distribution of m⁵dC thus determines the identity of a given cell. Recently, 5-hydroxymethyldeoxycytidine (hmdC) was found as a sixth base of

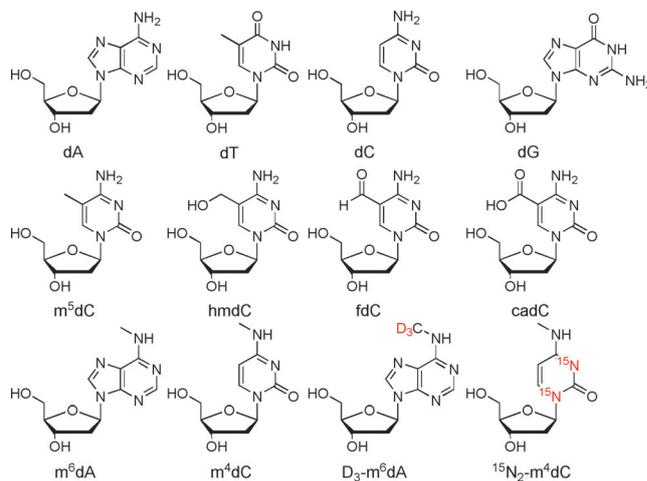


Figure 1. Depiction of the four canonical DNA bases and the epigenetic DNA marks m⁵dC, hmdC, fdC, and cadC, as well as the bases m⁶dA and m⁴dC together with the synthesized isotopologues.

the genetic system^[2,3] and in 2011, 5-formyldeoxycytidine (fdC)^[4,5] and 5-carboxydeoxycytidine (cadC)^[5,6] were also discovered, particularly in DNA isolated from stem cells, but also in brain DNA. It is currently believed that fdC and cadC are intermediates in an active demethylation process that allows cells to change the methylation pattern and hence the activity state of specific genes.^[7,8] For fdC, separate epigenetic functions are also envisaged.^[9]

While the genomes of bacteria are known to also contain N⁴-methyldeoxycytidine (m⁴dC)^[10] and N⁶-methyldeoxyadenosine (m⁶dA),^[11] attempts to detect these bases in the DNA of higher organisms have failed until recently.^[12–15] m⁶dA has now been found in algae (0.4 mol % m⁶dA/A),^[12] fruit flies (0.001%–0.07% m⁶dA/A),^[14] and *C. elegans* (0.01%–0.4% m⁶dA/A),^[13] and its presence has even been reported in the DNA of vertebrates (0.00009% in *X. laevis*^[16] and 0.00019–0.003% of dA in murine cells and tissue^[17]). These discoveries, especially concerning the DNA of vertebrates, have spurred a worldwide research interest in unraveling the function of these new bases in human genomic DNA.^[18–20]

In this study, we developed an ultrasensitive triple quadrupole mass spectrometry (QQQ-MS) method, which in combination with ultra-high-pressure chromatography (UHPLC) enables m⁴dC and m⁶dA to be searched for and quantified in parallel to the more established new epigenetic DNA marks m⁵dC, hmdC, fdC and cadC.

*] S. Schiffrers, C. Ebert, R. Rahimoff, O. Kosmatchev, J. Steinbacher, Dr. F. Spada, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPSM) at the Department of Chemistry, LMU München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: <http://www.carellgroup.de>
Dr. A.-V. Bohne, Prof. Dr. J. Nickelsen
Biocenter of the LMU
Dept. Biologie 1 – Botanik, Molekulare Pflanzenwissenschaften
Grosshaderner Strasse 2–4, 82152 Planegg-Martinsried (Germany)
Dr. S. Michalakis
CiPSM, Department of Pharmacy, LMU München
Pharmacology for Life Sciences
Butenandtstr. 7, 81377 München (Germany)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<http://dx.doi.org/10.1002/anie.201700424>.

For the quantitative measurements, we first chemically synthesized the two isotopologues of m^6dA and m^4dC shown in Figure 1 as internal standards for the analytical method. The prepared compounds D_3 - m^6dA and $^{15}N_2$ - m^4dC are three and two mass units heavier, respectively, than the natural bases. Despite these molecular-weight differences, they have identical properties during the UHPLC chromatography step so that they strictly coelute with their natural counterparts, thus allowing them to enter the mass spectrometer at exactly the same time as the internal standards. The availability of these isotopologues makes the method highly reliable and strictly quantitative. The syntheses of the two compounds, together with all analytical data, are given in the Supporting Information.

We first benchmarked our study with an investigation of genomic DNA isolated from the unicellular green algae *Chlamydomonas reinhardtii* and the cyanobacterium *Synechocystis*. In both cases, DNA was isolated after cell lysis using a standard method (see the Supporting Information). The isolated DNA was subsequently digested with a mixture of three commercially available digestion enzymes (Nuclease S1, Antarctic Phosphatase, and Snake Venom Phosphodiesterase; see the Supporting Information). We next added the isotope-labelled standards D_3 - m^6dA and $^{15}N_2$ - m^4dC to the obtained digestion mixture and performed UHPLC-QQQ analysis. For the mass spectrometry detection, we selected fragmentation of the glycosidic bond as the indicative and hence recorded mass transition. This is $m/z = 266.12 \rightarrow 150.08$ for m^6dA and $m/z = 269.14 \rightarrow 153.10$ for its isotopologue D_3 - m^6dA . For m^4dC , we also used fragmentation of the glycosidic bond, which gives a mass transition of $m/z = 242.11 \rightarrow 126.07$ for the natural compound m^4dC and $m/z = 244.11 \rightarrow 128.07$ for its isotopologue $^{15}N_2$ - m^4dC (Figure 2A).

We next modified the reported UHPLC-QQQ method^[10] for the simultaneous quantification of m^4dC and m^6dA , together with the other epigenetically relevant bases m^5dC , hmdC, fdC, and cadC. To this end, the UHPLC gradient was fine-tuned to enable full separation of all six compounds. Finally, we measured precise calibration curves for all of the compounds (see Figures S1 and S2 in the Supporting Information). This subsequently allowed exact quantification of all of the discussed epigenetic DNA marks in a given sample (Figure 2B–D).

Since m^4dC and m^6dA are well known in bacteria, we first analysed the cyanobacterium *Synechocystis* (PCC6803), and we indeed found both bases (Figure 2B). The base m^6dA was detected at a level of 8.4×10^{-3} per dN and for m^4dC we measured a value of 5.9×10^{-3} per dN. The constitutional isomer m^5dC and all other dC-derived epigenetic DNA marks were detectable, but were not quantified in this experiment.

Next, we analyzed two different strains of *Chlamydomonas reinhardtii* (CC-3491 and wt 7d+), in which m^6dA has just recently been discovered,^[12] and the levels of m^6dA were determined to be 8.4×10^{-4} per dN for CC-3491 and 6.9×10^{-4} per dN for wt 7d+ (Figure 2C). This corresponds to about 3000 m^6dA bases per *Chlamydomonas* genome (genome size 1.2×10^8), which at 0.7% of the dA is a relatively high number. In both strains, m^4dC was not detected, thus showing that this base is unlikely to be a component of the genetic

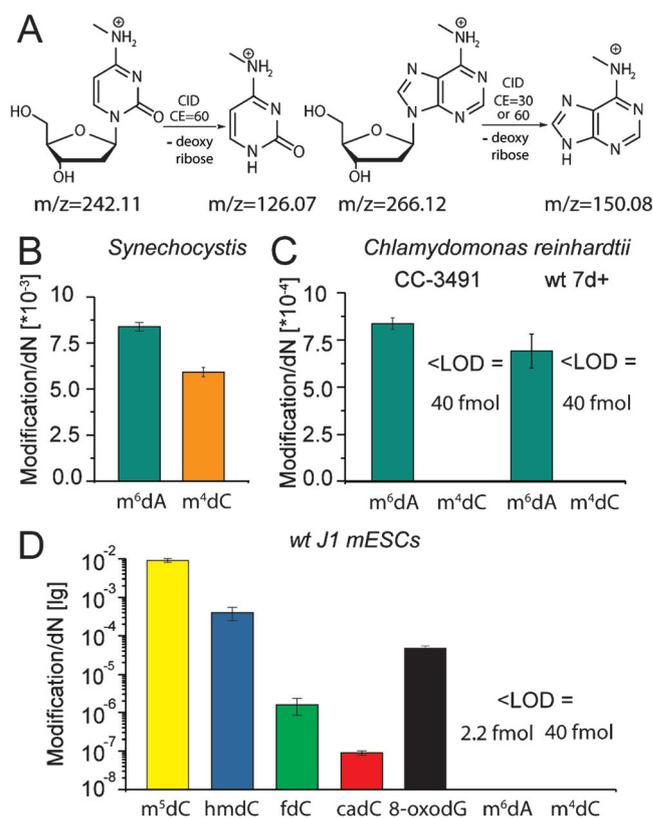


Figure 2. A) Fragmentation patterns of m^4dC and m^6dA . B–D) Quantitative data of the bases m^4dC and m^6dA in *Synechocystis* (B), *Chlamydomonas* (C) and of these bases and the other epigenetic DNA marks hmdC, fdC, cadC, and m^5dC in mouse embryonic stem cells (D).

material of *Chlamydomonas*. This is interesting because *Synechocystis* is considered a relative of the chloroplasts present in *Chlamydomonas*.

With these positive results in hand, we extended our study to mouse embryonic stem cells (ESCs; wt J1, Figure 2D). m^6dA in particular was recently reported to occur as an epigenetically relevant DNA mark in mouse ESCs (mESC cell line wt TT2).^[17] When performing the measurements, we turned the mass spectrometer to maximum sensitivity. But even in this mode, we were unable to detect a signal for m^6dA within the detection limits of our system Table S2. In contrast, the other epigenetically relevant bases hmdC, fdC, cadC, and even the oxidative lesion 8-oxodG, which we also quantified in parallel, were clearly detectable. The 8-oxodG level was 4.8×10^{-5} per dN. The rare and difficult to detect cadC was clearly seen even at levels of only 9.0×10^{-8} per dN. For m^6dA , in contrast, a signal did not appear. We also re-measured wt TT2 cells as described and still did not detect m^6dA over background levels (see Figure S4).^[17] To obtain unequivocal proof that m^6dA is not present in stem cells, we added $^{13}CD_3$ -methionine to the mESC culture. Methionine provides the methyl group for the biosynthesis of m^6dA . With $^{13}CD_3$ -methionine, this would lead to an m/z -shift of +4. We tuned the mass spectrometer to the new m/z -transition and again were unable to see any signal for $^{13}CD_3$ - m^6dA (Figure S8).

We subsequently turned our attention to adult mouse tissue and analysed DNA isolated from liver and whole brain

(Figure S3) using our UHPLC-QQQ method. Figure 3 shows the data obtained from mouse liver. The middle column shows the data we obtained for hmdC. The already reported D_2 - $^{15}N_2$ -hmdC standard elutes at a retention time of 2.25 min

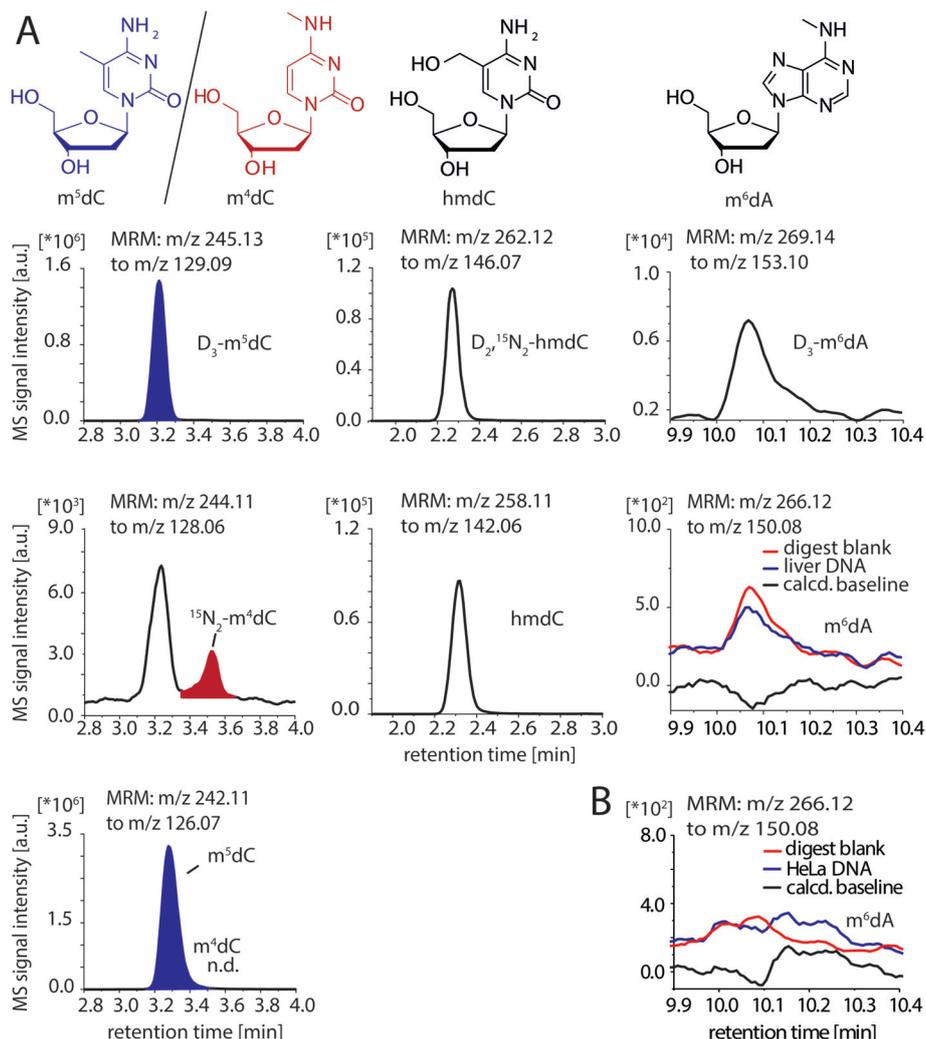


Figure 3. A) Chromatograms of the mass signal of mouse liver DNA. UHPLC-QQQ data obtained for m^5dC and m^4dC (left), hmdC (middle), and m^6dA (right, blue line) and their corresponding isotopic standards are shown. Additionally, for m^6dA , the chromatogram for the digest blank is shown (red line) and a computed baseline (black line), which was determined by subtracting the digest blank from the sample. B) A chromatogram of the mass signal from UHPLC-QQQ data obtained for m^6dA in DNA from HeLa cells.

and shows the expected fragmentation of the glycosidic bond, providing the fragmentation signature $m/z = 262.12 \rightarrow 146.07$, which allowed assignment of the signal. The naturally occurring hmdC is detected at exactly the same retention time with a mass transition of $m/z = 258.11 \rightarrow 142.06$, thus unequivocally demonstrating the presence of hmdC in murine liver DNA. Regarding the different monomethylated dC compounds m^5dC and m^4dC (left column), the epigenetic DNA mark m^5dC is clearly detected at a retention time of 3.2 min, but for m^4dC with a retention time of 3.5 min, there is obviously no signal present.

The m^6dA data are highly interesting (Figure 3 right column). While the D_3 - m^6dA standard was clearly detectable at a retention time of 10.1 min, the unlabelled m^6dA provided a very weak signal. We then performed a control experiment to determine the limit of detection and investigated the digestion solution alone, which contains all of the commercial enzymes but no isolated DNA (red chromatogram). A weak signal for m^6dA was again detectable at a retention time of 10.1 min. After subtracting this background signal (red) from the measured chromatogram (blue), we obtained the black line showing that the original signal at 10.1 min is purely caused by background derived from the enzyme mixture. Here it is important to note that most of these proteins are recombinant proteins obtained from bacterial expression systems and bacterial DNA contains plenty of m^6dA . To support the evidence that m^6dA is not present in vertebrate DNA, we performed the same experiment with HeLa cells and also observed no signal for m^6dA (Figure 3B).

We then determined our limit of detection for m^6dA to be 3.5×10^{-7} per dN, which corresponds to 170 m^6dA bases per murine genome. This is a very small number and demonstrates the excellent sensitivity of our method. It shows that the maximum number of m^6dA that could be theoretically present and would not be detectable by our method is less than 170 m^6dA bases per genome, which led us to conclude that m^6dA is likely not epigenetically relevant but rather formed as a DNA lesion, perhaps by misguided methyltransferases. Spiking tests with synthetic nucleoside and DNA from *Chlamydomonas* nevertheless confirmed the sensitivity of the method, since the input amount equalled the found amount (see Figure S6).

To find a potential source for m^6dA in mESC DNA that could explain previous sequencing data,^[17] we thought that m^6dA -containing bacterial DNA that gets degraded could provide the m^6dA nucleoside, which then might get incorporated into mESC DNA. This is indeed a possibility. When we added the m^6dA nucleoside to a mESCs culture, we indeed saw incorporation of some m^6dA into the genome (Figure S7).

We also cannot fully exclude the possibility that the presence of few m⁶dA bases, at levels below our detection limit, could have a biological function. In addition, it is possible that at certain stages of organismal development, certain methyltransferases are activated that may induce high m⁶dA levels at specific time points that may have escaped our detection.^[18] Our data, however, show clearly that the maximum possible levels of m⁶dA in the analysed organisms and mESCs under normal conditions are far lower than so far believed.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for financial support via SFB1032 (TP-A5), SFB749 (TP-A4), SPP1784, CA275 and the Excellence Cluster CiPS^M.

Conflict of interest

The authors declare no conflict of interest.

Keywords: DNA · epigenetics · methyldeoxyadenosine · methyldeoxycytidine · mass spectrometry

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 11268–11271
Angew. Chem. **2017**, *129*, 11422–11425

-
- [1] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204–220.
[2] 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.
[3] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
[4] 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; *Angew. Chem.* **2011**, *123*, 7146–7150.
[5] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
[6] 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.
[7] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472–479.
[8] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
[9] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797–11800; *Angew. Chem.* **2016**, *128*, 11974–11978.
[10] M. Yu, L. Ji, D. A. Neumann, D. H. Chung, J. Groom, J. Westpheling, C. He, R. J. Schmitz, *Nucleic Acids Res.* **2015**, *43*, e148.
[11] M. Ehrlich, M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, C. W. Gehrke, *Nucleic Acids Res.* **1985**, *13*, 1399–1412.
[12] Y. Fu, G. Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L. C. Dore, X. Weng, Q. Ji, L. Mets, C. He, *Cell* **2015**, *161*, 879–892.
[13] E. L. Greer, M. A. Blanco, L. Gu, E. Sendinc, J. Liu, D. Aristizabal-Corrales, C. H. Hsu, L. Aravind, C. He, Y. Shi, *Cell* **2015**, *161*, 868–878.
[14] G. Zhang, H. Huang, D. Liu, Y. Cheng, X. Liu, W. Zhang, R. Yin, D. Zhang, P. Zhang, J. Liu, C. Li, B. Liu, Y. Luo, Y. Zhu, N. Zhang, S. He, C. He, H. Wang, D. Chen, *Cell* **2015**, *161*, 893–906.
[15] S. Hattman, C. Kenny, L. Berger, K. Pratt, *J. Bacteriol.* **1978**, *135*, 1156–1157.
[16] M. J. Koziol, C. R. Bradshaw, G. E. Allen, A. S. Costa, C. Frezza, J. B. Gurdon, *Nat. Struct. Mol. Biol.* **2016**, *23*, 24–30.
[17] T. P. Wu, T. Wang, M. G. Seetin, Y. Lai, S. Zhu, K. Lin, Y. Liu, S. D. Byrum, S. G. Mackintosh, M. Zhong, A. Tackett, G. Wang, L. S. Hon, G. Fang, J. A. Swenberg, A. Z. Xiao, *Nature* **2016**, *532*, 329–333.
[18] J. Liu, Y. Zhu, G. Z. Luo, X. Wang, Y. Yue, X. Wang, X. Zong, K. Chen, H. Yin, Y. Fu, D. Han, Y. Wang, D. Chen, C. He, *Nat. Commun.* **2016**, *7*, 13052.
[19] W. Huang, J. Xiong, Y. Yang, S.-M. Liu, B.-F. Yuan, Y.-Q. Feng, *RSC Adv.* **2015**, *5*, 64046–64054.
[20] G. Z. Luo, F. Wang, X. Weng, K. Chen, Z. Hao, M. Yu, X. Deng, J. Liu, C. He, *Nat. Commun.* **2016**, *7*, 11301.

Manuscript received: January 13, 2017

Revised manuscript received: February 15, 2017

Version of record online: March 30, 2017

Supporting Information

Quantitative LC–MS Provides No Evidence for m⁶dA or m⁴dC in the Genome of Mouse Embryonic Stem Cells and Tissues

*Sarah Schiffers, Charlotte Ebert, René Rahimoff, Olesya Kosmatchev, Jessica Steinbacher, Alexandra-Viola Bohne, Fabio Spada, Stylianos Michalakis, Jörg Nickelsen, Markus Müller, and Thomas Carell**

anie_201700424_sm_miscellaneous_information.pdf

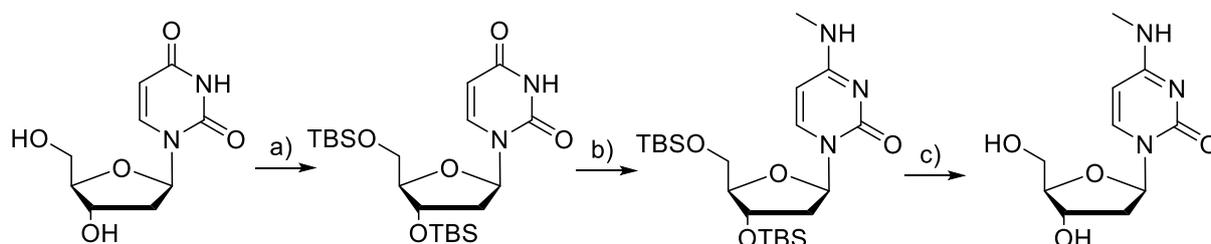
Supporting information

General Methods

Chemicals were purchased from Sigma-Aldrich, Fluka, Alfa Aesar, ChemGenes or Acros and used without further purification. Solutions were concentrated *in vacuo* on a Heidolph rotary evaporator. The solvents were of reagent grade and purified by distillation. Dry solvents were bought from Sigma-Aldrich. Acetonitrile for HPLC-purification was purchased from VWR. Water was purified by a Milli-Q Plus system from Merck Millipore. 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 F254) plates. ^1H and ^{13}C -NMR spectra were recorded in deuterated solvents on *Bruker ARX 300*, *Varian VXR400S*, *Varian Inova 400*, *Bruker AMX 600* and *Bruker Avance 800* spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, hept = septet, m = multiplet, br. = broad. For assignment of the structures, additional 2D NMR spectra (COSY, HSQC, HMBC) were measured. High resolution electrospray ionization mass spectra (HRMS-ESI) were recorded on a Thermo Finnigan LTQ-FT (ESI-FTICR).

Synthesis of 2'-deoxy-4-methyl-cytidine

The final compounds were purified by reverse phase HPL chromatography (Macherey-Nagel, Nucleosil 100-7 C18, 10 \times 250 mm, linear gradient from 0% to 10% acetonitrile in water in 45 min for 2'-deoxy-4-methyl-cytidine or 0% to 15% acetonitrile in water in 45 min for 2'-deoxy-6-methyladenosine).



Scheme S1: Synthesis of 2'-deoxy-4-methyl-cytidine. TBS-Cl, imidazole, pyridine, 0 $^{\circ}\text{C}$ to r.t., o/n, quant b) 1. NaH, THF, r.t., 45 min, 2. TPS-Cl, r.t., 15 h, 3. MeNH₂, r.t., 5 min, 86% c) HF in pyridine, EtOAc, r.t., o/n, quant.

Synthesis of 2'-deoxy-4-methyl-cytidine

O^{3'},*O*^{5'}-TBS-2'-deoxyuridine

2'-Deoxyuridine (3.00 g, 13.1 mmol, 1.0 eq) was dissolved in pyridine (50 mL) and cooled to 0 °C. Imidazole (4.00 g, 59.0 mmol, 4.5 eq) and TBS-Cl (5.90 g, 39.3 mmol, 3.0 eq) were then added. The mixture was allowed to warm to rt overnight and volatiles were subsequently removed *in vacuo*. The gummy residue was taken up in DCM (100 mL) and washed with sat. NH₄Cl (3x 100 mL). The organic phase was dried over Na₂SO₄, the organic solvent removed under reduced pressure and the resulting oil was finally coevaporated with toluene (3x) to yield a white powder in quantitative yield (5.98 g, 13.1 mmol) that was used without further purification.

O^{3'},*O*^{5'}-TBS-4-methyl-2'-deoxycytidine

O^{3'},*O*^{5'}-TBS-2'-deoxyuridine (1.00 g, 2.19 mmol, 1.0) was dissolved in THF (50 mL) and cooled to 0 °C. NaH (60% in mineral oil, 434 mg, 10.95 mmol, 5.0 eq) was added in one portion and the grey suspension was kept at 0 °C for 45 minutes. Subsequently, TPS-Cl (1.33 g, 4.38 mmol, 2.0 eq) was added and the mixture was warmed to rt and stirred for 15 hours. After all the starting material was consumed (as judged by TLC), ice cold sat. NH₄Cl (200 mL) was added cautiously and the aqueous layer was extracted with EtOAc (2x 100 mL). After drying the organic phase over Na₂SO₄, volatiles were removed *in vacuo* and the residue was charged with MeNH₂ (50 mL, 33% in EtOH). After stirring at rt for 5 minutes, the reaction was aborted through addition of NH₄Cl (200 mL) and the aqueous phase was again extracted with EtOAc (2x 50 mL). After drying over Na₂SO₄ and removal of EtOAc *in vacuo*, the compound was purified by column chromatography (DCM --> 2.5% MeOH --> 4% MeOH) and the product was yielded as a colourless foam (887 mg, 1.80 mmol, 86%).

¹H-NMR (599 MHz, CDCl₃, ppm): δ = 7.87 (d, *J* = 7.8 Hz, 1H, 6-H), 6.29 (dd, *J* = 7.8 Hz, 6.3 Hz, 1H, 1'-H), 5.55 (d, *J* = 7.8 Hz, 1H, 5-H), 4.35 (q, *J* = 5.5 Hz, 1H, 3'-H), 3.90 (dd, *J* = 11.4 Hz, 2.6 Hz, 1H, 5'-H), 3.86 (m, 1H, 4'-H), 3.75 (dd, *J* = 11.1 Hz, 2.3 Hz, 1H, 5'-H), 3.0 (ds, *J* = 4.9 Hz, 3H, CH₃-NH), 2.44 - 2.36 (m, 1H, 2'-H), 2.07 (ddd, *J* = 13.5, 6.4, 4.9 Hz, 1H, 2'-H), 0.96 - 0.82 (m, 18H, TBS-C(CH₃)₃), 0.14 - 0.01 (m, 12H, TBS-CH₃).

¹³C-NMR (151 MHz, CDCl₃, ppm): δ = 164.2 (C4), 156.2 (C2), 139.5 (C6), 94.4 (C5), 87.1 (C4'), 85.7 (C3'), 70.1 (C1'), 61.9 (C5'), 42.1 (C2'), 28.0 (CH₃), 25.9 - 25.7 (TBS-C(CH₃)₃), 18.3 - 18.0 (TBS-C(CH₃)₃), -4.6 - -5.6 (TBS-CH₃).

HR-MS (ESI⁺): *m/z* calculated for [C₂₂H₄₄N₃O₄Si₂⁺] ([M+H]⁺): 470.2865, found: 470.2864.

2'-deoxy-4-methyl-cytidine

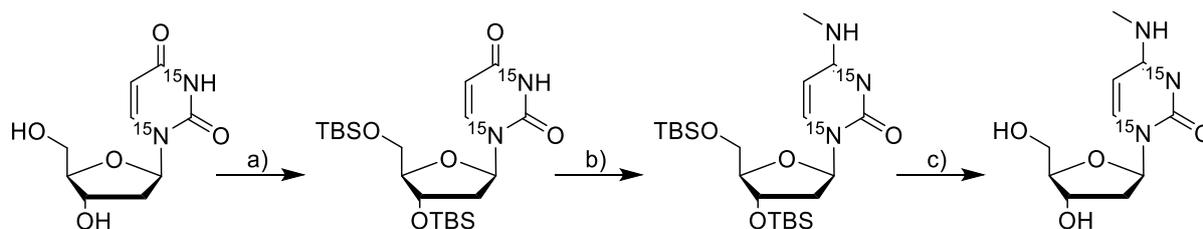
In a falcon tube, the TBS-protected nucleoside (300 mg, 0.64 mmol, 1.0 eq) was dissolved in EtOAc (5 mL). The mixture was then charged with HF•pyridine (70%, 374 μ L, 14.4 mmol, 22.5 eq) and stirred at rt overnight. Then TMS•OMe (3.46 mL, 25.6 mmol, 40 eq) was added whereas a colorless precipitate was formed. The white solid was isolated through repeated centrifugation of the mixture (2x 10 minutes, 6000 rpm) and subsequent washing with EtOAc (2x 5 mL). The supernatant was discarded and the compound (154 mg, 0.64 mmol, quant.) was yielded as a white powder, that was further purified by HPLC.

^1H NMR (400 MHz, D_2O) δ 7.68 (d, J = 7.6 Hz, 1H, C6-H), 6.29 (t, J = 6.7 Hz, 1H, C1'-H), 5.98 (d, J = 7.6 Hz, 1H, C5-H), 4.45 (dt, J = 6.5, 3.9 Hz, 1H, C3'-H), 4.05 (dt, J = 5.1, 3.7 Hz, 1H, C4'-H), 3.84 (dd, J = 12.4, 3.7 Hz, 1H, C5'-H), 3.76 (dd, J = 12.4, 5.2 Hz, 1H, C5'-H), 2.89 (s, 3H, CH_3), 2.41 (ddd, J = 14.1, 6.4, 4.0 Hz, 1H, C2'-H), 2.29 (dt, J = 13.9, 6.8 Hz, 1H, C2'-H) ppm.

^{13}C NMR (101 MHz, D_2O) δ 164.5 (C4), 157.8 (C2), 139.3 (C6), 97.3 (C5), 86.4 (C4'), 85.8 (C1'), 70.6 (C3'), 61.3 (C5'), 39.1 (C2'), 27.1 (CH_3) ppm.

HR-MS (ESI+): m/z calculated for $[\text{C}_{10}\text{H}_{16}\text{N}_3\text{O}_4]^+$ ($[\text{M}+\text{H}]^+$): 242.1136, found: 242.1137.

Synthesis of 2'-deoxy-4-methyl- $[\text{N}^1, \text{N}^3\text{-}^{15}\text{N}_2]$ -cytidine



Scheme S2: Synthesis of 2'-deoxy-4-methyl- $[\text{N}^1, \text{N}^3\text{-}^{15}\text{N}_2]$ -cytidine. a) TBS-Cl, imidazole, pyridine, r.t., 72 h, 64% b) 1. NaH, THF, r.t., 45 min, 2. TPS-Cl, r.t., 17 h, 3. MeNH_2 , r.t., 10 min, 49% c) HF in pyridine, EtOAc, r.t., 17 h, 64%.

O^3, O^5 -TBS-2'-deoxy- $[\text{N}^1, \text{N}^3\text{-}^{15}\text{N}_2]$ -uridine

2'-Deoxy- $[\text{N}^1, \text{N}^3\text{-}^{15}\text{N}_2]$ -uridine (51 mg, 0.22 mmol, 1.0 eq), TBS-chloride (109 mg, 0.726 mmol, 3.3 eq) and imidazole (99 mg, 1.4 mmol, 6.6 eq) were dissolved in 1 mL pyridine and stirred at

r.t. for 72 h. The solvent was removed *in vacuo*, the residue was dissolved in DCM and washed with sat. NaHCO₃ (aq). *O*^{3'},*O*^{5'}-TBS-2'-deoxy-[*N*¹,*N*³-¹⁵N₂]-uridine (59 mg, 0.13 mmol, 64%) was obtained as a white solid.

HR-MS (ESI⁺): *m/z* calculated for [C₂₁H₄₁¹⁵N₂O₅Si₂]⁺ ([M+H]⁺): 459.2489, found: 459.2493.

***O*^{3'},*O*^{5'}-TBS-2'-deoxy-[*N*¹,*N*³-¹⁵N₂]-cytidine**

O^{3'},*O*^{5'}-TBS-2'-deoxy-[*N*¹,*N*³-¹⁵N₂]-uridine (59 mg, 0.13 mmol, 1.0 eq) was dissolved in THF (2 mL) and cooled to 0 °C. Sodium hydride (16 mg, 0.65 mmol, 5.0 eq) was added and the mixture was stirred at r.t. for 45 minutes. Then TPS-chloride (79 mg, 0.26 mmol, 2.0 eq) was added and the mixture was stirred at r.t. for another 17 h. The reaction was quenched with saturated NH₄Cl_(aq), extracted with EtOAc and washed three times with saturated NH₄Cl_(aq). The organic phase was dried with Na₂SO₄ and the solvent was removed *in vacuo*. The residue was dissolved in MeNH₂ (5 mL, 33% in EtOH) and stirred at r.t. for 10 minutes. The solvent was removed *in vacuo* and the residue was purified via column chromatography (DCM/MeOH 0.5% to 1%, SiO₂). *O*^{3'},*O*^{5'}-TBS-2'-deoxy-4-methyl-[*N*¹,*N*³-¹⁵N₂]-cytidine (30 mg, 0.064 mmol, 49%) was obtained as a white solid. Unreacted *O*^{3'},*O*^{5'}-TBS-2'-deoxy-[*N*¹,*N*³-¹⁵N₂]-uridine could be recovered via column chromatography.

HR-MS (ESI⁺): *m/z* calculated for [C₂₂H₄₄N¹⁵N₂O₄Si₂]⁺ ([M+H]⁺): 472.2806, found: 472.2804.

2'-deoxy-4-methyl-[*N*¹,*N*³-¹⁵N₂]-cytidine

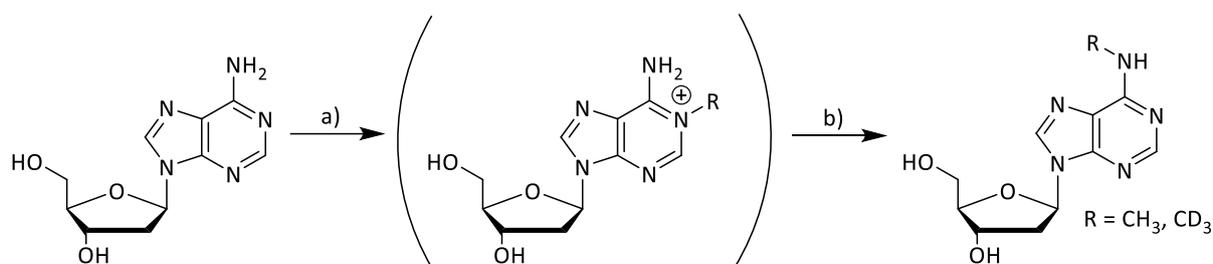
O^{3'},*O*^{5'}-TBS-2'-deoxy-4-methyl-[*N*¹,*N*³-¹⁵N₂]-cytidine (30 mg, 0.064 mmol, 1.0 eq) was dissolved in EtOAc (3 mL), HF in pyridine (40 μL, 1.4 mmol, 22.5 eq) was added and the reaction mixture was stirred at r.t. for 17 h. The mixture was cooled to 0 °C, quenched with TMS-OMe and stirred at r.t. for 4 h. The precipitation was centrifuged and the supernatant was removed and the solid was washed three times with EtOAc. The solvent was removed *in vacuo* and 2'-deoxy-4-methyl-[*N*¹,*N*³-¹⁵N₂]-cytidine (10 mg, 0.041 mmol, 64%) was obtained as a white solid.

¹H NMR (599 MHz, D₂O) δ 7.68 (d, *J* = 7.4 Hz, 1H, C6-H), 6.29 (d, *J* = 7.7 Hz, 1H, C1'-H), 5.99 (dd, *J* = 7.8, 3.9 Hz, 1H, C5-H), 4.45 (h, *J* = 4.1, 3.5 Hz, 1H, C3'-H), 4.05 (m, 1H, C4'-H), 3.84 (m, 1H, C5'-H), 3.77 (m, 1H, C5'-H), 2.90 (d, *J* = 3.3 Hz, 3H, CH₃), 2.41 (m, 1H, C2'-H), 2.31 (q, *J* = 7.3 Hz, 1H, C2'-H) ppm.

^{13}C NMR (101 MHz, D_2O) δ 164.4 (d, $J = 9.7$ Hz, C4), 157.8 (C2), 139.2 (d, $J = 13.2$ Hz, C6), 97.3 (C5), 86.4 (C4'), 85.7 (d, $J = 11.7$ Hz, C1'), 70.6 (C3'), 61.3 (C5'), 39.0 (C2'), 27.0 (d, $J = 2.0$ Hz, CH_3) ppm.

HR-MS (ESI+): m/z calculated for $[\text{C}_{10}\text{H}_{16}\text{N}^{15}\text{N}_2\text{O}_4]^+$ ($[\text{M}+\text{H}]^+$): 244.1076, found: 244.1078.

Synthesis of 2'-deoxy-6-methyladenosine



Scheme S3: Synthesis of 2'-deoxy-6-methyladenosine and 2'-deoxy-6-[D_3]-methyladenosine.

a) Methyl-*p*-toluenesulfonate, DMF, r.t., 20 h ($\text{R}=\text{CH}_3$); CD_3I , *N,N*-dimethylacetamide 28 °C, 20 h ($\text{R}=\text{CD}_3$), b) 2 M NaOH, 28 °C, 1 h, 13% ($\text{R}=\text{CH}_3$), 14% ($\text{R}=\text{CD}_3$).

2'-Deoxy-6-methyladenosine was synthesized as published in ^[1].

Methyl-*p*-toluenesulfonate (600 mg, 3.22 mmol, 4.0 eq) was dissolved in 1 mL *N,N*-dimethylformamide and 2'-Deoxyadenosine (200 mg, 0.796 mmol, 1.0 eq) was added. The solution was stirred at r.t. for 20 h. Celite was added and the solution was filtered. Acetone (30 mL) was added to the filtrate and the mixture was stirred for 1 h. The precipitation was filtered, washed with acetone and dried on high vacuum. The white solid was dissolved in 5 mL 2 M NaOH and stirred at 28 °C for 1 h. The solution was neutralized with 10% aqueous *p*-toluenesulfonic acid and the solvent was removed *in vacuo*. The residue was purified via column chromatography (DCM/MeOH 10:1 → 9:1, SiO_2) and 2'-deoxy-6-methyladenosine (27 mg, 0.10 mmol, 13%) was obtained as a white solid.

^1H NMR (800 MHz, $\text{DMSO-}d_6$) δ 8.32 (s, 1H, C8-H), 8.22 (s, 1H, C2-H), 7.77 (s, 1H, NH), 6.35 (dd, $J = 7.9, 6.2$ Hz, 1H, C1'-H), 5.30 (d, $J = 4.0$ Hz, 1H, C3'-OH), 5.24 (t, $J = 5.9$ Hz, 1H, C5'-OH), 4.41 (dq, $J = 6.3, 3.1$ Hz, 1H, C3'-H), 3.88 (q, $J = 3.8$ Hz, 1H, C4'-H), 3.62 (dq, $J = 13.2, 4.4$ Hz, 1H, C5'-H), 3.52 (ddt, $J = 10.8, 7.7, 4.0$ Hz, 1H, C5'-H), 2.95 (s, 3H, CH_3), 2.72 (ddd, $J = 13.4, 7.7, 5.6$ Hz, 1H, C2'-H), 2.26 (ddd, $J = 13.2, 6.2, 2.8$ Hz, 1H, C2'-H) ppm.

^{13}C NMR (201 MHz, $\text{DMSO-}d_6$) δ 155.0 (C6), 152.4 (C2), 147.9 (C4), 139.3 (C8), 119.8 (C5), 88.0 (C4'), 84.0 (C1'), 71.0 (C3'), 61.9 (C5'), 39.4 (C2'), 27.0 (CH_3) ppm.

HR-MS (ESI+): m/z calculated for $[\text{C}_{11}\text{H}_{16}\text{N}_5\text{O}_3]^+$ ($[\text{M}+\text{H}]^+$): 266.1248, found: 266.1248.

Synthesis of 2'-deoxy-6-[D₃]-methyladenosine

2'-Deoxyadenosine (200 mg, 0.796 mmol, 1.0 eq) was dissolved in 1 mL *N,N*-dimethylacetamide and CD₃I (0.15 mL, 2.4 mmol, 3.0 eq) was added. The yellow solution was stirred at 28 °C for 20 h. Acetone was added and the precipitation was filtered, washed with acetone and dried on high vacuum. The white solid was dissolved in 5 mL 2 M NaOH and stirred at 28 °C for 1 h. The solution was neutralized with 10% aqueous *p*-toluenesulfonic acid and the solvent was removed *in vacuo*. The residue was purified via column chromatography (DCM/MeOH 10:1 → 9:1, SiO₂) and 2'-deoxy-6-[D₃]-methyladenosine (29 mg, 0.11 mmol, 14%) was obtained as a white solid.

¹H NMR (400 MHz, DMSO-*d*₆) δ 8.32 (s, 1H, C8-H), 8.22 (s, 1H, C2-H), 7.74 (s, 1H, NH), 6.35 (dd, *J* = 7.9, 6.1 Hz, 1H, C1'-H), 5.30 (d, *J* = 4.0 Hz, 1H, C3'-OH), 5.24 (s, 1H, C5'-OH), 4.41 (dq, *J* = 6.0, 2.9 Hz, 1H, C3'-H), 3.88 (td, *J* = 4.2, 2.5 Hz, 1H, C4'-H), 3.62 (m, 1H, C5'-H), 3.52 (d, *J* = 11.6 Hz, 1H, C5'-H), 2.72 (ddd, *J* = 13.4, 8.0, 5.7 Hz, 1H, C2'-H), 2.26 (ddd, *J* = 13.1, 6.1, 2.9 Hz, 1H, C2'-H) ppm.

¹³C NMR (101 MHz, DMSO-*d*₆) δ 155.0 (C6), 152.3 (C2), 147.8 (C4), 139.2 (C8), 119.7 (C5), 88.0 (C4'), 83.9 (C1'), 70.9 (C3'), 61.9 (C5'), 39.4 (C2'), 26.4 (CD₃) ppm.

HR-MS (ESI⁺): *m/z* calculated for [C₁₁H₁₃D₃N₅O₃]⁺ ([M+H]⁺): 269.1436, found: 269.1437.

Cyanobacterial and Algal Strains, Culture Conditions and DNA Extraction

For DNA extractions, 150 mL of wild-type *Synechocystis* sp. PCC6803 cultures were grown at 30 °C in BG11 medium containing 5 mM glucose.^[2] For *Chlamydomonas reinhardtii* the cell-wall-deficient strain 3491 (*cw15 mt-*), obtained from the *Chlamydomonas* Center (<http://www.chlamy.org/>), and the cell-walled laboratory strain 7d+ were cultured in 150-300 mL Tris-acetate-phosphate medium containing 1% sorbitol^[3] at 25°C. All cultures were grown under orbital shaking and continuous irradiation with 30 μmol of photons m⁻² s⁻¹.

For determination of *N*⁶-methyldeoxyadenosine and *N*⁴-methyldeoxycytidine, DNA was isolated by harvesting cells at 3000 g for 20 min, followed by resuspension of the pellets in 4-8 mL DNA extraction buffer ((2% (w/v) cetyltrimethylammonium bromide (CTAB), 100 mM Tris-HCl, pH 8; 1.4 M NaCl; 20 mM EDTA pH 8; 2% (v/v) β-mercaptoethanol)). Cells were lysed upon addition of ~300 μL glass beads for 1 h at 65 °C and vortexed every 15 min. Proteins were removed by extraction with phenol:chloroform:isoamylalcohol 25:24:1 solution. RNA in the water phase was digested at 37 °C for 1 h using 0.1 mg/mL RNase A. RNase A was removed

by extraction with phenol:chloroform:isoamylalcohol 25:24:1 solution, followed by a chloroform/isoamylalcohol 24:1 extraction and isopropanol precipitation of the DNA. DNAs were dissolved in appropriate volumes of water at 4 °C overnight.

Preparation of mouse tissues

All mice used were on the wild-type C57-BL6 genetic background. All procedures concerning animals were performed with permission of the local authority (*Regierung von Oberbayern*). Brain and liver tissues of male wt mice (C57-BL6/N) were dissected at postnatal day 51 and immediately placed in 2 mL *Eppendorf*-tubes, snap frozen in liquid nitrogen and stored at -80 °C until use.

Genomic DNA isolation

Genomic DNA from murine tissues was extracted using the Qiagen Blood and Cell Culture DNA Midi Kit. Extraction was performed following the manufacturer's instructions for genomic DNA isolation from tissue samples. 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 desferoxamine mesylate salt (desferal, 200 µM). G2-buffer was additionally supplemented with the deaminase inhibitor tetrahydrouridine (THU, 200 µM), according to published methods, to reduce background oxidation or deamination.^[4] Elution buffer QF was supplemented with 200 µM BHT.

mESC and HeLa samples differentiated in the presence of growth factors 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 performed with 100 µl of bidest. water supplemented with BHT (0.2 µM).

Cell culture

Feeder independent wt J1 cells (strain 129S4/SvJae)^[5] were cultured in the presence of serum and LIF as previously described^[6] and routinely maintained on gelatinized plates in DMEM (PAA or Sigma) supplemented with 10% FBS, 1×MEM-nonessential amino acids (NEAA), 0.2 mM *L*-alanyl-*L*-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin (all from PAA), 0.1 mM β-mercaptoethanol, 20 ng/ml ($\geq 1 \times 10^3$ U/ml) mouse recombinant LIF (ORF Genetics), 1 µM PD 0325901 and 3 µM CHIR 99021 (2i; both from Axon Medchem). Before DNA isolation, 2i cultures were passaged twice (over 4 d and 5 d, respectively) 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.

As an additional experiment for determination of m⁶dA as a marker of early embryogenesis, we cultured wt J1 cells in serum-free 2i/LIF on GelTrex-coated plates for more than 5 passages and harvested them with RLT buffer (Qiagen) supplemented with BHT and desferal.

Feeder independent wt TT2 cells were cultured like wt J1 cells (as described above) and primed for 4 d, 10 d, 12 d and 14 d respectively. Feeding experiments with dimethylsulfate (DMS, <0.01% DMSO) were performed on day four of the priming process and the cells were incubated with DMS for 24 h.

For isotope tracing with heavy methionine in serum-primed mESCs, 2i cultures of wt J1 cells were passaged twice (over 4 d or 5 d) without 2i in L-Metfree DMEM (Sigma) supplemented as above and with 0.2 mM of either [*methyl*-¹³C,³D₃]L-Met or natural L-Met. The medium was changed after 24 h. The experiment was performed in a biological duplicate.

Feeding of wt J1 cells with m⁶dA was performed once by adding m⁶dA nucleoside (1 μM, 0.1% DMSO) to the medium on day 0 of culturing in FBS/LIF and increasing the concentration on day 2 for another two days (1 mM, 1% DMSO). DMSO controls were only treated with the respective concentration of DMSO.

HeLa cells were cultivated at 37 °C in water saturated, CO₂-enriched (5%) atmosphere. DMEM (10% FBS) was used as growing medium. When reaching a confluence of 70% to 80% the cells were passaged or harvested with RLT buffer (Qiagen) supplemented with BHT and desferal, respectively.

Digest of genomic DNA

1 μg (10 μg) of genomic DNA in 35 μl H₂O were digested as follows: An aqueous solution (7.5 μl) of 480 μM ZnSO₄, containing 42 U (63 U) nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U (7.5 U) Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards 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 0.15 U (0.2 U) snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for another 3 h (overnight) at 37 °C. Prior to LC-MS/MS analysis, samples were filtered by using an AcroPrep Advance 96 filter plate 0.2 μm Supor (Pall Life Sciences).

For dilution experiments, m⁶dA and m⁴dC were not spiked during the digest step, but after a dilution of the digest of 1:20, or 1:200 respectively.

Furthermore, a recovering experiment was performed as follows: a digest of 10 µg of murine DNA was spiked with a small amount of *Chlamydomonas* (CC-3491) DNA or synthetic m⁶dA-nucleoside, which corresponds to a level of m⁶dA as described earlier in mouse tissue.^[7]

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,^[8] was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature 80 °C, gas flow 15 l/min (N₂), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 l/min (N₂), 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 Table S1.

Exact quantification of the nucleosides m⁵dC, hmdC, fdC, cadC and 8oxodG was performed with a previously published method.^[6]

Table S1: 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 1.5–6 min								
$[^{15}N_2]$ -cadC	274.08	Wide	158.03	Wide	50	5	5	Positive
cadC	272.09	Wide	156.04	Wide	50	5	5	Positive
$[D_2-^{15}N_2]$ -hmdC	262.12	Wide	146.07	Wide	50	27	1	Positive
hmdC	258.11	Wide	142.06	Wide	50	27	1	Positive
$[D_3]$ -m ⁵ dC	245.13	Wide	129.09	Wide	55	60	1	Positive
$[^{15}N_2]$ -m ⁴ dC	244.11	Wide	128.06	Wide	55	60	1	Positive
m ⁵ dC/m ⁴ dC	242.11	Wide	126.07	Wide	55	60	1	Positive
dC	228.1	Wide	112.05	Wide	1	1	0	Positive
$[^{15}N_2]$ -cadC ₂	158.03	Wide	140.09	Wide	50	13	7	Positive
cadC ₂	156.04	Wide	138.03	Wide	50	13	7	Positive
Time segment 6-8 min								
$[D_2]$ -hmdU	259.09	Wide	216.08	Wide	48	7	5	Negative
$[D_2]$ -hmdU ₂	259.09	Wide	126.05	Wide	48	7	5	Negative
hmdU	257.08	Wide	214.07	Wide	48	7	5	Negative
hmdU ₂	257.08	Wide	124.04	Wide	48	7	5	Negative
$[^{15}N_2]$ -fdU	257.06	Wide	213.05	Wide	48	6	5	Negative
$[^{15}N_2]$ -fdU ₂	257.06	Wide	141	Wide	48	6	5	Negative
fdU	255.06	Wide	212.06	Wide	48	6	5	Negative
fdU ₂	255.06	Wide	140	Wide	48	6	5	Negative
$[^{15}N_2]$ -dU	229.06	Wide	185.06	Wide	48	5	5	Negative
dU	229.06	Wide	185.06	Wide	48	5	5	Negative
Time segment 8-12 min								
$[^{15}N_5]$ -8oxodG	289.09	Wide	173.04	Wide	60	9	7	Positive
8oxodG	284.1	Wide	168.05	Wide	60	9	7	Positive
$[D_3]$ -m ⁶ dA sensitive	269.14	Wide	153.1	Wide	55	30	1	Positive
m ⁶ dA sensitive	266.12	Wide	150.08	Wide	55	30	1	Positive
$[^{13}CD_3]$ -m ⁶ dA	270.15	Wide	154.14	Wide	55	30	1	Positive
$[D_3]$ -m ⁶ dA insensitive	269.14	Wide	153.1	Wide	55	60	5	Positive
m ⁶ dA insensitive	266.12	Wide	150.08	Wide	55	60	5	Positive
$[^{15}N_2]$ -fdC	258.09	Wide	142.04	Wide	60	5	5	Positive
fdC	256.09	Wide	140.05	Wide	60	5	5	Positive
$[^{15}N_2]$ -fdC ₂	142.04	Wide	98.04	Wide	60	13	7	Positive
fdC ₂	140.05	Wide	97.04	Wide	60	13	7	Positive

Chromatography was performed by a Poroshell 120 SB-C18 column (Agilent, 2.7 μm , 2.1 mm \times 150 mm) at 30 °C using a previously published gradient^[8] of water and MeOH, each containing 0.0075% (v/v) formic acid, at a flow rate of 0.35 ml/min. The effluent up to 1.5 min and after 12 min was diverted to waste by a Valco valve. The autosampler was cooled to 4 °C. The injection volume was amounted to 39 μl . Calibration curves are shown in Fig S4.

Method validation and data processing were performed as described in earlier published work. Calibration curves were obtained by analyzing each standard (5–8 standard concentrations) as technical triplicate and applying linear regression Origin® 6.0 (Microcal™). Therefore, the ratio of the area under the curve (A/A^*) of the natural nucleoside to the internal standard was plotted against the ratio of the amount of substance (n/n^*) of the natural nucleoside to the internal standard (see Fig. S4). Additionally, acceptable accuracy (80–120%) as well as precision (<20% RSD) was required. Accuracy was proven by calculating 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 determined 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. A compilation of absolute and relative LLOQs is shown in Table S2.

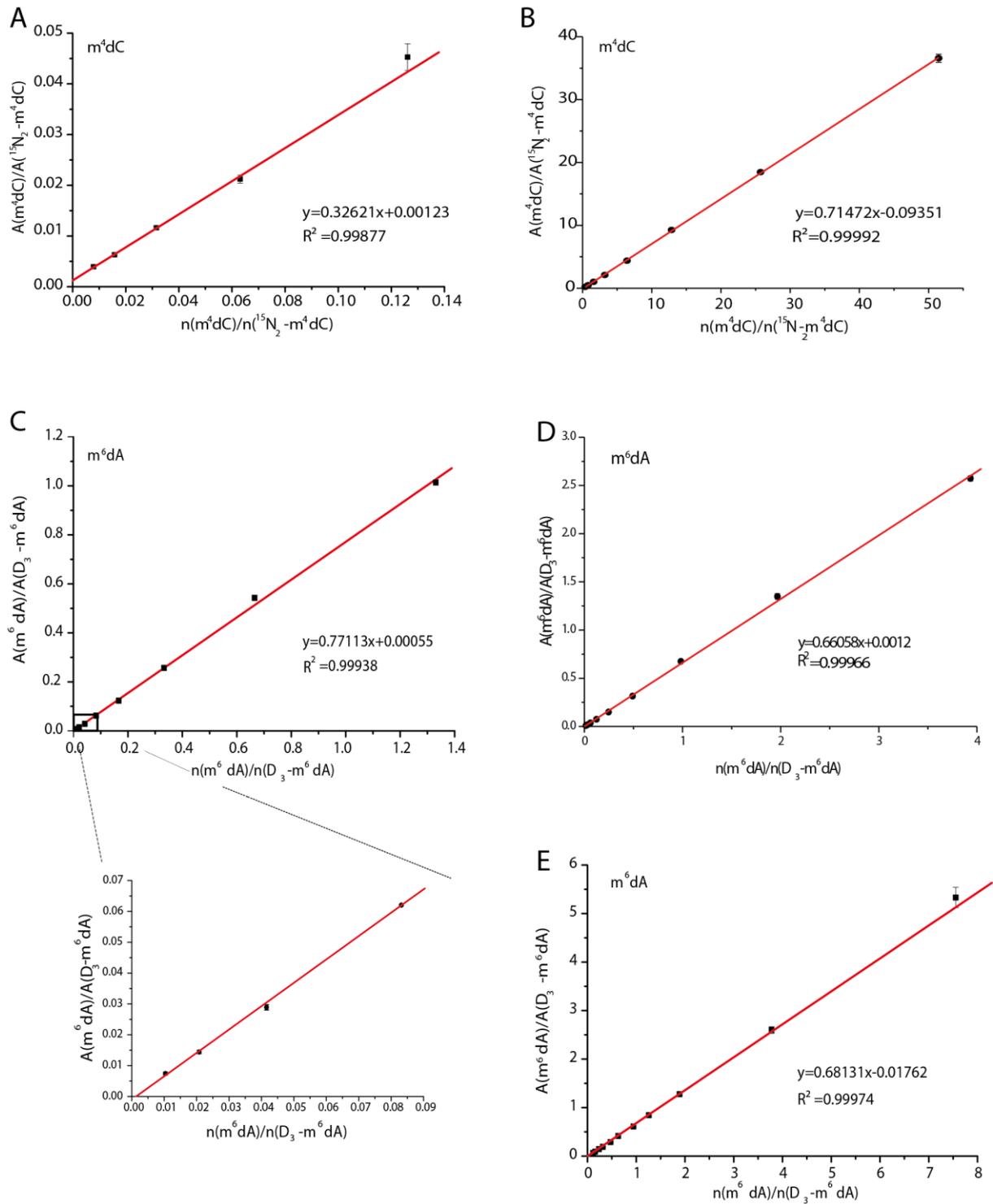


Figure S1: Calibration curves for m^4dC (upper row; A for murine, cell culture and *Chlamydomonas* DNA, B for *Synechocystis* DNA) and m^6dA (sensitive signals: C for mouse and cell culture DNA; insensitive signals: D for *Chlamydomonas* DNA and E for *Synechocystis* DNA); the ratio of the signal integral of unlabeled nucleoside towards labelled nucleoside is plotted against the ratio of the amount of natural nucleoside over the labelled nucleoside.

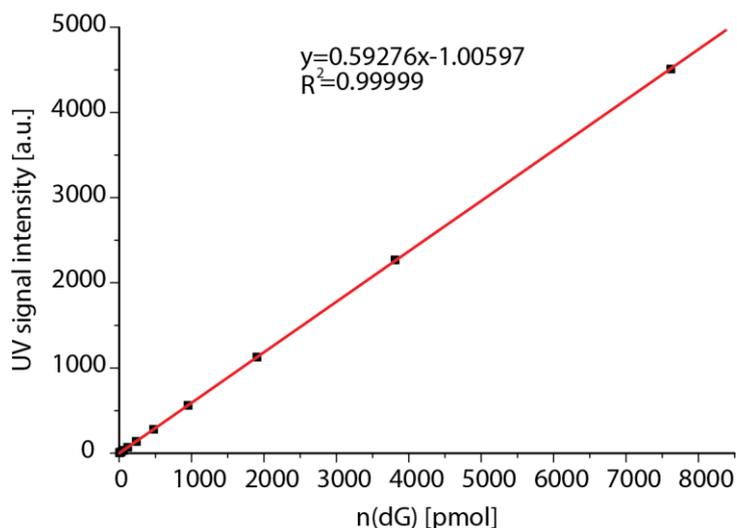


Figure S2: UV Calibration curve for dG with the UV signal intensity plotted against the amount of nucleoside [pmol].

Table S2: Compilation of absolute lower limits of quantification [fmol] (LLOQ; see Fig. S1) and relative LLOQs [per dN]. The relative LLOQs were computed by generating ratios of the absolute LLOQ [pmol] to the total amount of nucleosides (dN; [pmol]) in 10 µg of DNA (for calibration curve A and C) or 1 µg of DNA (for calibration curve B and D) or 50 ng (for calibration curve E). The total amount of nucleosides was obtained by calculating the content of dG using the respective calibration curve (see Fig. S2) and taking into account that murine DNA contains 21% dG, DNA of *Chlamydomonas* contains 32% dG and DNA of *Synechocystis* contains 23.9% dG.

	Absolute ULOQ [fmol]	Absolute LLOQ [fmol]	Relative ULOQ [per dN]	Relative LLOQ [per dN]	Relative LOD [per dN]	Absolute LOD [fmol]
m ⁴ dC (small, A)	302.1	18.8	1.1*10 ⁻⁵	6.6*10 ⁻⁷	1.4*10 ⁻⁶ (mouse/cell culture) 4.8*10 ⁻⁵ (<i>Chlamydomonas</i>)	40
m ⁴ dC (big, B)	48100	376	5.1*10 ⁻²	3.9*10 ⁻⁴	4.2*10 ⁻⁵	40
m ⁶ dA (sensitive, C)	203.8	1.6	7.4*10 ⁻⁶	5.8*10 ⁻⁸	8.0*10 ⁻⁸	2.2
m ⁶ dA (insensitive, D)	1162	18.2	1.4*10 ⁻³	2.2*10 ⁻⁵	2.0*10 ⁻⁵	17
m ⁶ dA (insensitive, E)	15500	60	2.6*10 ⁻¹	1.0*10 ⁻³	4.3*10 ⁻⁴	26

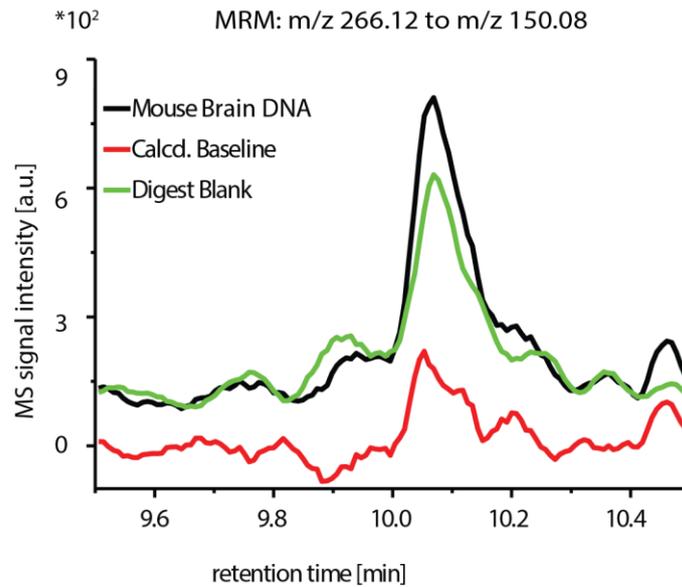


Figure S3: chromatogram of DNA from murine brain green: digest blank, in which everything was performed like in a real digest, but no DNA was added; red: calculated baseline, for which the chromatogram of the digest blank was subtracted from the chromatogram for DNA from murine brain.

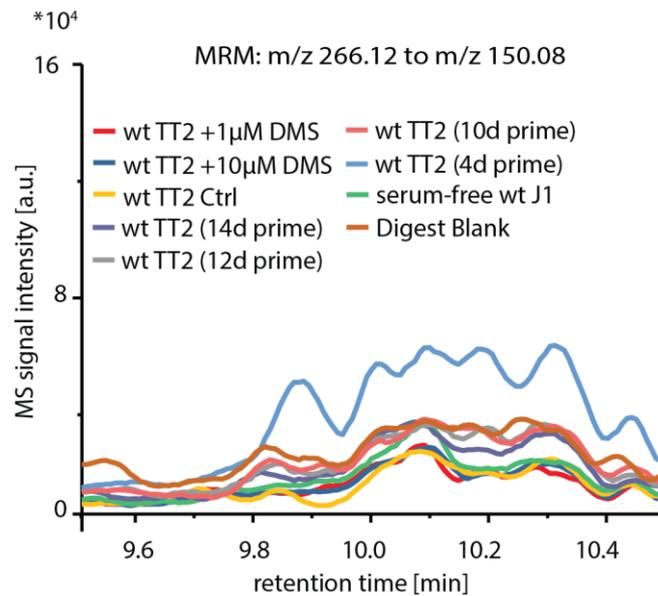


Figure S4: chromatogram of various cell culture samples; red: DNA from wt TT2 cells fed with 1 μM DMS, blue: DNA from wt TT2 cells fed with 10 μM DMS, yellow: DNA from a wt TT2 control, purple: DNA from wt TT2 cells after 14 d of priming, grey: DNA from wt TT2 cells after 12 d of priming, orange: DNA from wt TT2 cells after 10 d of priming, light blue: DNA from wt TT2 after 4 d of priming, green: DNA from wt J1 cells cultured in serum-free 2i/LIF, brown: digest blank, in which everything was performed like in a real digest, but no DNA was added

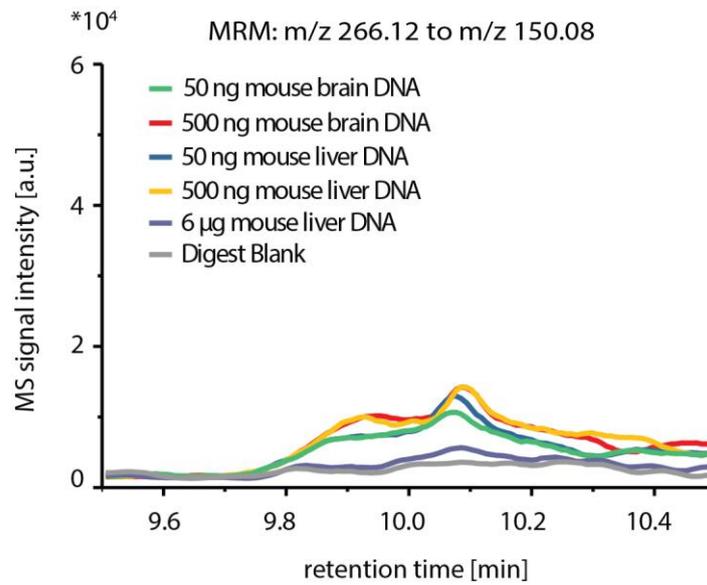


Figure S5: chromatogram of murine DNA from different tissues in different dilutions.; green: 50 ng of DNA from murine brain, red: 500 ng of DNA from murine brain, blue: 50 ng of DNA from murine liver, yellow: 500 ng of DNA from murine liver, 6 µg of DNA from murine liver, grey: digest blank, in which everything was performed like in a real digest, but no DNA was added.

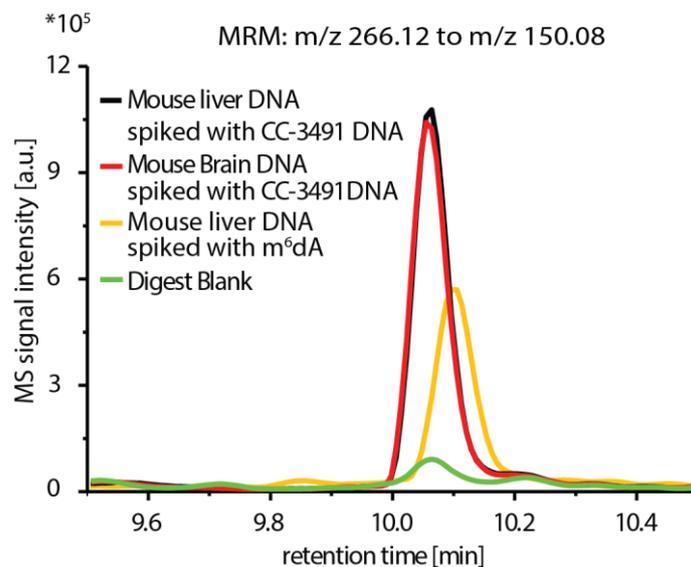


Figure S6: chromatogram of murine DNA from different tissues spiked with DNA from *Chlamydomonas reinhardtii* (CC-3491) and m⁶dA nucleoside, respectively.; black: DNA from murine liver spiked with CC-3491; green: digest blank, in which everything was performed like in a real digest, but no DNA was added, red: DNA from murine brain with CC-3491, yellow: DNA from murine liver spiked with m⁶dA nucleoside. The shown peak seems shifted in contrast to the other two peaks, but this is due to a measurement at a different time.

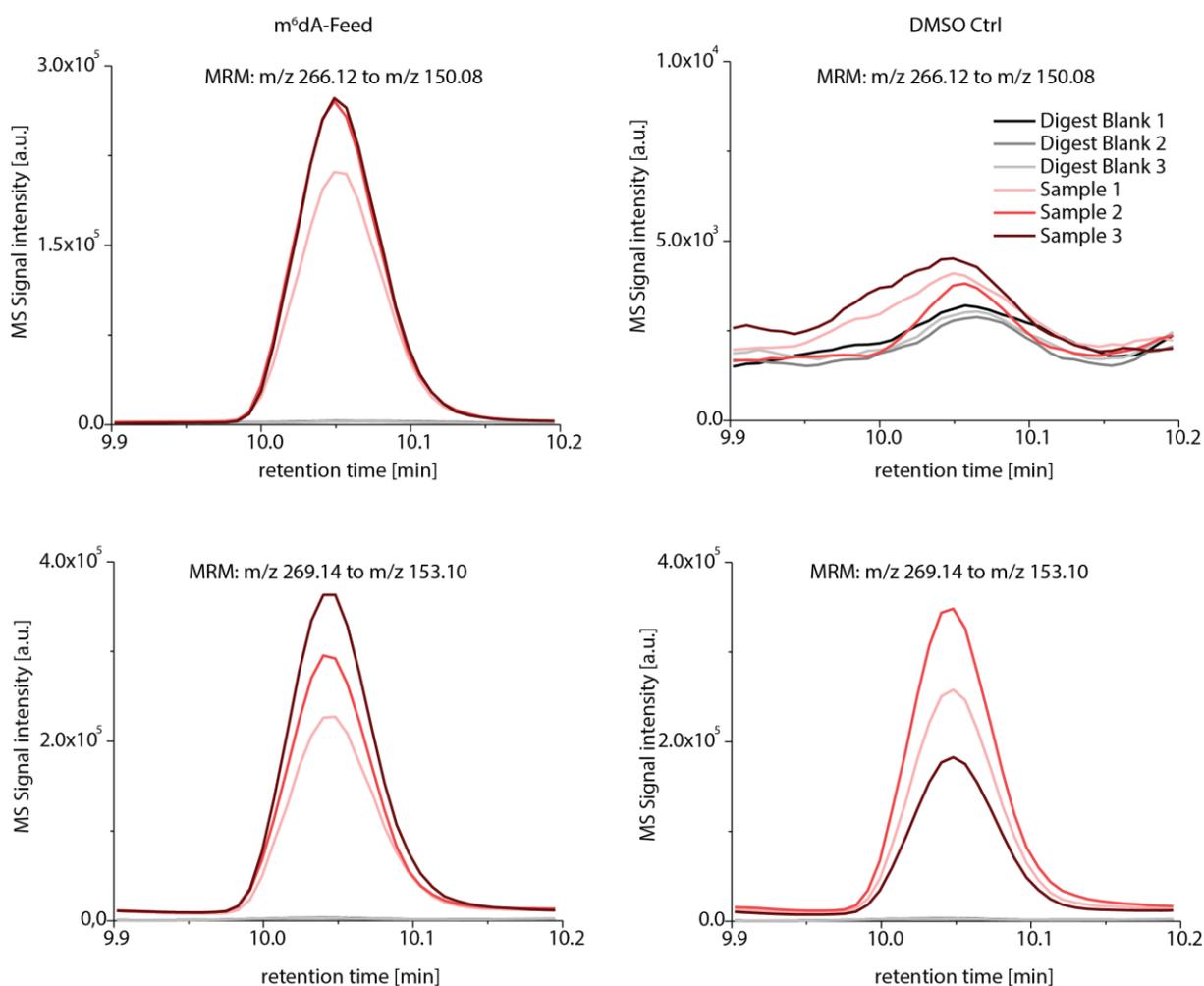


Figure S7: chromatograms for m⁶dA-fed cells (left) and the respective DMSO control (right) with the transition for the natural nucleoside (m/z 266.12 to m/z 150.08) displayed in the upper row and the transition for the spiked heavy labeled nucleoside (m/z 269.14 to m/z 153.10) displayed in the lower row. The red lines indicate the chromatogram for the measured samples, the black line indicates the respective digest blank, in which everything was performed like in a real digest, but without addition of DNA or D₃-m⁶dA).

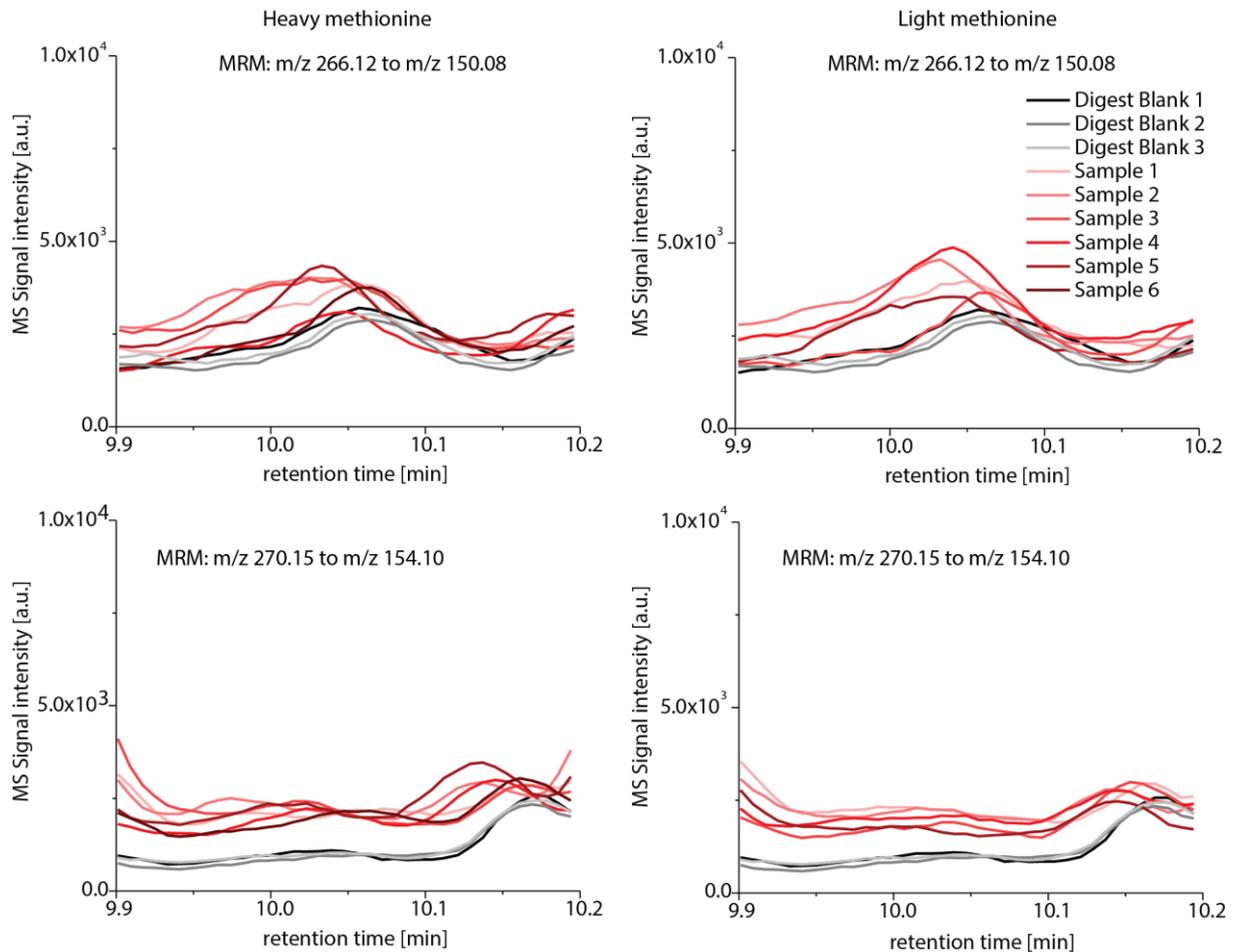


Figure S8: chromatograms for heavy methionine-fed cells (left) and the respective light methionine control (right) with the transition for the natural nucleoside (m/z 266.12 to m/z 150.08) displayed in the upper row and the transition for the heavy labeled nucleoside (m/z 270.15 to m/z 154.10; not formed) displayed in the lower row. The red lines indicate the chromatogram for the measured samples, the black line indicates the respective digest blank, in which everything was performed like in a real digest, but without addition of DNA.

- [1] J. W. Jones, R. K. Robins, *J. Am. Chem. Soc.* **1963**, *85*, 193-201.
- [2] J. D. Rosmarie Rippka, John B. Waterbury, Michael Herdman, Roger Y. Stanier, *J. Gen. Microbiol.* **1979**, *111*, 1-61.
- [3] D. S. Gorman, R. P. Levine, *Proc. Natl. Acad. Sci. U S A* **1965**, *54*, 1665-1669.
- [4] K. Taghizadeh, J. L. McFaline, B. Pang, M. Sullivan, M. Dong, E. Plummer, P. C. Dedon, *Nat. Protoc.* **2008**, *3*, 1287-1298.
- [5] E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915-926.
- [6] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeyer, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012.
- [7] M. J. Koziol, C. R. Bradshaw, G. E. Allen, A. S. Costa, C. Frezza, J. B. Gurdon, *Nat. Struct. Mol. Biol.* **2016**, *23*, 24-30.
- [8] M. Yu, L. Ji, D. A. Neumann, D. H. Chung, J. Groom, J. Westpheling, C. He, R. J. Schmitz, *Nucleic Acids Res.* **2015**, *43*, e148.

4 Unveröffentlichte Arbeiten

4.1 Entwicklung und Synthese C5- und C6-deuterierter Cytidine und deren epigenetischer Metabolismus in embryonalen Stammzellen

In diesem Projektabschnitt werden die Entwicklung und Synthese von vier verschiedenen Testmolekülen vorgestellt, welche im Rahmen dieser Dissertation durchgeführt wurden. Mit Hilfe einer Deuterierung an der C5- und C6- Position von dC sollten weitere Hinweise auf die Beteiligung eines direkten C-C-Bindungsbruchs an der aktiven Demethylierung gefunden werden. Der Hauptfokus lag dabei auf einer möglichst hohen Isotopenreinheit der an C5/C6-deuterten Verbindung. Vor allem sollte sichergestellt sein, dass keine oder eine möglichst geringe Verunreinigung an C5-¹H-dC-Spezies (**Abbildung 4.1**, blauer Kreis) vor Durchführung des Experiments vorhanden ist. Diese C5-¹H-dC-Spezies ist das Produkt einer direkten Demodifizierung über einen C-C-Bindungsbruch und könnte somit einen Hinweis auf die Beteiligung eines solchen Mechanismus liefern (**Abb. 4.1**). Daher sollte die Menge an Verunreinigung der C5-¹H-dC-Spezies, welche während der Synthese entsteht, vor Beginn des Experiments idealerweise gegen Null gehen. Dadurch entstünde ein niedriges Hintergrundsignal des eigentlich zu messenden Produkts, wodurch die massenspektrometrische Analyse einer möglichen aktiven Demethylierung mittels UHPLC-ESI-MS/MS und die Interpretierbarkeit der erhaltenen Daten deutlich erleichtert wird. **Abbildung 4.1** zeigt eine idealisierte Darstellung des durchzuführenden Experiments.

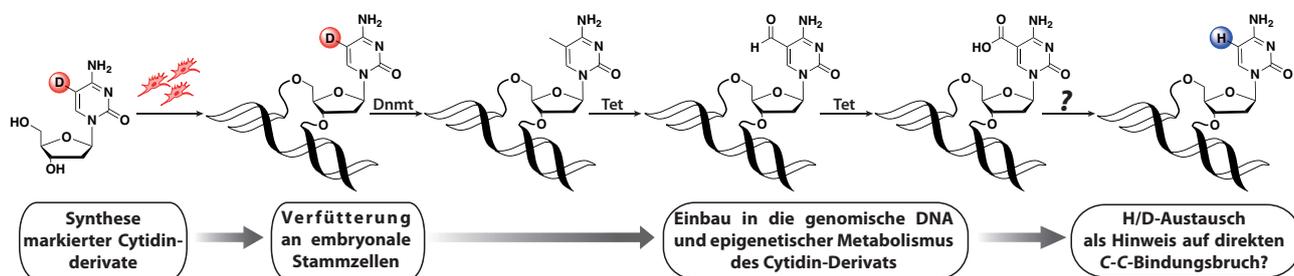
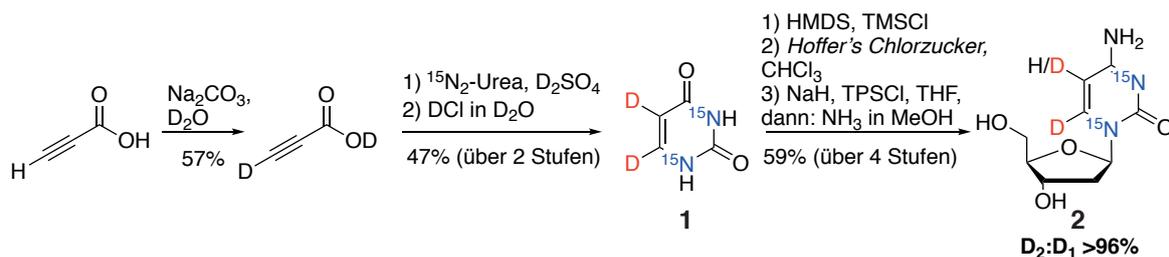


Abbildung 4.1: Die Verfütterung von an C5-markierten Cytidinderivaten (hier sei vereinfacht eine Monodeuterierung gezeigt) könnte durch Nachweis einer undeuterten Spezies (blau markiert) Hinweise auf die Beteiligung eines direkten C-C-Bindungsbruchs bei der aktiven Demethylierung liefern.

Außerdem muss die Massendifferenz des zugegebenen dC-Derivats groß genug gewählt sein, so dass sich die erwarteten Signale von denen des natürlich vorkommenden dCs und dessen Derivaten unterscheiden und es somit nicht zu einer Signalüberlappung bei der späteren massenspektrometrischen Analyse mittels UHPLC-MS/MS kommt.

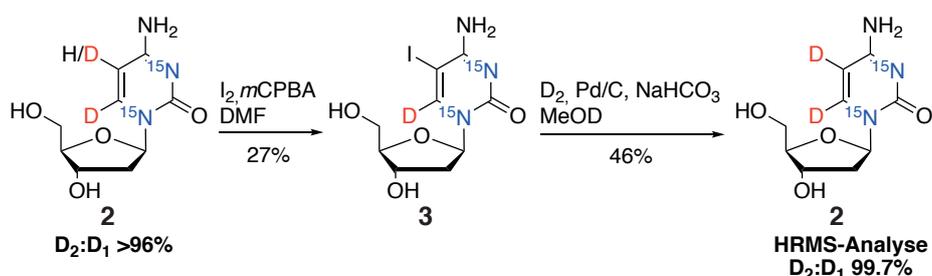
In einer ersten Generation deuterierter dC Derivate wurde zusammen mit *Arne Schröder* ein Testmolekül entwickelt, welches durch die Einführung stabiler Isotopologe um 4u schwerer war als natürlich vorkommendes dC („[M+4]⁺“). Durchläuft das Nucleosid den in **Abbildung 4.1** gezeigten Metabolismus,

so entsteht nach C-C-Bindungsbruch ein [M+3]-dC Fragment, dessen natürliches Vorkommen gegen Null geht. Die von *Arne Schröder* entwickelte Synthese ist in **Schema 4.1** gezeigt.



Schema 4.1: Synthese von D₂-¹⁵N₂-dC.

Ausgehend von Propiolsäure wurde hier unter basischen Bedingungen durch Rühren in D₂O das C3-Wasserstoff-Atom der Propiolsäure mit Deuterium ausgetauscht, wobei nach mehreren Iterationen des Vorgangs ein kompletter H/D-Austausch erreicht werden konnte (ermittelt *via* NMR/EI).^[199,200] Nachfolgende Zyklisierung mit ¹⁵N₂-Harnstoff und D₂SO₄ lieferte zunächst ein Gemisch aus D₁- und D₂-¹⁵N₂-Uracil **1**,^[201] welches zur weiteren Deuterierung in konzentrierter DCl in D₂O erhitzt wurde. Dabei konnte mittels NMR festgestellt werden, dass für die C6-Position bei **1** kein ¹H-Signal mehr erkennbar war, lediglich die C5-Position zeigte eine noch nicht vollständig erfolgte Deuterierung. Da in der Literatur bereits Deuterierungen an der C5-C6-Position am freien Nucleosid beschrieben sind,^[202-204] wurde das nicht vollständig deuterierte Uracil **1** in einer Glykosidierungsreaktion mit Hoffer's Chlorzucker umgesetzt, mittels TPSCl und darauffolgender Behandlung mit methanolischen Ammoniak zum dC-Derivat aminiert und gleichzeitig entschützt. Das somit erhaltene freie Nucleosid **2** wurde von *Jessica Steinbacher* massenspektrometrisch auf seine isotopische Reinheit geprüft und dabei konnte ein Verhältnis [M+4]/[M+3] von ca. 96% ermittelt werden. Da dies einen relativ hohen Hintergrund an zu beobachtender [M+3]-Spezies von 4% im unter **Abbildung 4.1** geplanten Experiment bedeuten würde, sollte in einem letzten Schritt eine weitere Anreicherung der doppelt deuterierten [M+4]-Spezies erreicht werden. Dazu wurde in einem ersten Versuch mittels Pd/C unter D₂-Atmosphäre bei einer Temperatur von 120 °C ein nachträglicher H/D-Austausch angestrebt. Dabei wurde jedoch ausschließlich eine vollständige Desaminierung des dC-Derivats **2** zu einem dU-Derivat beobachtet. In Zusammenarbeit mit *Arne Schröder* wurde daraufhin eine Derivatisierung der C5-Position angestrebt, um restliches ¹H zunächst zu funktionalisieren, dabei zu eliminieren und in einem späteren Schritt mit Deuterium auszutauschen. In diesem Zusammenhang bietet sich die Iodierung der C5-Position an, da diese erstens am freien Nucleosid und zweitens selektiv an dieser Position möglich ist.^[205,206] Unter Verwendung von *m*CPBA und I₂ konnte 5-Iodo-D₁-¹⁵N₂-dC erhalten werden, welches mittels Säulenchromatographie aufgereinigt und mit einer Ausbeute von 29% in hoher Reinheit erhalten werden konnte. Schließlich wurde das an C5-funktionalisierte Iodid **3** mit Hilfe von Pd/C in einer D₂-Atmosphäre (Normaldruck) vollständig mit Deuterium ausgetauscht (**Schema 4.2**).^[207]



Schema 4.2: Post-Deuterierungs Strategie von **2** zur weiteren Anreicherung von Deuterium an *C5*.

Zur Kontrolle der damit erzielten Isotopenreinheit und Anreicherung an der *C5*-Position von **2** wurde das post-deutериerte Nucleosid mittels NMR (**Abb. 4.2 A**) und hochauflösender Massenspektrometrie (**Abb. 4.2 B**) analysiert.

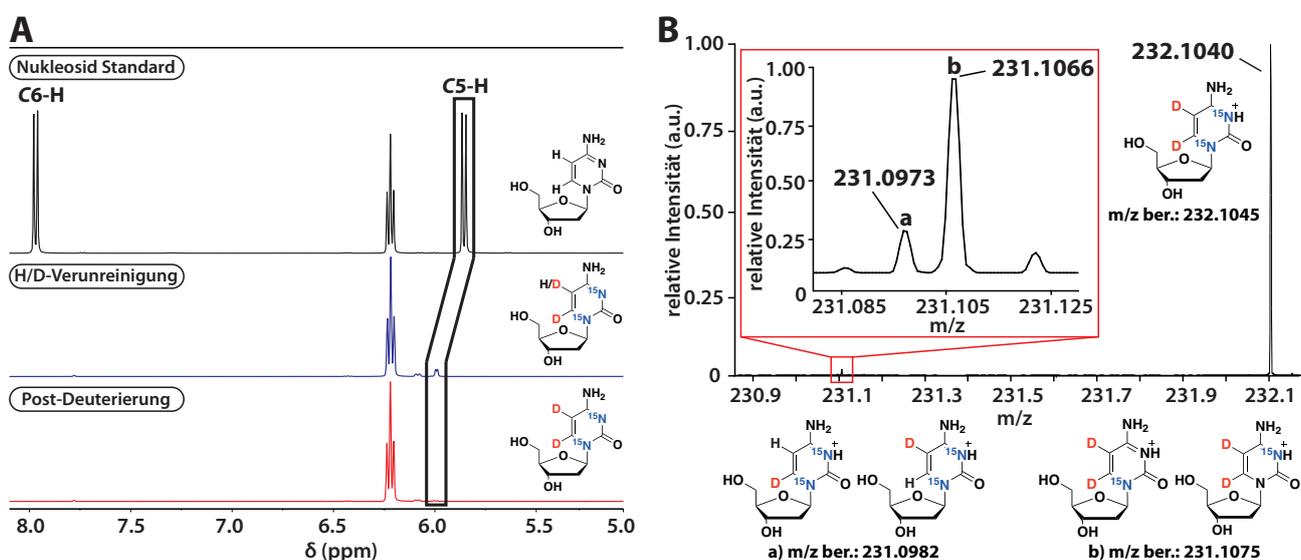


Abbildung 4.2: Analyse zur Post-Deuterierung der *C5*-Position von **2**. **A** Vergleich des NMR-Spektrums von dC, **2** vor und nach der Deuterierung über das Iodid **3**. **B** Hochauflösende Massenspektrometrie.

Anhand des NMRs ist klar zu erkennen, dass die Strategie zur Funktionalisierung und darauffolgendem Halogen-Deuterium-Austausch zur vollständigen Deuterierung der *C5*-Position führt, da das Restsignal von **2** (Kasten blaues Spektrum) nach der in **Schema 4.2** beschriebenen Synthese nicht mehr sichtbar war (Kasten rotes Spektrum). Somit kann gefolgert werden, dass auf diese Art und Weise eine selektive Deuterierung mit hoher Isotopenreinheit an *C5* erreicht werden kann. Zur genauen Bestimmung der Zusammensetzung des Testmoleküls **2** wurde außerdem ein hochaufgelöstes Massenspektrum der Substanz an einer *LTQ Orbitrap* aufgenommen (siehe **Abb. 4.2 B**). Das Signal bei $m/z = 232.1040$ war dem gewünschten Produkt mit doppelter Deuterierung an *C5-C6*-Position zuzuordnen und zeigte, dass diese Spezies das Hauptprodukt darstellte. Außerdem waren kleine Signale bei $m/z = 231.0973$ und 231.1066 messbar, welche von einer noch nicht vollständig deuterierten (**a**), beziehungsweise Mono- ^{15}N -substituierten Spezies (**b**) stammen. Das reine D_2/D_1 -Verhältnis konnte somit mit 99.7% bestimmt

werden. Da jedoch das bei der späteren Analyse verwendete Triplequadrupol-Massenspektrometer lediglich ein Auflösungsvermögen im niederen Bereich besitzt (ca. 0.4 amu), konnte aufgrund der kommerziell erhältlichen geringeren Isotopenreinheit des $^{15}\text{N}_2$ -Harnstoffs (angegeben mit >98%) im Triplequadrupol (QQQ) nur ein [M+4]/[M+3]-Verhältnis von $\sim 97\%$ erreicht werden. Im Falle eines D/H-Austausches an Spezies (**b**) würde ein [M+2]-Fragment entstehen, welches für die gewünschte Beobachtung eines [M+4]- zu [M+3]-Fragments im Rahmen einer aktiven Demethylierung über einen C-C-Bindungsbruch im QQQ-Experiment unproblematisch ist. Eine Detektion von aus **2** stammender, neu gebildeter [M+3]-Spezies im Rahmen des mdC-Metabolismus sollte damit über dem Hintergrundsignal quantifizierbar sein. Daher wurde in einem nächsten Schritt **2** (100 μM) von *Fabio Spada* an embryonale Stammzellen der Maus (mESC) verfüttert. Dabei wurde sowohl eine heterozygote (Tdg $^{+/-}$) als auch eine Tdg *Knockout* (KO) Zelllinie (Tdg $^{-/-}$) verwendet. Der Hintergrundgedanke hierbei war, dass der in der Literatur beschriebene BER-*Pathway*, bei dem fdC komplett ausgeschnitten wird, im Tdg KO nicht stattfinden kann und somit nur der antizipierte Demethylierungsweg über einen C-C-Bindungsbruch beobachtet werden könnte.^[124,166,173] Um eine möglichst hohe Markierung der genomischen DNA mit **2** zu erreichen, wurde die Fütterung des Nukleosids weiterhin über mehrere Tage durchgeführt und an Tag 0 für drei Tage ausdifferenziert. Schließlich wurde die DNA zusammen mit *Jessica Steinbacher* enzymatisch total bis auf die Nukleosidebene verdaut und die verschiedenen DNA Modifikationen mittels UHPLC-MS/MS quantifiziert. Zunächst wurde der erreichte Markierungsgrad von **2** in genomischer DNA mit 39% beziffert. Somit war mehr als jedes dritte dC durch das isotope markierte **2** ausgetauscht, was eine genaue Untersuchung der Zusammensetzung an [M+4]- und [M+3]-Spezies erlaubt. In **Abbildung 4.3** sind die gemessenen Level an dC-Derivaten gezeigt. Bei direktem Vergleich zwischen natürlichen und isotope markierten dC-Derivaten wird sofort deutlich, dass sich das Testmolekül **2** nach Verfütterung an Stammzellen in Bezug auf Methylierung und darauffolgende epigenetische Oxidation analog zum natürlich vorkommenden dC verhält. Das eingebaute [M+4]-Nukleotid lag im Wildtyp ferner mit 3.9% zu mdC+3 methyliert, mit 0.7% zu hmdC+3 hydroxymethyliert und insgesamt mit 0.02% zu fdC+3 formyliert vor. Diese Werte sind vergleichbar mit den in der Literatur bekannten.^[125] Folglich können durch Deuterium und $^{15}\text{N}_2$ hervorgerufene Isotopeneffekte bei der Interpretation der erhaltenen Daten vermutlich vernachlässigt werden.

Betrachtet man zunächst die mdC und mdC+3 Werte, so fällt auf, dass sich die Werte während der Differenzierung um ca. das 6-fache erhöhen, was vermutlich auf die erhöhte epigenetische Kontrolle der Genexpression im Laufe dieses Vorgangs zurückzuführen ist. Die hmdC und hmdC+3 Level bleiben im Wildtyp über den Zeitraum stabil, wohingegen in der KO-Zelllinie nach drei Tagen eine deutliche Akkumulation dieser Modifikation ersichtlich wird. Diese Beobachtung bestätigt sich auch in den fdC und fdC+3 Werten des Tdg *Knockouts*, bei dem die fdC und fdC+3 Mengen schon zu Beginn der

Ausdifferenzierung in etwa doppelt so hoch sind wie im Tdg Wildtyp. Da fdC im Tdg *Knockout* nicht mehr über den BER-Mechanismus prozessiert werden kann, kommt es folglich zu einer Akkumulation dieser Modifikation, welche sich außerdem im hmdC Wert widerspiegelt. Diese Daten lassen vermuten, dass hmdC somit über fdC und nicht direkt entfernt wird.

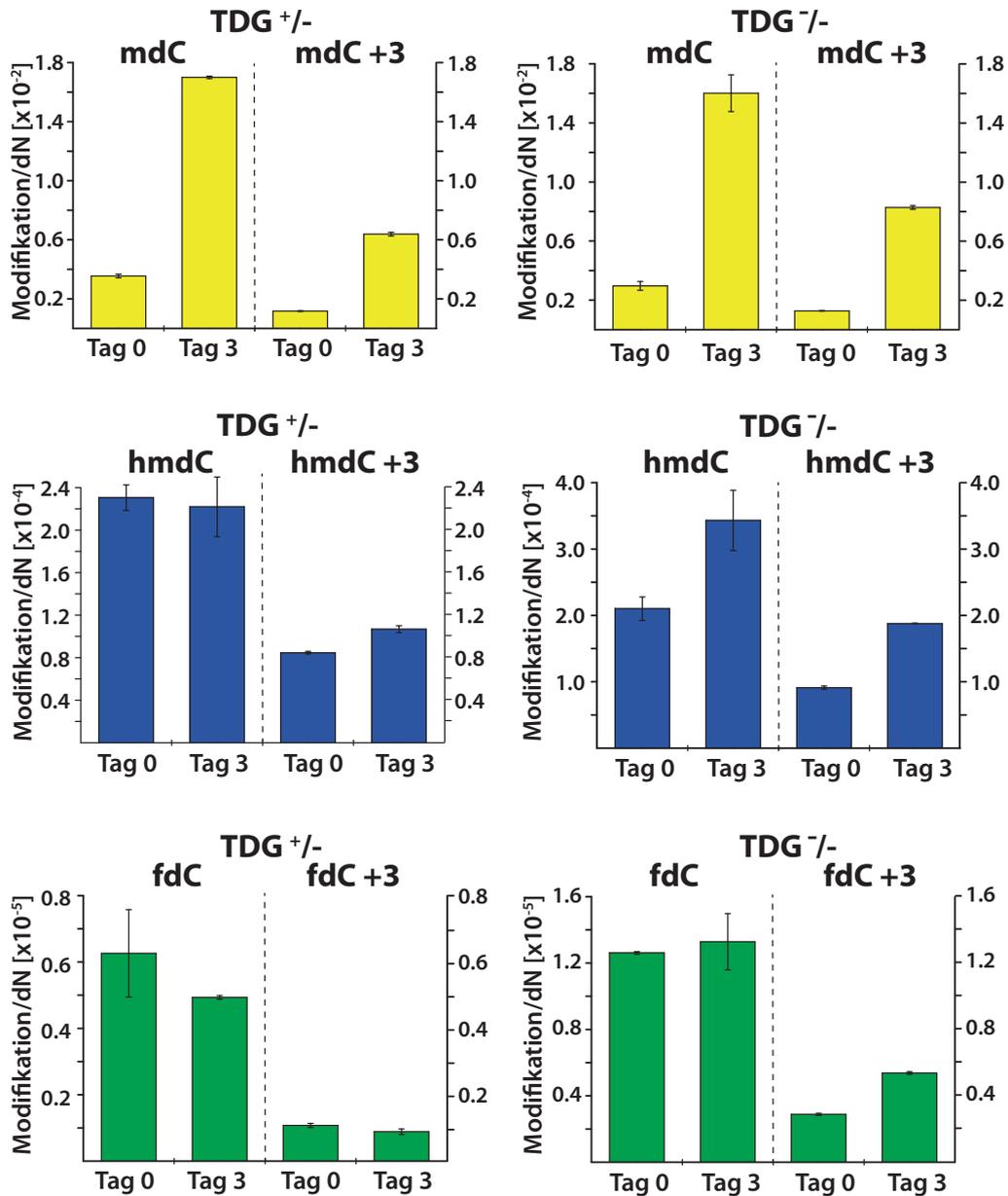


Abbildung 4.3: Quantifizierte Mengen an natürlichen und isotonenmarkierten dC Derivaten (mdC, hmdC und fdC respektive mC+3, hmdC+3 und fdC +3) in Tdg wt und KO Zelllinien nach Verfütterung von 2.

Neben den verschiedenen Modifikationen konnten im selben Experiment die Mengen der [M+4]- und [M+3]-dC Spezies in beiden Zelllinien bestimmt werden, mit welchen erste Hinweise auf eine Beteiligung eines direkten C-C-Bindungsbruchs bei der aktiven Demethylierung erhalten werden sollten. Die quantitativen Daten sind in **Abbildung 4.4** gezeigt. Die Menge an eingebautem [M+4]-dC in den

Tdg profizienten Zellen stieg von Tag 0 auf Tag 3 um ca. 8% an. Ein gleiches Bild ergibt sich für die [M+3]-dC Spezies, bei der im Wildtyp innerhalb von drei Tagen ebenfalls eine 8%-ige Erhöhung der Einbaumenge festgestellt werden kann. Somit ist davon auszugehen, dass es sich in diesem Fall beim quantifizierten [M+3]-dC lediglich um die direkt eingebaute Verunreinigung handelte und kein neues [M+3]-dC gebildet wurde. Das fdC kann an dieser Stelle auch noch über den BER-*Pathway* prozessiert werden.^[124,166,173]

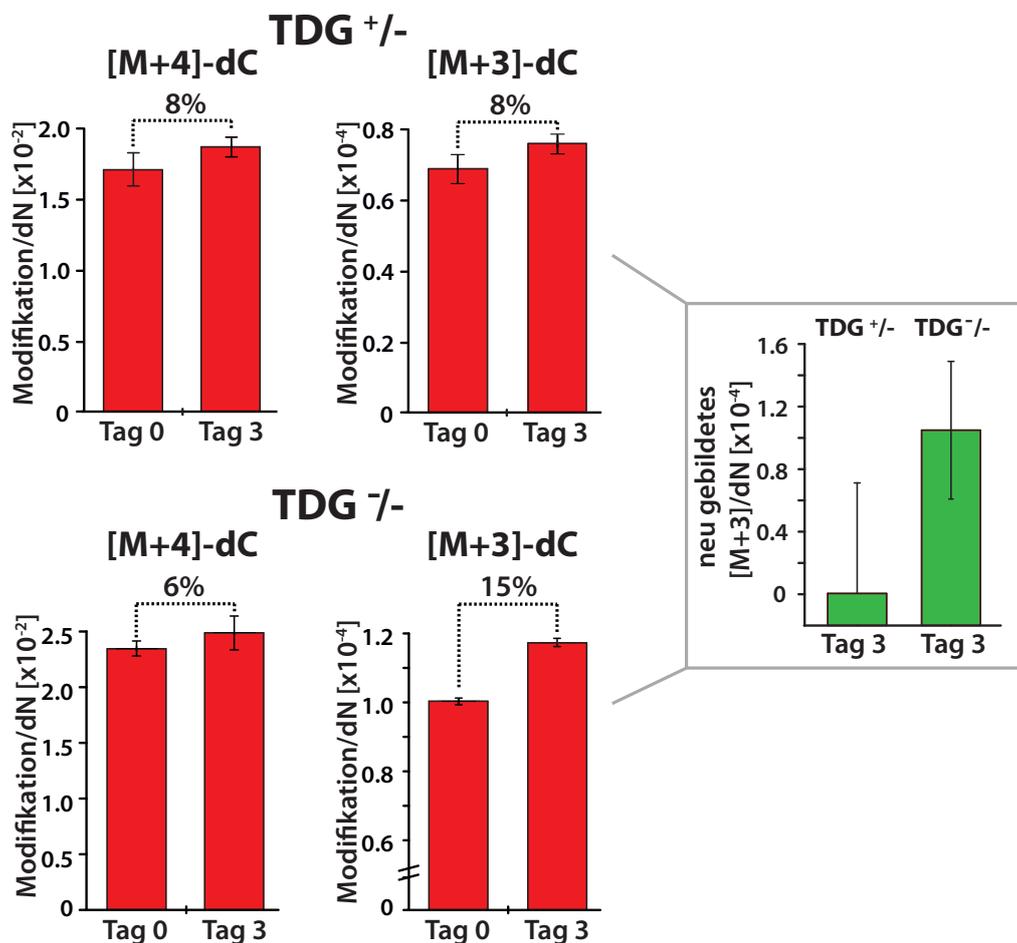


Abbildung 4.4: Quantifizierung der [M+4]- und [M+3]-Level konnte erste Hinweise auf einen direkten C-C-Bindungsbruch bei der aktiven Demethylierung zeigen.

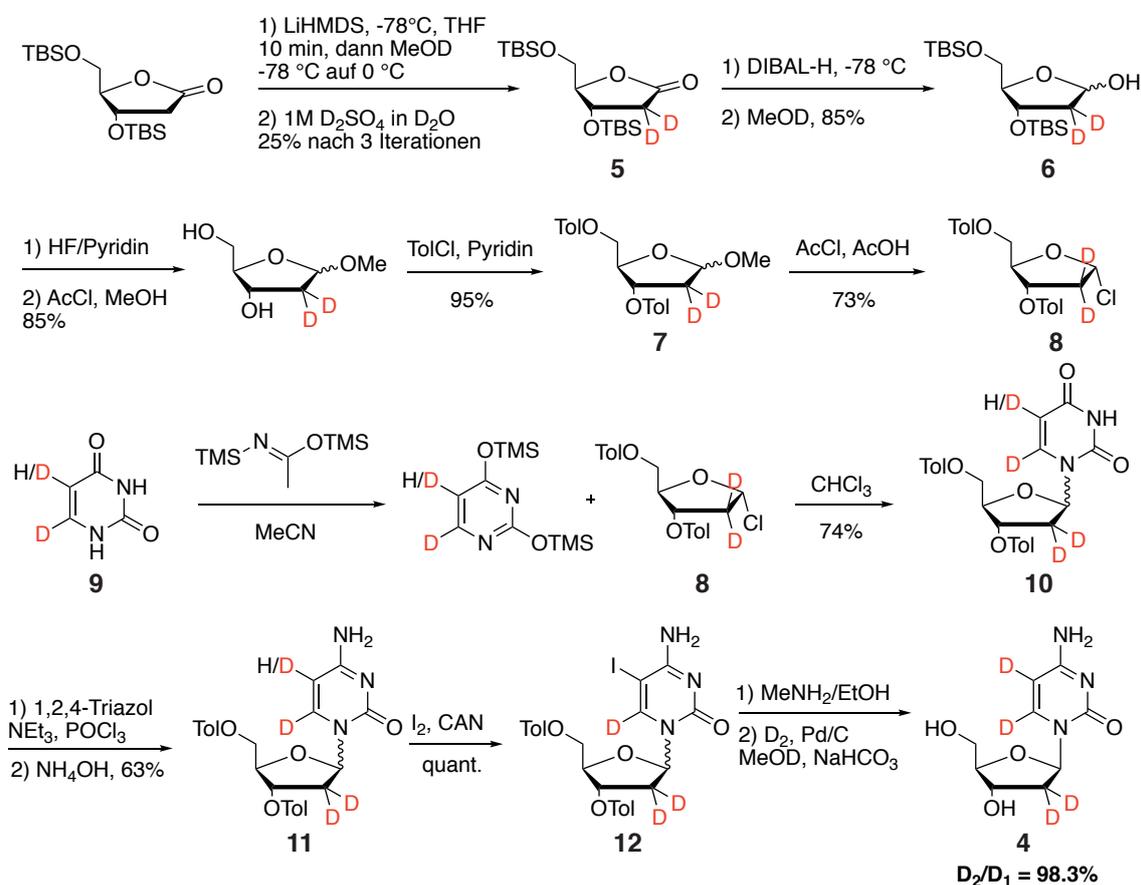
Betrachtet man nun die Werte im Tdg *Knockout*, bei dem wie zuvor beschrieben fdC nicht mehr über BER entfernt werden kann, ergibt sich ein anderes Bild. Auch hier konnte beim Einbau des [M+4]-dCs nach drei Tagen eine erhöhte Einbaumenge von ca. 6% festgestellt werden. Im Hinblick auf das quantifizierte [M+3]-Fragment sieht man im Tdg *Knockout* jedoch eine Erhöhung der [M+3]-Spezies um ca. 15%. Die Daten bestätigen somit einen Austausch von D mit H und deuten auf einen direkt erfolgten C-C-Bindungsbruch hin, wie er im Rahmen von **Abbildung 4.1** beschrieben wurde. Da sich, wie *vide supra* schon diskutiert, das isotopenmarkierte Nukleosid **2** äquivalent zu dem natürlichen, unmarkierten dC verhält, können Isotopeneffekte ausgeschlossen werden. Ein bevorzugter Einbau der Verunreinigung

in der Tdg KO Zelllinie kann ebenfalls außer Betracht gelassen werden, da man eine Erhöhung von [M+3]-Spezies sonst auch in Tdg $^{+/-}$ Zellen beobachtet hätte. Somit kann es sich hierbei um eine Erzeugung von [M+3]-Spezies durch einen C-C-Bindungsbruch im Rahmen der aktiven Demethylierung handeln. Beim Vergleich des [M+3]-dC Gehalts an Tag 3 im Wildtyp und im *Knockout* konnte somit neu gebildetes [M+3]-dC in einer Größenordnung von $1.0 \cdot 10^{-4}/\text{dN}$ ermittelt werden. Dieses Level erscheint im Vergleich mit der normalerweise im Gleichgewicht vorkommenden fdC Menge von ca. $6 \cdot 10^{-6}/\text{dN}$ relativ groß, könnte jedoch einen dynamischen Reaktionsweg der fdC-Entfernung beschreiben.

Um diese zunächst vielversprechenden Ergebnisse zu validieren, wurde das Experiment noch ein weiteres Mal durchgeführt. In diesem biologischen Replikat konnte keine Erhöhung von [M+3]-dC über den eingebauten [M+3]-Hintergrund im Tdg KO ermittelt werden (Daten nicht gezeigt). Dies könnte zum einen an unbekanntem Sekundäreffekten liegen, welche im ersten Fall eine Anreicherung oder vermehrte Bildung an [M+3]-dC begünstigten. Weiterhin handelte es sich bei diesen Experimenten um Momentaufnahmen, bei denen die Zellen nicht zu den exakt gleichen Zeitpunkten geerntet wurden. Geht man bei der aktiven Demethylierung via C-C-Bindungsbruch von einem dynamischen Vorgang aus, könnten kleine Unterschiede beim Erntezeitpunkt der Zellen eine mögliche Beobachtung dieses Vorgangs erschweren. Zum anderen ist die Detektion solcher kleinen Mengen in diesem Isotopenverfolgungsexperiment mit einer Unsicherheit behaftet, wenn über einen relativ hohen Hintergrund von schon vorhandenem [M+3]-dC, wie es bei **2** zu Beginn der Verfütterung der Fall war, eine kleinste Veränderung festgestellt werden soll.

Daher sollte in der nächsten Phase dieses Projekts die Synthese von Testmolekülen mit einer höheren Isotopenreinheit fokussiert werden. Massenspektrometrische Messungen wurden von diesem Zeitpunkt an von *Katharina Iwan* durchgeführt und damit verbundene Details sind ihrer Dissertation zu entnehmen. Ein Hauptgrund für den hohen chemischen Hintergrund an [M+3]-Fragmenten war die Tatsache, dass keine ^{15}N -Quelle mit höherer Isotopenreinheit als 98% zugänglich war. Die kommerziell am reinsten erhältlichen Isotopenquellen sind in jedem Fall deuterierte Lösungsmittel ($>99.8\%$) oder D_2 -Gas (99.96%). Das heißt, dass eine höchst mögliche Isotopenreinheit im Testmolekül vermutlich durch Deuterium-Substitution zugänglich ist. Daher wurde die Synthese eines dC-Derivats angestrebt, welches an der 2'-Position der Ribose doppelt und an der C5- und C6-Position der Base mit Deuterium substituiert sein sollte. Am Ausgangspunkt des Experiments sollte dabei eine minimale Verunreinigung an [M+3]-dC vorhanden sein, um kleinste Mengen an neu gebildetem [M+3]-dC leichter ermitteln zu können. Im Rahmen seiner Masterarbeit konnte zusammen mit *Alexander Schön* unter meiner Anleitung die Synthese eines tetradeuterierten D_4dCs (**4**) entwickelt werden, dessen Route in **Schema 4.3** vorgestellt ist. Zunächst wird hierbei das TBS-geschützte Ribose-Lakton über eine regioselektive, iterative Deprotonierung und darauffolgende Deuterierung mittels D_2SO_4 an der 2'-Position zu **5** deuteriert. Nach Reduktion mittels

DIBAL-H zum Halbacetal **6** konnte über drei Schritte die Toluoyl-geschützte Verbindung **7** und letztlich der zur Glykosylierung verwendete Chlorzucker **8** erhalten werden. Darauf folgende Silylierung des C5-C6-deuterierten Uracils **9**, welches analog zur im **Schema 4.1** gezeigten Route dargestellt werden kann, ergab das geschützte Nucleosid **10**, welches mit Hilfe von 1,2,4-Triazol und wässrigem Ammoniak zum dC-Derivat **11** umgesetzt werden konnte. Auch hier wurde zur Anreicherung der C5-Deuterierung die in **Schema 4.2** verwendete Strategie zur Funktionalisierung zum C5-Iodid angewendet. Hier konnte aufgrund der Löslichkeit des geschützten dC-Derivats **11** eine alternative, oxidative Iodierung mittels CAN anstelle von *m*CPBA angewandt werden, welche weiterhin eine quantitative Ausbeute des Iodids **12** lieferte. In einem letzten Schritt wurde das Nucleosid entschützt und mit Pd/C der zuvor gezeigte Iod-Deuterium-Austausch in D₂-Atmosphäre durchgeführt. Somit konnte **4** in einer isotopischen Reinheit von $[M+4]/[M+3] = 98.3\%$ erhalten werden.



Schema 4.3: Syntheseroute eines tetradeuterierten dC-Derivats **4**.

Damit war eine Verringerung des [M+3]-Hintergrunds im Vergleich zu **2** um ca. 1.3% erreicht. Testmolekül **4** wurde daraufhin an Tdg^{+/-} und Dnmt TKO Zellen verfüttert, wobei erneut erste Hinweise einer möglichen direkten Demethylierung über einen C-C-Bindungsbruch erhalten wurden (siehe Punkt 4.1.1). Ein Nachteil von Verbindung **4** war jedoch, dass es sich bei dem 1.3%-igen chemischen Hinter-

grundsignal entweder um eine $C5$ -undeuterierte oder $C6$ -undeuterierte Spezies handeln kann (siehe **Abbildung 4.5**). Das bedeutet, dass es während eines erfolgten $C-C$ -Bindungsbruchs auch zur Bildung einer $[M+2]$ -Spezies kommen könnte, welche dann bei der Quantifizierung zusätzlich berücksichtigt werden muss.

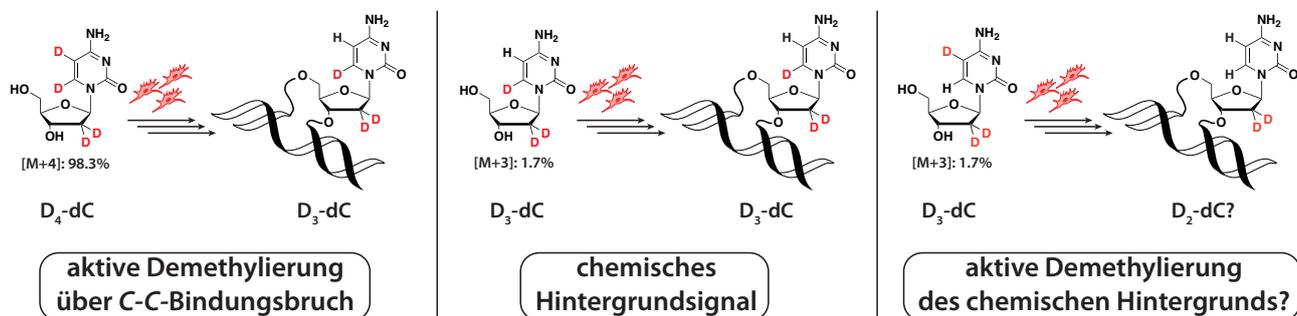


Abbildung 4.5: Mögliche massenspektrometrisch nachweisbare Produkte nach Verfütterung von **4** an mESC und deren jeweiliger Ursprung.

Dadurch wird die Interpretierbarkeit der Daten noch weiter erschwert. Daher wurde in der nächsten Phase des Projekts ein weiteres Testmolekül entworfen, um die vielversprechenden Hinweise auf eine ablaufende $C-C$ -Bindungsbruchreaktion zu untermauern.

Startpunkt war die Synthese eines neuen Testmoleküls **13**. Dieses sollte zum einen wesentlich schwerer als natürliches dC und außerdem nur spezifisch an der $C5$ -Position deuteriert sein. Daher sollte ein Derivat synthetisiert werden, welches an der Ribose komplett mit ^{13}C und an der Base mit einem $C5$ -Deuterium substituiert ist (siehe **Abbildung 4.6**). Diese Verbindung ist insgesamt um 6u schwerer als natürlich vorkommendes dC, der natürliche Hintergrund an $[M+6]$ -dC Spezies geht damit gegen Null. Ein zur Glykosylierung verwendetes $^{13}\text{C}_5$ -Ribosefragment kann retrosynthetisch aus $^{13}\text{C}_6$ -Glucose dargestellt werden.^[197] Diese ist zwar auch nur in einer Isotopenreinheit von $>99\%$ erhältlich, bei der Triplequadrupol-Messung ist jedoch nur wichtig, dass vor allem die Base an der $C5$ -Position eine hohe Deuterium-Anreicherung aufweist, weil dieses Fragment letztlich das gemessene Signal ergibt. Zur Veranschaulichung ist eine schematische Darstellung des Experiments in **Abbildung 4.6** illustriert.

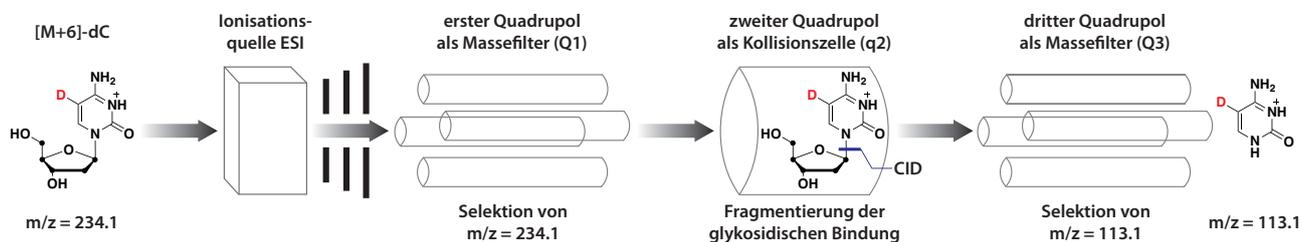
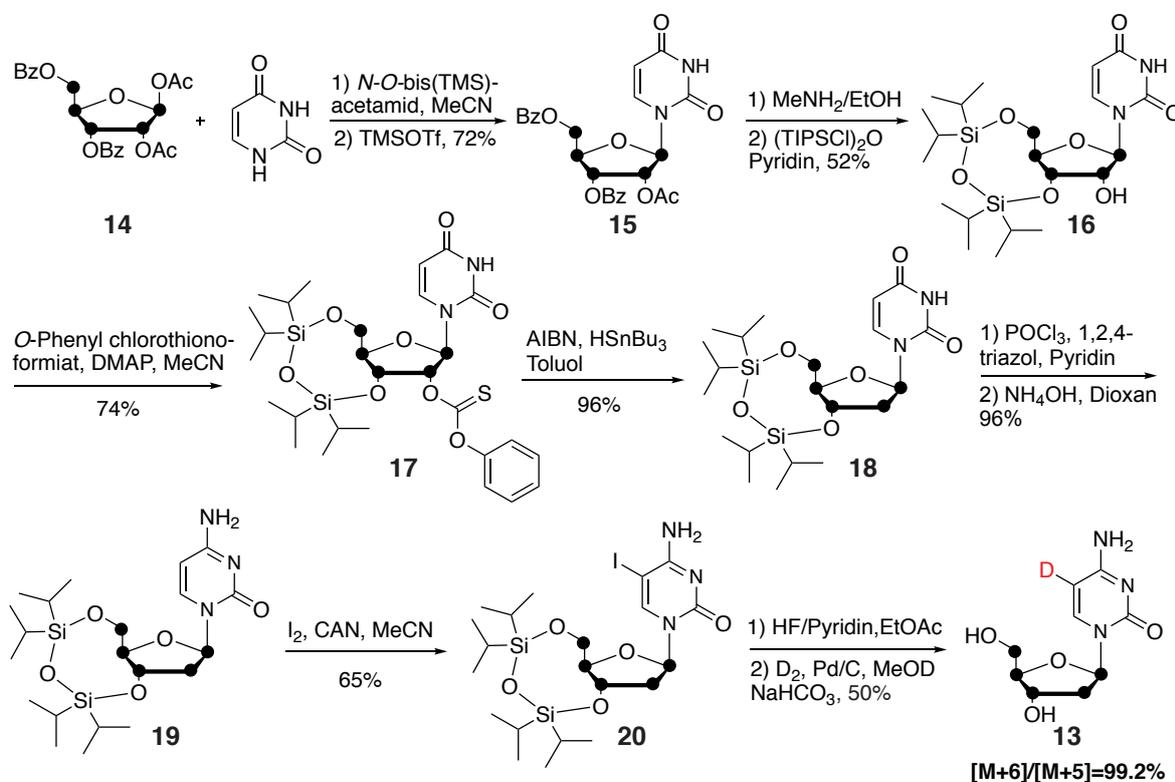


Abbildung 4.6: Schematische Darstellung des Triplequadrupol-Experiments am Beispiel der $[M+6]$ -dC Spezies **13**. • = ^{13}C -Isotope.

Zunächst wird das von der UHPLC getrennte $[M+6]$ -Fragment in der ESI-Quelle ionisiert und erreicht danach den ersten Quadrupol. Dieser selektiert nun ausschließlich die $[M+6]$ -Spezies mit $m/z = 234.1$ und

lenkt alle anderen ionisierten Verbindungen ab. Daraufhin wird die Verbindung im zweiten Quadrupol (oder auch Kollisionszelle genannt) bei einer für diesen Vorgang ermittelten, bestimmten Energie an der glykosidischen Bindung fragmentiert. Bei dieser homolytischen Spaltung entsteht dann das gut ionisierbare, an C5-deuterierte Basenfragment mit $m/z = 113.1$, welches letzten Endes im zweiten Massefilter, dem dritten Quadrupol, selektiert wird. Über *electron multiplier* (hier nicht gezeigt) wird das Signal letztlich noch amplifiziert. Das heißt, dass selbst ein komplett an C-5 deuteriertes Fragment, welches aber aufgrund der isotopischen Reinheit der verwendeten $^{13}\text{C}_5$ -Ribose von 99% zu einem Prozent als [M+5]-Spezies vorhanden ist, bei diesem gemessenen Übergang nicht zum chemischen Hintergrund dazugezählt wird. Es wird folglich nur ein Signal für ein C5-deuteriertes Basenfragment erhalten, wenn dieses aus einem [M+6]-Fragment, welches zu Beginn im ersten Massefilter isoliert wurde, stammt. Die Synthese des in **Abbildung 4.6** vorgeschlagenen Testmoleküls **13** ist in **Schema 4.4** gezeigt.



Schema 4.4: Synthese von $^{13}\text{C}_5$ -D₁-dC **13**. • = ^{13}C .

Die zur Glykosylierung benötigte $^{13}\text{C}_5$ -Ribose **14** wurde analog zur beschriebenen, literaturbekannten Syntheseroute dargestellt und ist in diesem Schema nicht extra gezeigt.^[197] Uridin wurde daraufhin mit *N,O*-bis(TMS-)acetamid silyliert und **14** in einem Ein-Topf-Verfahren mit Hilfe von TMSOTf aktiviert, wobei Uridin **15** in einer guten Ausbeute erhalten werden konnte. Ethanolisches Methylamin wurde zur Entfernung der Benzoyl- und Acetylschutzgruppen verwendet und die daraus resultierenden freien 5'- und 3'-OH-Gruppen selektiv mit der (TIPSCl)₂O-Klammer zu **16** geschützt. Die freie 2'-OH

Gruppe in **16** wurde daraufhin für die nachfolgende *Barton-McCombie*-Desoxygenierungsreaktion mit einem Chlorothionoformiat zu **17** carbonyliert und letztlich radikalisch mit Hilfe von AIBN und HSnBu_3 entfernt, wobei die silylgeschützte 2'-Desoxyuridinverbindung **18** in sehr guter Ausbeute erhalten werden konnte. Darauf folgende Aminierung unter Verwendung von POCl_3 , 1,2,4-Triazol und Austausch des damit erhaltenen Triazolats von **18** mit wässrigem Ammoniak in Dioxan erzeugte das silylgeschützte 2'-Desoxycytidin **19** ebenfalls in sehr guter Ausbeute. Durch die im **Schema 4.3** verwendete Iodierungsreaktion zur Funktionalisierung der C5-Position konnte das Iodo-dC **20** in guter Ausbeute erhalten werden, welches schließlich noch mit Hilfe von HF·Pyridin entschützt und über Pd/C in einer D_2 -Atmosphäre bei Normaldruck zum monodeutierten [M+6] **13** defunktionalisiert werden konnte. Die Isotopenreinheit von **13** wurde von *Katharina Iwan* zu 99.2% bestimmt, wobei im Vergleich zum bis dato isotopenreinsten D_4 -Testmolekül **4** eine weitere Verbesserung um 0.9% erreicht werden konnte. Der dabei vorhandene $^{13}\text{C}_5$ -dC-Hintergrund von 0.8% könnte dabei aus einer Verschleppung der Synthese stammen. Da die iodierter Verbindung **20** jedoch sehr gut vom silylgeschützten dC-Fragment **19** mittels säulenchromatographischer Aufreinigung abzutrennen war, ist dies eher unwahrscheinlich. Jedoch könnte man nach Entschützung von Iodoverbindung **20** noch zusätzlich eine HPL-chromatographische Aufreinigung vornehmen, um eine Verschleppung ausschließen zu können, was an dieser Stelle nicht durchgeführt wurde. Mit Verbindung **13** wurden weiterhin Verfütterungsexperimente an Wildtyp und Dnmt TKO-Zelllinien durchgeführt, die erneut das Potential des synthetisierten Testmoleküls für den Nachweis eines C-C-Bindungsbruchs zeigten (siehe **4.1.2**). Im Falle der Verfütterung von **13** an Tet *Knockout* Zelllinien konnte jedoch zunächst keine eindeutige C-C-Bindungsbruchaktivität der Zellen gezeigt werden. Als möglicher limitierender Faktor der Verfütterungsexperimente könnte eine geringe Einbaurrate der zur Zellkultur gegebenen Nukleoside angeführt werden. Obwohl relativ hohe Konzentrationen an **4** und **13** (100 - 200 μM) verfüttert wurden, waren jeweils nur ~25% aller natürlichen dCs durch die Testmoleküle ausgetauscht. Höhere Einbauraten der isotopenmarkierten dC-Testmoleküle, bei denen alle natürlichen dCs ausgetauscht werden, könnten die Effekte deutlicher darstellen und dadurch ein einfacher zu interpretierendes Ergebnis liefern. Ein Grund der relativ geringen Einbaurrate von **13** könnte in den verschiedenen Umwandlungsprozessen begründet sein, denen das dC-Nukleosid im löslichen Zytosol-Pool unterliegt (schematische Darstellung siehe **Abbildung 4.7**). Das freie Nukleosid kann dabei zunächst zum Monophosphat phosphoryliert, jedoch auch gleichzeitig durch Cda (Cytidin Desaminase) enzymatisch zu dU desaminiert werden. An dieser Stelle wird aktiv gebildetes dU zu dUMP phosphoryliert und kann mit Hilfe der Thymidylat Synthase (Ts) zum dT-Monophosphat methyliert und letztlich als Triphosphat in die DNA eingebaut werden. Das dC-Monophosphat kann entweder über das Diphosphat zum Triphosphat phosphoryliert und in die DNA eingebaut werden, oder durch die Desoxycytidylat Desaminase (Dctd) zu dUMP desaminiert werden, welches dann als dT in das Genom

eingebaut werden kann. Das heißt, dass es relativ viele alternative Mechanismen und Wege gibt, die stattfinden, bevor zugeführtes isotopenmarkiertes dC als solches in das Genom eingebaut wird.

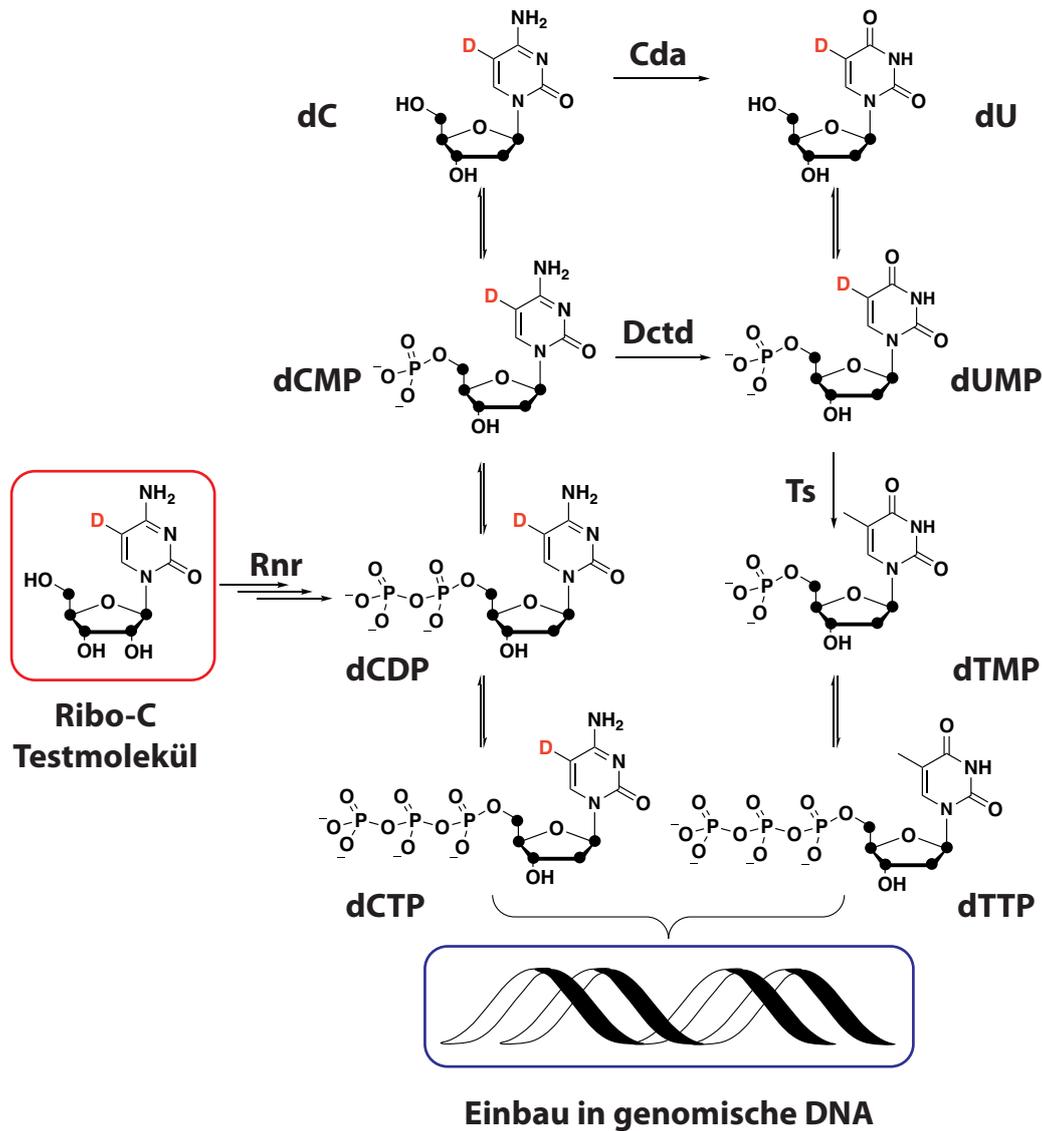
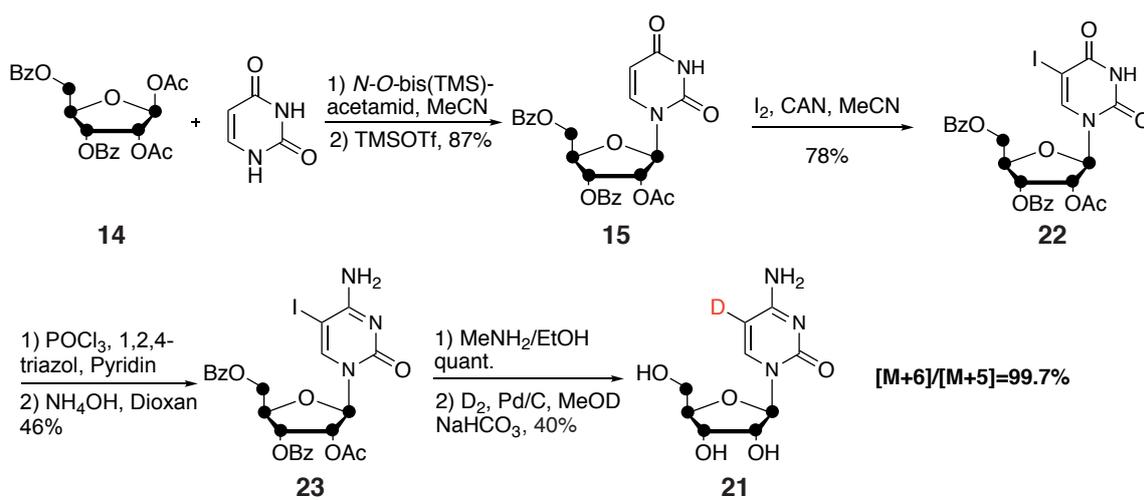


Abbildung 4.7: Verschiedene Wege der Umwandlung von dC auf Nucleosidebene vor dem Einbau in genomische DNA am Beispiel von **21** (● = ^{13}C).

Dabei wird weiterhin die Komplexität der Nucleosidmischung erhöht, was die Auswertung der Daten potentiell erschweren könnte. An diesem Punkt ist es nun wichtig zu erkennen, dass die Desoxycytidin-spezifischen Desaminasen jedoch nur auf Nucleosid- bzw. Monophosphat Ebene arbeiten. Füttert man die Zellen nun mit Ribonucleotiden, könnten diese zuvor dargestellten Desaminierungsprozesse umgangen und somit die Einbaurate von markierten Cytidinen als Testmoleküle erhöht werden. Ribonucleotide werden nämlich erst auf Diphosphat-Ebene durch die Ribonucleotidreduktase (Rnr) an der 2'-Position desoxygeniert und an dieser Stelle ist keine enzymatische Desaminierung mehr möglich. Daher könnte die Verfütterung eines RNA-Testmoleküls, welches in **Abbildung 4.7** gezeigt ist (roter Kasten), potentiell zu einem hohen Austausch von natürlichem dC in genomischer DNA führen. Die Komplexität

des erhaltenen Nucleosidgemisches würde dabei ebenfalls verringert werden. Die Synthese dieses isotonenmarkierten RNA-Derivats ist im Vergleich zur Synthese von **13** außerdem unkomplizierter, da an dieser Stelle keine *Barton-McCombie*-Desoxygenierung mehr durchgeführt werden muss und die Darstellung somit um mehrere Stufen verkürzt wird. Weiterhin sollte in diesem Fall eine Route entwickelt werden, bei der in der Synthese kein freies Cytidin in irgendeiner Art vorkommt. Eine zuvor bei **13** (*vide supra*) diskutierte, jedoch unwahrscheinliche Verschleppung, die zu einem Hintergrundsignal von undeutertem dC führen könnte, sollte so ausgeschlossen werden. **Schema 4.5** zeigt, wie dieses neue C-Derivat **21** erfolgreich synthetisiert werden konnte.



Schema 4.5: Synthese von $^{13}\text{C}_5\text{-D}_1\text{-C}$ **21**. • = ^{13}C .

Die Glykosylierungsreaktion wurde analog zu der in **Schema 4.4** gezeigten durchgeführt, allerdings wurde hier die Silylierungsreaktion mit *N,O*-bis(TMS)-acetamid nur für 10 Minuten, anstatt wie weiter oben gezeigt, für 30 Minuten durchgeführt, wobei eine höhere Ausbeute von 87% erreicht werden konnte. Darauf folgende elektrophile Iodierung mit Iod und CAN erzeugte die an *C5*-funktionalisierte Iodospezies **22** in guter Ausbeute, welche dann über das Triazolol zum Cytidin-Derivat **23** aminierte wurde. Die relativ geringe Ausbeute von nur 46% könnte in einer partiellen Entschützung der Benzoyl-Acetylgeschützten Verbindung begründet sein. Zwar konnten keine mono-, di- oder vollentschützten Spezies bei einer durchgeführten LC-MS-Analyse ermittelt werden, was auf eine durch das im Gemisch vorhandene NH_4OH erzeugte Ionensuppression zurückzuführen sein könnte. Dies erscheint jedoch unwahrscheinlich, da im nächsten Schritt **23** mit ethanolischer Methylaminlösung über Nacht voll entschützt wurde und in diesem Fall bei einer LC-MS-Analyse alle denkbaren Entschützungsfragmente detektiert werden konnten. Im letzten Schritt wurde das vollentschützte Iodo-C-Derivat mit Hilfe von Pd/C in einer D_2 -Atmosphäre mit Deuterium zu **21** isotonenmarkiert. Die Deuterium-Anreicherung an *C5* konnte von *Katharina Iwan* zu 99.7% bestimmt werden, dadurch wurde eine weitere Verbesserung der Reinheit

um 0.5% erzielt. Mit **21** wurden weiterhin Verfütterungsexperimente von *Angie Kirchner* durchgeführt. Bei Zugabe von $100 \mu\text{M}$ **21** zu Dnmt pro- und defizienten Zellen wurden 43% beziehungsweise im TKO sogar 46% aller natürlichen dCs mit dem isotope markierten Testmolekül ausgetauscht. Damit konnte gezeigt werden, dass die Verwendung eines RNA-Derivats von **13** selbst in einem nicht-optimierten Verfütterungsprotokoll schon zu einer annähernden Verdoppelung der Einbauraten führte, was für zukünftige Experimente einen großen Vorteil darstellt. Außerdem konnte mit dieser Synthese eine 99.7%-ige Deuterierung an C-5 erzeugt werden, womit weitere Untersuchungen zur Demethylierung über einen direkten C-C-Bindungsbruch verbessert werden sollten.

Zusammengefasst war es durch stetige Verbesserungen der Verfütterungsexperimente und der Weiterentwicklung der verwendeten dC-Testmoleküle möglich, erste Hinweise auf die Beteiligung eines direkten C-C-Bindungsbruches zu erhalten. Im Laufe der Umgestaltungen der Testmoleküle und Synthesen konnte die Deuterium-Anreicherung an C5 von ursprünglich 97% (**2**) auf letztendlich 99.7% (**21**) verbessert werden und die Einbauraten in das Genom der jeweiligen Zelllinien deutlich erhöht werden.

4.1.1 Zellbiologischen Daten der verfütterten Testmoleküle

Um das Potential der synthetisierten Testmoleküle zu zeigen, werden im nächsten Punkt noch einige Beispiele der nach der Verfütterung von **4**, **13** und **21** erhaltenen Daten aufgeführt. Quantifizierungen der erhaltenen Nukleosidgemische wurden von *Katharina Iwan* durchgeführt und detaillierte Ausführungen dazu sind ihrer Dissertation zu entnehmen.

Zunächst wurde **4** an Tdg^{+/-}- und Tdg^{-/-}-Zelllinien verfüttert. Die dadurch erhaltenen Daten sind in **Abbildung 4.8** gezeigt.

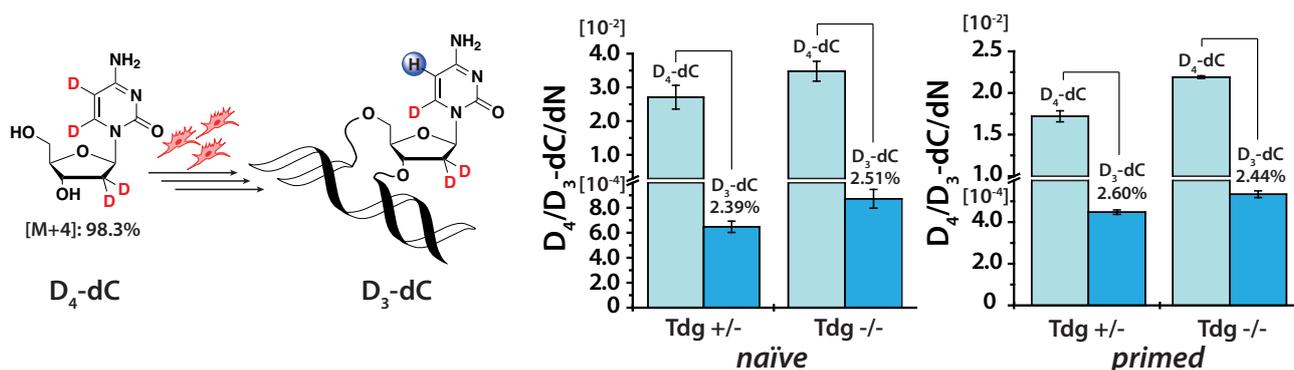


Abbildung 4.8: Quantitative Daten zur Verfütterung von **4** an Tdg pro- und defiziente Zellen. Gezeigt sind absolute Werte, die auf die Menge aller im Gemisch vorkommenden Nukleoside normiert sind.

Es konnte dabei eine Markierungseffizienz von 8-17% ermittelt werden. Bei Betrachtung des prozentualen Anteils des pro D₄-dC erhaltenen C-C-Bindungsbruchprodukts D₃-dC erkennt man, dass bei den Zellen im *naive* Zustand, bei dem eine geringere dC-Methylierungsrate im Vergleich zum *primed* Zustand

vorliegt, der Anteil von D₃-dC im Vergleich zur Tdg^{+/-}-Zelllinie höher ist. Da im Falle von Tdg^{+/-}-fdC und cadC über die BER entfernt werden können, würde man im Tdg^{-/-} einen höheren Anteil an D₃-dC erwarten, was hier gezeigt werden konnte. Außerdem ist der Anteil an D₃dC mit 2.39% (Tdg^{+/-}) beziehungsweise 2.51% (Tdg^{-/-}) höher als der Anteil von D₃-dC im Testmolekül **4**, welcher zu 1.70% bestimmt wurde. Somit scheint es, dass es bereits im *naïve* Zustand zu einem C-C-Bindungsbruch an C5 von D₄-dC kommt. Im *primed* Zustand konnte im Vergleich mit Tdg^{+/-} jedoch keine vermehrte Bildung von D₃-dC im Tdg^{-/-} gefunden werden. Diese teils widersprüchlichen Ergebnisse könnten unterschiedliche Gründe haben. Zum einen war die Markierungseffizienz bei diesem Experiment mit maximal 17% relativ gering. Somit könnte die Detektion von kleinen Veränderungen am D₃-dC-Level, welche im Rahmen einer aktiven Demethylierung entstehen würden, schwierig sein. Außerdem wurden die Zellen bei diesen Versuchen die ganze Zeit über mit **4** inkubiert. Zu den Zeitpunkten der Zellernte handelte es sich somit um Momentaufnahmen, da die Zellen durch die hohe zugeführte D₄-dC Konzentration immer wieder neues Nukleosid in ihr Genom einbauen konnten.

Wie *vide supra* gezeigt, wurde für eine weitere Untersuchung das Testmolekül **13** synthetisiert, welches neben einer großen Massendifferenz von 6u im Vergleich zu natürlichem dC auch noch eine höhere Isotopenreinheit an C5 aufwies. Außerdem wurde das Verfütterungsexperiment noch einmal überarbeitet, um die reine Beobachtung einer Momentaufnahme der ablaufenden Prozesse zu vermeiden. Eine schematische Darstellung dieses Experiments ist in **Abbildung 4.9** gezeigt.

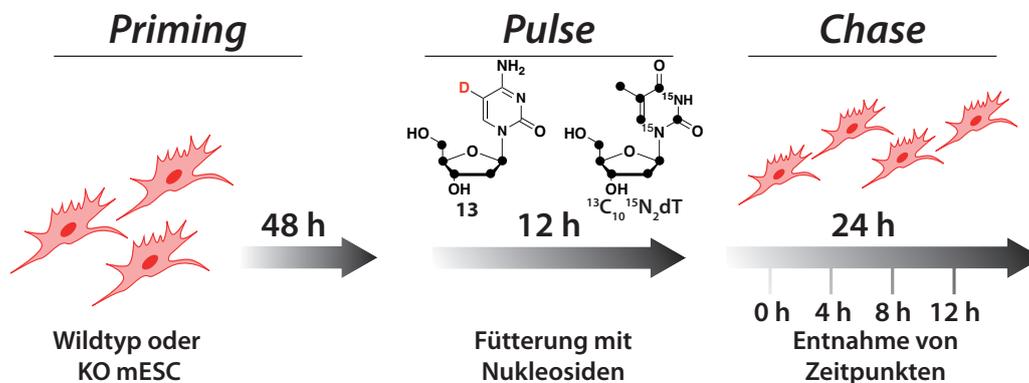


Abbildung 4.9: Schematischer Ablauf des *Pulse-Chase*-Experiments, bei dem die Zellen zunächst **13** und den internen Referenzstandard ¹³C₁₀¹⁵N₂-dT (● = ¹³C) für 12 h inkorporieren und dann zu bestimmten Zeitpunkten Proben zur massenspektrometrischen Analyse entnommen werden.

Die Zellen wurden in den nachfolgenden Experimenten zunächst für 48 Stunden in den *primed* Zustand versetzt und für eine bestimmte Zeit mit Medium, welches **13** enthielt, inkubiert. Daraufhin wurde das Medium durch frisches, nukleosidfreies Medium ausgetauscht. Zu bestimmten Zeitpunkten wurden dann Zellen abgeerntet und lysiert, sodass eine Analyse der unterschiedlichen dC-Spezies an diesen Punkten möglich war. Um etwaige Schwankungen bei der Einbau- und Metabolisierungsrate des

zugefütterten dC-Derivats **13** ausgleichen zu können, wurde außerdem ein Referenzstandard zugefüttert. Dabei handelte es sich um $^{13}\text{C}_{10}^{15}\text{N}_2\text{-dT}$, da dT zum einen beim Einbau nicht chemisch verändert wird (im Gegensatz zu dC, welches zum Beispiel auf Nukleosidebene teils enzymatisch desaminiert wird), stabil gegenüber spontaner Depyrimidinierung ist und in relativ geringen Konzentrationen eine hohe Einbaurate in genomische DNA aufweist. Somit wird der Gehalt an dT nur passiv im Rahmen der Zellteilung und nicht durch chemische oder spontane Veränderungen verringert und bietet daher einen stabilen internen Normierungsstandard. Unterschiedliche Eliminierungskinetiken der dC-Derivate, welche durch die Zellteilung bedingt sind, können dabei ebenfalls berücksichtigt werden.

13 wurde daraufhin an Dnmt Wildtyp und *Knockout* Zellen verfüttert. In Dnmt-TKO Zellen kann dC nicht zu mdC methyliert werden. Das heißt, dass epigenetische Prozesse zur Entfernung von mdC in diesem Fall nicht mehr stattfinden sollten und dieses System quasi eine Art Nullpunktsreferenz darstellt. Der Gehalt an neu gebildetem $^{13}\text{C}_5\text{-dC}$, welches das Produkt einer aktiven Demethylierung durch C-C-Bindungsbruch ist, sollte daher in den *Knockout* Zellen geringer als in den Wildtyp-Zellen sein. **Abbildung 4.10** zeigt die ersten erhaltenen quantitativen Daten. Zunächst konnte eine $^{13}\text{C}_5\text{-D}_1\text{-dC}$ Markierungsrate von 32% in den Wildtyp- und von 31% in den Dnmt TKO Zellen ermittelt werden. In **Abbildung 4.10 A** kann man gut erkennen, dass sowohl im Wildtyp als auch im *Knockout* über einen Zeitraum von 12 h stetig **13** eingebaut wird. Im Dnmt TKO ist die Menge an $^{13}\text{C}_5\text{-D}_1\text{-dC}$ 19% geringer als im Wildtyp. Anhand der in **Abbildung 4.10 B** dargestellten Werte (in Form einer Geraden) sieht man deutlich, dass die Menge an $^{13}\text{C}_5\text{-D}_1\text{-dC}$ bezogen auf dT+12 in gleichem Maße ansteigt. Für die Beobachtung einer möglichen Erhöhung an C-C-Bindungsbruchprodukt $^{13}\text{C}_5\text{-dC}$ zeigen die in **Abbildung 4.10 C** abgebildeten Werte einen vielversprechenden Trend. Während die Werte an $^{13}\text{C}_5\text{-dC}$ im Laufe der ersten 12 h nur leicht ansteigen, kann man im Wildtyp einen Anstieg an potentielltem C-C-Bindungsbruchprodukt $^{13}\text{C}_5\text{-dC}$ beobachten. Dieser Trend ist in **Abbildung 4.10 D** noch klarer zu sehen. Die Steigung der Geraden im Wildtyp (schwarz) unterscheidet sich deutlich von der im Dnmt TKO (rot), welche in etwa die gleiche Steigung wie die in **Abbildung 4.10 B** gezeigte (rot) aufweist. Das heißt, dass im Dnmt TKO eine Korrelation zwischen Einbau von $^{13}\text{C}_5\text{-D}_1\text{-dC}$ und dem chemischen Hintergrund an $^{13}\text{C}_5\text{-dC}$ beobachtet wird, wobei im Wildtyp eine deutliche Neubildung von $^{13}\text{C}_5\text{-dC}$ erkennbar ist. Diese korreliert nicht mit dem Einbau von **13** und zeigt somit vermutlich das Produkt eines erfolgten direkten C-C-Bindungsbruchs, welcher über eine Deformylierung von fdC oder Decarboxylierung von cadC ablaufen kann.

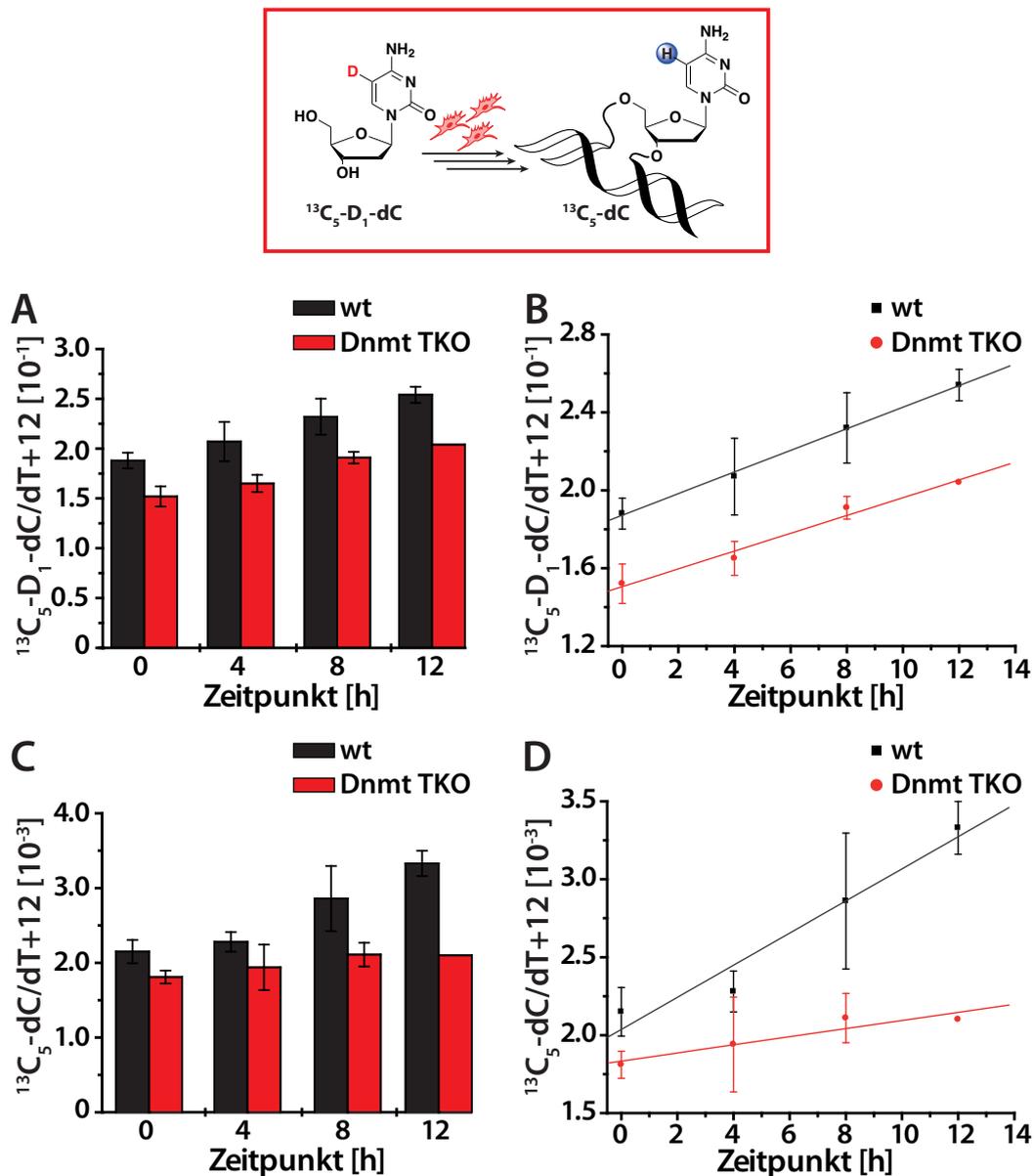


Abbildung 4.10: Absolute Level der verschiedenen dC-Spezies normiert auf $^{13}\text{C}_{10}^{15}\text{N}_2\text{-dT}$. **A** und **B** $^{13}\text{C}_5\text{-D}_1\text{-dC/dT+12}$. **C** und **D** Das Produkt einer möglichen Deformylierung/Decarboxylierung $^{13}\text{C}_5\text{-dC/dT+12}$. Zur verständlicheren Darstellung sind die in **A** und **C** dargestellten Werte in **B** und respektive **D** als Geraden dargestellt.

Um diese Daten zu bestätigen, wurde das Experiment weiterhin an Tet pro- und defizienten Zellen untersucht. In Tet *Knockout* Zellen können die ausgehend von mdC oxidierten Derivate hmdC, fdC und cadC nicht mehr erzeugt werden. Das bedeutet, dass eine Demethylierung über direkten C-C-Bindungsbruch hier nicht stattfinden kann. Daher würde im Wildtyp wie bei **Abbildung 4.10** eine Erhöhung des Produkts $^{13}\text{C}_5\text{-dC}$ im Vergleich mit dem *Knockout* erwartet werden. **Abbildung 4.11** zeigt erneut die nach Verfüterung mit **13** erhaltenen quantitativen Daten, die auf den internen Standard dT normiert wurden.

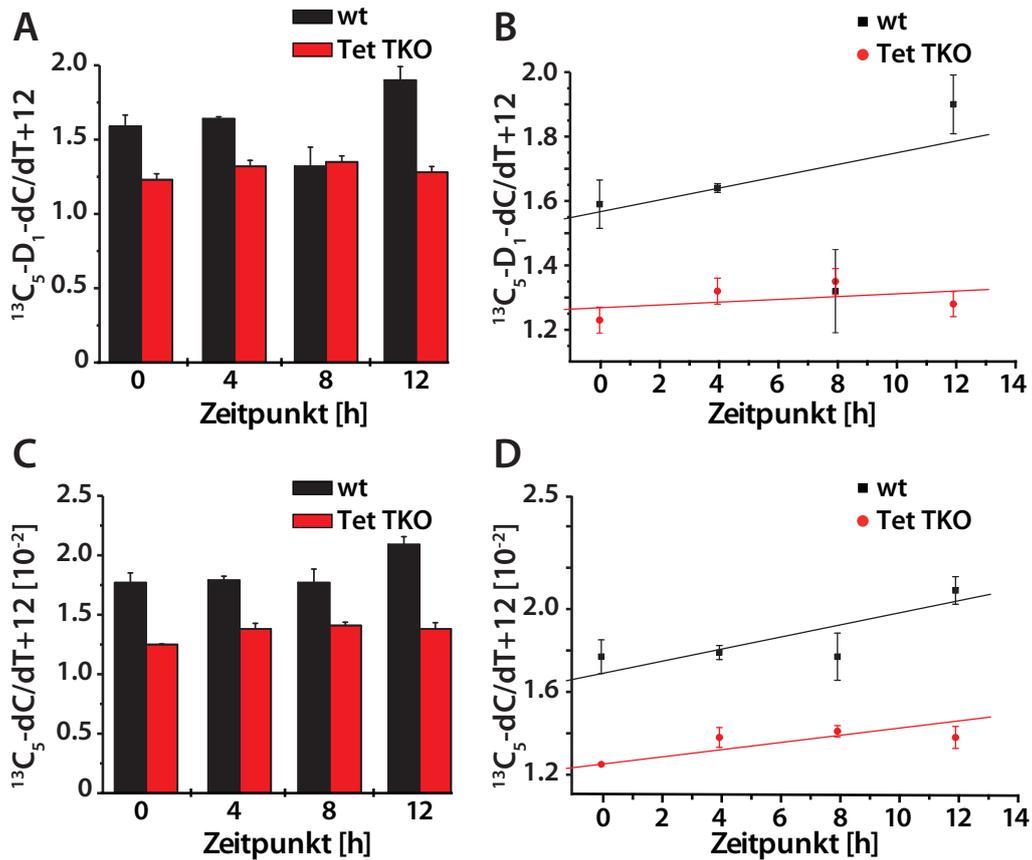


Abbildung 4.11: Absolute Level der verschiedenen dC-Spezies normiert auf $^{13}\text{C}_{10}^{15}\text{N}_2\text{-dT}$ in E14- (Wildtyp) und Tet TKO Zellen. **A** und **B** $^{13}\text{C}_5\text{-D}_1\text{-dC/dT+12}$. **C** und **D** Das Produkt einer möglichen Deformylierung/Decarboxylierung $^{13}\text{C}_5\text{-dC/dT+12}$. Zur verständlicheren Darstellung sind die in **A** und **C** dargestellten Werte in **B** und respektive **D** als Geraden dargestellt.

In diesem Fall wurde leider kein deutlicherer Anstieg des Produkts $^{13}\text{C}_5\text{-dC}$ ermittelt. Zwar ist die Steigung der Geraden (**Abb. 4.11 D**) im Wildtyp minimal steiler als im Tet TKO, jedoch ist der Effekt hier nicht so deutlich wie im Falle des Dnmt TKOs (**Abbildung 4.10**). Ein Grund hierfür könnte die wesentlich geringere Einbaurrate von **13** sein, die im Tet pro- und defizienten Zellen ermittelt wurde. Diese war mit ca. 20% um 1/3 geringer als in den Dnmt Wildtyp- und TKO-Zellen.

Um das Problem schwankender Einbauraten zu umgehen, wurde das RNA-Derivat **21** an Dnmt pro- und defiziente Zellen verfüttert. Wie oben beschrieben konnte eine Erhöhung der Einbaurrate von **21** auf 43% (Wildtyp) und 46% (TKO) ermittelt werden, was erneut eine Verbesserung darstellte. In **Abbildung 4.12 A** und **B** sind die quantitativen Daten von $^{13}\text{C}_5\text{-D}_1\text{-dC}$ und dem möglichen C-C-Bindungsbruchprodukt $^{13}\text{C}_5\text{-dC}$ abgebildet.

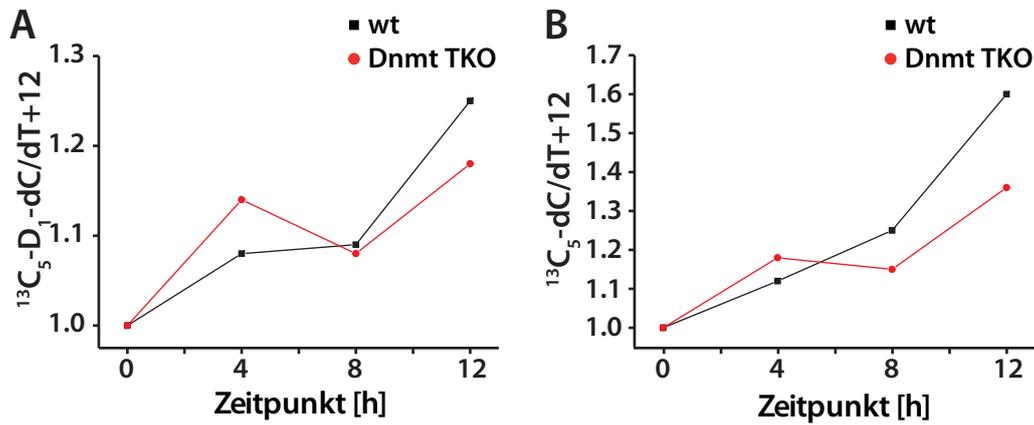


Abbildung 4.12: Normierung der absoluten Mengen an dC-Derivaten auf zugeführtes $^{13}\text{C}_{10}^{15}\text{N}_2\text{-dT}$. Abgebildet sind die absoluten Mengen als Geraden pro Zeitpunkt. Die einzelnen Werte wurden auf den $t = 0$ Ausgangswert normiert. **A** $^{13}\text{C}_5\text{-D}_1\text{-dC}$ in Wildtyp- und Dnmt-*Knockout* Zellen. **B** Mögliches *C-C*-Bindungsbruchprodukt $^{13}\text{C}_5\text{-dC}$.

Zunächst kann man erkennen, dass der Einbau von **21** nicht linear verläuft. Der Grund hierfür ist vermutlich, dass **21** als RNA-Spezies erst auf Diphosphatenebene zum DNA-Derivat reduziert wird (siehe **Abbildung 4.7**). So kann es sein, dass es zu bestimmten Zeitpunkten im Nukleosidpool zu einer unterschiedlichen Verfügbarkeit des DNA-Bausteins kommt, welche in diesem Fall vor allem von der Aktivität der Ribonukleotidreduktase abhängt. Betrachtet man nun die Werte von $^{13}\text{C}_5\text{-dC}$ (**Abb. 4.12 B**) im Dnmt TKO, so steigen diese stetig, aber nur leicht an. Im Wildtyp sieht es hingegen so aus, als würde $^{13}\text{C}_5\text{-dC}$ sehr stark ansteigen, vor allem bei Betrachtung der Werte zwischen 8 und 12 Stunden. Dies deutet erneut auf eine aktive Bildung des möglichen *C-C*-Bindungsbruchprodukts hin. Anhand der Daten kann man ebenfalls schließen, dass signifikante Effekte eventuell erst nach dem 12 Stunden Zeitpunkt zu beobachten sind, da in diesem Fall ein deutlicher Anstieg von $^{13}\text{C}_5\text{-dC}$ beobachtet wurde (**Abb. 4.12 B**). Für weitere Untersuchungen sollte daher in Zukunft das *Pulse-Chase*-Experiment über einen längeren Zeitraum durchgeführt werden.

Durch Verfütterung des RNA-Testmoleküls **21** konnte folglich eine deutliche Erhöhung der Einbaumenge von $^{13}\text{C}_5\text{-D}_1\text{-dC}$ erreicht werden, wodurch in Zukunft weitere wichtige Hinweise auf eine aktive Demethylierung mittels *C-C*-Bindungsbruchs erhalten werden könnten. Vor allem werden dadurch interessante Untersuchungen in Zelllinien ermöglicht, welche bis dato nur einen geringen Einbau von DNA-Testmolekülen aufwiesen (*vide infra*).

4.1.2 Zusammenfassung und Ausblick

Zusammengefasst wurden in Dnmt pro- und defizienten Zellen konkrete Hinweise auf die Beteiligung eines direkten *C-C*-Bindungsbruchs bei der aktiven Demethylierung beobachtet, welcher aus einer Deformylierungsreaktion von fdC oder aus einer Decarboxylierungsreaktion von cadC stammen könnte.

Bei Tet Wildtyp- und TKO-Zellen war diese Beobachtung jedoch nicht immer eindeutig. Mit diesen ersten Experimenten wurde jedoch eine solide Grundlage für eine genaue Untersuchung der einzelnen, zur aktiven mdC-Entfernung begleitenden Prozesse geschaffen.

Die Strategie zur Verfütterung eines RNA-C-Nukleosids **21** könnte aufgrund der erhöhten Einbauraten nun die Untersuchung eines aktiven C5-D/H-Austausches in Zellen ermöglichen, welche sich nicht mehr teilen. Die zu beobachtenden Effekte wären somit replikationsunabhängig und in diesem System würde man ausschließlich Vorgänge untersuchen, bei denen mdC aktiv demethyliert wird. Bei sogenannten iNGN-Zellen handelt es sich um humane Stammzellen, welche Neurogenin (NGN) 1 und 2 als Transgene enthalten. In diesem System können innerhalb von nur vier Tagen Stammzellen in bipolare Neuronen differenziert werden, wobei eine Zellteilung ab dem Start der Differenzierung nicht mehr stattfindet.^[208] Damit eignet sich dieses System hervorragend für *Pulse-Chase*-Experimente, außerdem wäre eine gleichzeitige Verfütterung von schwerem dT in diesem Fall obsolet. *Jessica Steinbacher* konnte in ihrer Dissertation zeigen, dass bereits nach 24 Stunden der Differenzierung der iNGN-Zellen die fdC- und cadC-Level deutlich abnahmen, was auf eine aktive Entfernung beider Spezies hindeutet.^[209] Damals wurde zwar auch versucht, das isotope markierte dC-Derivat der ersten Generation (**2**) an diese Zellen zu verfüttern, jedoch waren Einbauraten zwischen 2.5% - 7% zu gering, um signifikante Änderungen im Hinblick auf einen C5-D/H-Austausch im Rahmen aktiver Demethylierung zu zeigen. Außerdem sei nicht außer Acht gelassen, dass der chemische Hintergrund der dedeuterierten dC-Spezies zu Beginn schon mit einer Gesamtmenge von 3-4% beziffert werden konnte. Mit dem zuletzt vorgestellten Cytidin-Derivat **21** könnten genau diese Probleme umgangen werden, da zum einen eine sehr viel höhere Isotopenreinheit an C5- (99.7%) und somit ein geringeres Hintergrundsignal vorhanden ist. Zum anderen wurde gezeigt, dass der Einbau dieses Ribonukleosids, verglichen mit den 2'-Desoxy-dC-Derivaten, viel effizienter erfolgt. Auf diese Art und Weise könnte es in zukünftigen Experimenten möglich sein, eine Beteiligung des direkten C-C-Bindungsbruchs bei der aktiven Demethylierung zeigen zu können.

4.1.3 Projektbeiträge

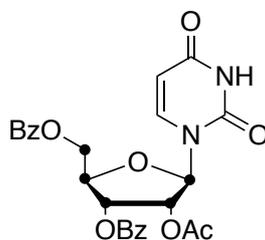
Im Rahmen dieser Arbeit war ich an der Entwicklung der vorgestellten dC-Derivate beteiligt. Die ursprüngliche Synthese des ersten dC-Derivats **2** wurde von *Arne Schröder* entwickelt, in dessen enger Zusammenarbeit die spätere C5-Funktionalisierungs-Deuterierungsmethodik zur erhöhten Deuteriumanreicherung angewandt wurde. Zusammen mit *Jessica Steinbacher* wurden dazu die ersten Daten in Wildtyp- und Tdg KO Zellen massenspektrometrisch ermittelt, wobei ich an der Probenvorbereitung, dem enzymatischen Verdau, der massenspektrometrischen Messung und der Auswertung beteiligt war. Die Synthese des zweiten dC-Derivats **4** wurde zusammen mit *Alexander Schön* im Rahmen seiner Masterarbeit entwickelt, welche von ihm unter meiner Aufsicht und Anleitung durchgeführt wurde und

welches für die Evaluierung der ersten biologischen Daten von mir synthetisiert und aufgereinigt wurde. Die Synthese der 3. und 4. Generation an dC (**13**) beziehungsweise C (**21**) Testmolekülen wurde von mir entwickelt und durchgeführt. Die Diskussion der erhaltenen Daten wurde in enger Zusammenarbeit mit *Thomas Carell*, *Katharina Iwan*, *Angie Kirchner*, *Fabio Spada* und *Markus Müller* durchgeführt.

4.1.4 Synthesevorschriften und Analytik

Synthese und Analytik der Verbindungen **1-3** wurden in Zusammenarbeit mit *Arne Schröder* durchgeführt und sind in seiner Dissertation zu finden.^[200] Verbindungen **4-12** wurden unter meiner Anleitung zusammen mit *Alexander Schön* synthetisiert und deren analytische Details sind in dieser aufgelistet.^[210]

Synthese von 2'-*O*-Acetyl-3',5'-di-*O*-benzoyl(1',2',3',4',5'-¹³C₅)-uridin (**15**)



¹³C₅-Ribosefragment **14**^[197] (500 mg, 1.12 mmol, 1.0 Äq) und Uracil (126 mg, 1.12 mmol, 1.0 Äq) wurden in MeCN (7.0 mL) suspendiert, *N,O*-Bis(trimethylsilyl)acetamid (0.82 mL, 3.36 mmol, 3.0 Äq) wurde zugegeben und das Gemisch für 30 Minuten bei 60 °C gerührt. Nachdem sich die trübe Suspension zu einer klaren Lösung entwickelte, wurde bei 60 °C TMSOTf (0.71 mL, 3.92 mmol, 3.5 Äq) hinzugegeben und das Gemisch für weitere 30 Minuten bei 60 °C verrührt. Das abgekühlte Gemisch wurde dann in gesättigte, wässrige NaHCO₃-Lösung gegeben (20 mL), mit DCM (3 x 50 mL) extrahiert und die vereinigten organische Phasen über Na₂SO₄ getrocknet. Das Lösemittel wurde *in vacuo* entfernt und das Rohprodukt säulenchromatographisch (50% EtOAc/ⁱHex) aufgereinigt, wobei **15** als farbloser Schaum isoliert wurde (410 mg, 0.82 mmol, 72%).

¹H-NMR (400 MHz, CDCl₃), δ/ppm = 8.09 - 8.03 (m, 4H, Bz-H), 7.62 (q, ³J = 7.6 Hz, 2H, Bz-H), 7.49 (td, ³J = 7.8 Hz, ⁴J = 2.9 Hz, 4H, Bz-H), 7.35 (dd, ³J = 8.2 Hz, ⁴J = 2.7 Hz, 1H, 5-H), 6.41 (d, ¹J_{C-H} = 171.1 Hz, 1H, 1'-H), 5.96 (d, ¹J_{C-H} = 150.3 Hz, 1H, 3'-H), 5.74 (d, ¹J_{C-H} = 153.0 Hz, 2'-H), 5.57 (d, ³J = 8.2 Hz, 6-H), 4.97 (d, ¹J_{C-H} = 154.6 Hz, 1H, 4'-H), 4.82 (d, ¹J_{C-H} = 150.1 Hz, 1H, 5'-H), 2.06 (s, 3H, Ac-CH₃).

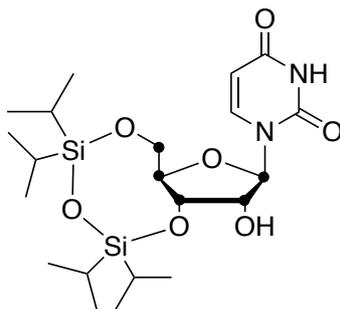
¹³C-NMR (101 MHz, CDCl₃), δ/ppm = 169.7 (Ac-C=O), 166.0 (Bz-C=O), 165.3 (Bz-C=O), 162.3 (C-4), 150.1 (C-2), 139.1 (C-5), 133.9 (C_{ar}), 133.8 (C_{ar}), 129.9 (C_{ar}), 129.7 (C_{ar}), 128.8 (C_{ar}), 128.7

(C_{ar}), 103.5 (C-6), 87.2 (dd, $J = 43.5$ Hz, 3.7 Hz, 1'-C), 80.4 (dd, $J = 43.0$ Hz, 38.3 Hz, 4'-C), 72.9 (dd, $J = 44.6$, 40.0, 2'-C), 70.9 (td, $J = 39.3$ Hz, 3.8 Hz, 3'-C), 63.7 (d, $J = 43.0$ Hz, 5'-C), 20.4 (Ac-CH₃).

HRMS (ESI): berechnet für $[\text{C}_{20}^{13}\text{C}_5\text{H}_{22}\text{N}_2\text{O}_9\text{Na}]^+$ ($\text{M}+\text{Na}$)⁺: 522.1385; gefunden: 522.1383.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1} = 1715.9$ (m), 1691.0 (m), 1600.8 (w), 1450.4 (w), 1373.5 (w), 1314.6 (w), 1262.5 (m), 1176.2 (w), 1107.9 (m), 1068.0 (m), 1025.0 (m), 806.0 (w), 768.0 (w), 708.0 (s), 686.8 (w).

Synthese von 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-(1',2',3',4',5'-¹³C₅)-uridin (**16**)



Zunächst wurde **15** (380 mg, 0.76 mmol, 1.0 Äq) durch Rühren in einer Lösung aus MeNH₂ in Ethanol (10 mL, 33%) über Nacht bei Raumtemperatur voll entschützt. Lösemittel wurden daraufhin *in vacuo* entfernt, der Rückstand in H₂O (10 mL) aufgenommen und mit DCM (5 x 10 mL) extrahiert. Nachdem die wässrige Phase eingeeengt wurde, wurde das freie Nucleosid in Pyridin (3.0 mL) gelöst und auf 0 °C gekühlt. 1,3-Dichlor-1,1,3,3-tetraisopropyldisiloxan (0.30 mL, 0.94 mmol, 1.2 Äq) wurde daraufhin tropfenweise hinzugegeben und das Gemisch auf Raumtemperatur aufgewärmt. Nach weiterem Rühren für 6 Stunden bei Raumtemperatur wurde die Reaktion durch Zugabe von H₂O (10 mL) gestoppt und die wässrige Phase mit EtOAc (2 x 15 mL) extrahiert. Die organische Phase wurde dann über Na₂SO₄ getrocknet und das Lösemittel *in vacuo* entfernt. Säulenchromatographische Aufreinigung (50% EtOAc/ⁱHex) des Rohprodukts lieferte 195 mg (0.40 mmol, 52%) von **16** in Form eines farblosen Schaums.

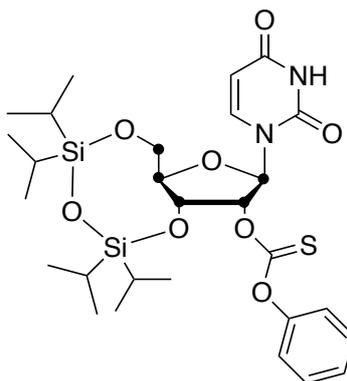
¹H-NMR (400 MHz, CDCl₃), $\delta/\text{ppm} = 8.54$ (bs, 1H, NH) 7.68 (dd, $^3J = 8.2$ Hz, $^4J = 2.4$ Hz, 1H, 5-H), 5.95 (d, $^1J_{\text{C-H}} = 174.0$ Hz, 1H, 1'-H), 5.69 (dd, $^3J = 8.2$ Hz, $^4J = 2.0$ Hz, 1H, 6-H), 4.60 - 3.76 (m, 5H, 2', 3', 4', 5'-H), 1.17 - 0.84 (m, 28H, SiCH(CH₃)₂, Si-CH).

¹³C-NMR (101 MHz, CDCl₃), $\delta/\text{ppm} = 162.9$ (C-4), 149.7 (2-C), 140.0 (C-5), 101.2 (C-6), 90.9 (dd, $J = 43.0$ Hz, 2.8 Hz, 1'-C), 81.9 (t, $J = 42.4$ Hz, 2'-C), 75.2 (dd, $J = 43.0$ Hz, 38.1 Hz, 3'-C), 69.0 (ddd, $J = 41.4$ Hz, 38.1 Hz, 3.2 Hz, 4'-C), 60.3 (d, $J = 43.2$ Hz, 5'-C), 17.4 - 16.8 (Si-C(CH₃)₂), 13.4 - 12.8 (Si-C(CH₃)₂).

HRMS (ESI): berechnet für $[\text{C}_{16}^{13}\text{C}_5\text{H}_{39}\text{N}_2\text{O}_7\text{Si}_2]^+$ ($\text{M}+\text{H}$)⁺: 492.2458; gefunden: 492.2458.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1} = 2942.4$ (w), 2924.8 (w), 2865.4 (w), 1701.5 (m), 1686.3 (s), 1680.9 (s), 1462.3 (m), 1453.4 (m), 1437.9 (m), 1384.6 (w), 1262.7 (m), 1199.0 (w), 1135.6 (m), 1097.3 (s), 1028.7 (s), 976.5 (m), 884.3 (s), 853.9 (m), 806.5 (m), 763.9 (m), 700.6 (s).

Synthese von 3',5'-(1,1,3,3,-Tetraisopropyldisiloxan-1,3-diyl)-2'-O-phenoxythiocarbonyl-(1',2',3',4',5'- $^{13}\text{C}_5$)-uridin (17)



16 (190 mg, 0.39 mmol, 1.0 Äq) wurde zusammen mit DMAP (152 mg, 1.25 mmol, 2.9 Äq) in MeCN (6.0 mL) gelöst, anschließend mit *O*-Phenylchlorothionoformiat (90 μL , 0.65 mmol, 1.5 Äq) versetzt und für 2 Stunden bei Raumtemperatur verrührt. Das Reaktionsgemisch wurde dann in gesättigte, wässrige NH_4Cl -Lösung (50 mL) gegeben und mit EtOAc (2 x 50 mL) extrahiert. Die vereinigten organischen Phasen wurden schließlich über Na_2SO_4 getrocknet, das Lösemittel *in vacuo* entfernt und das Produkt säulenchromatographisch (25% EtOAc/ i Hex) isoliert. So konnte **17** (180 mg, 0.29 mmol, 74%) als farbloses Öl isoliert werden.

$^1\text{H-NMR}$ (400 MHz, CDCl_3), $\delta/\text{ppm} = 8.32$ (bs, 1H, NH), 7.73 (dd, $^3J = 8.2$ Hz, $^4J = 2.5$ Hz, 1H, 5- H), 7.42 (t, $^3J = 7.8$ Hz, 2H, CH_{ar}), 7.30 (t, $^3J = 7.4$ Hz, 1H, CH_{ar}), 7.12 (d, $^3J = 7.9$ Hz, 2H, CH_{ar}), 6.21 (td, $^1J_{\text{C-H}} = 163.7$ Hz, $^3J = 5.3$ Hz, 1H, 2'- H), 6.15 (d, $^1J_{\text{C-H}} = 173.0$ Hz, 1H, 1'- H), 5.74 (d, $^3J = 8.1$ Hz, 1H, 6- H), 4.75 - 3.84 (m, 4H, 3', 4', 5'- H), 1.11 - 1.02 (m, 28H, $\text{SiCH}(\text{CH}_3)_2$, Si-CH).

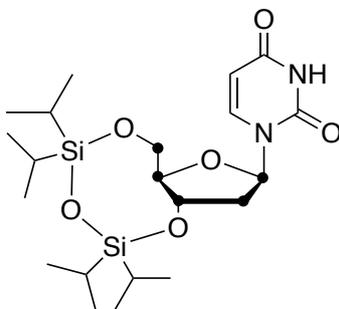
$^{13}\text{C-NMR}$ (101 MHz, CDCl_3), $\delta/\text{ppm} = 193.7$ ((O) $\text{C}=\text{S}$), 162.6 (C-4), 153.4 (C_{ar}), 149.4 (2- C), 139.5 (C-5), 129.6 (C_{ar}), 126.7 (C_{ar}), 121.8 (C_{ar}), 102.3 (C-6), 88.6 (d, $J = 43.6$ Hz, 1'- C), 83.7 (t, $J = 41.7$ Hz, 2'- C), 82.2 (t, $J = 42.2$ Hz, 3'- C), 68.1 (td, $J = 41.0$ Hz, 4.0 Hz, 4'- C), 59.4 (d, $J = 43.1$ Hz, 5'- C), 17.4 - 16.8 ($\text{Si-C}(\text{CH}_3)_2$), 13.4 - 12.8 ($\text{Si-C}(\text{CH}_3)_2$).

HRMS (ESI): berechnet für $[\text{C}_{23}^{13}\text{C}_5\text{H}_{42}\text{N}_2\text{O}_8\text{SSi}_2\text{Na}]^+$ ($\text{M}+\text{Na}$) $^+$: 650.2260; gefunden: 650.2260.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1} = 3192.9$ (bw), 2943.2 (m), 2866.6 (m), 1691.9 (s), 1489.9 (m), 1462.3 (m), 1385.9 (m), 1299.7 (m), 1273.3 (s), 1205.4 (s), 1139.8 (m), 1101.7 (s), 1042.9 (s), 926.6 (w), 884.7 (m), 855.5

(m), 810.0 (w), 766.1 (m), 689.9 (m).

Synthese von 3',5'-(1,1,3,3,-Tetraisopropylidisiloxan-1,3-diy1)-2'-desoxy-(1',2',3',4',5'-¹³C₅)-uridin (18)



Zuerst wurde **17** (170 mg, 0.27 mmol, 1.0 Äq) zusammen mit AIBN (9.0 mg, 0.05 mmol, 0.2 Äq) und *n*Bu₃SnH (110 μ L, 0.41 mmol, 1.5 Äq) in Toluol (5.0 mL) gelöst. Die Reaktionsmischung wurde daraufhin *freeze-pump-thaw*-entgast (3x) und für 90 Minuten auf 75 °C erhitzt. Lösemittel wurden *in vacuo* entfernt und das Rohprodukt über Säulenchromatographie (30% EtOAc/²Hex) aufgereinigt, wobei das Desoxygenierungsprodukt **18** (124 mg, 0.26 mmol, 96%) als farbloser Schaum isoliert werden konnte.

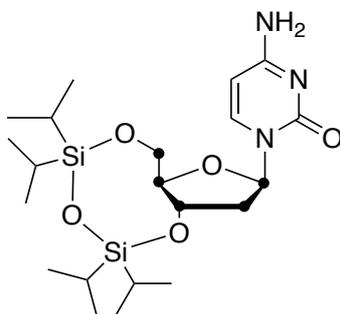
¹H-NMR (400 MHz, CDCl₃), δ /ppm = 8.42 (bs, 1H, NH), 7.78 (dd, ³*J* = 8.2 Hz, ⁴*J* = 2.4 Hz, 1H, 5-H), 6.25 (dd, ¹*J*_{C-H} = 167.2 Hz, ³*J* = 6.6 Hz, 1H, 1'-H), 5.69 (dd, ³*J* = 8.1 Hz, ⁴*J* = 2.1 Hz, 1H, 6-H), 4.66 - 3.48 (m, 4H, 3', 4', 5'-H), 2.74 - 2.00 (m, 2H, 2'-H), 1.19 - 0.86 (m, 28H, SiCH(CH₃)₂, Si-CH).

¹³C-NMR (101 MHz, CDCl₃), δ /ppm = 163.0 (C-4), 149.8 (2-C), 139.7 (C-5), 101.7 (C-6), 85.1 (ddd, *J* = 43.1 Hz, 40.6 Hz, 2.3 Hz 4'-C), 84.4 (d, *J* = 36.3 Hz, 1'-C), 66.9 (t, *J* = 40.2 Hz, 3'-C), 60.0 (d, *J* = 43.0 Hz, 4.0 Hz, 5'-C), 39.8 (t, *J* = 38.1 Hz, 2'-C), 17.8 - 16.5 (Si-C(CH₃)₂), 14.1 - 12.3 (Si-C(CH₃)₂).

HRMS (ESI): berechnet für [C₁₆¹³C₅H₃₉N₂O₆Si₂]⁺ (M+H)⁺: 476.2509; gefunden: 476.2510.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 2942.9 (m), 2866.4 (m), 1689.2 (s), 1462.2 (m), 1386.0 (w), 1266.4 (m), 1246.6 (w), 1121.8 (m), 1095.9 (m), 1065.1 (m), 1031.8 (s), 884.4 (m), 765.2 (w), 699.8 (m).

Synthese von 3',5'-(1,1,3,3,-Tetraisopropylidisiloxan-1,3-diyloxy)-2'-desoxy-(1',2',3',4',5'-¹³C₅)-cytidin (**19**)



18 (120 mg, 0.25 mmol, 1.0 Äq) und 1,2,4-Triazol (140 mg, 2.02 mmol, 8.0 Äq) wurden in Pyridin (3.0 mL) gelöst. POCl₃ (48 µL, 0.50 mmol, 2.0 Äq) wurde tropfenweise hinzugefügt und für 45 Minuten bei Raumtemperatur verrührt. Das Gemisch wurde in gesättigte, wässrige NH₄Cl-Lösung (25 mL) gegeben und mit EtOAc (3 x 25 mL) extrahiert. Die vereinigten organischen Lösungen wurden über Na₂SO₄ getrocknet und das Lösemittel *in vacuo* entfernt. Das orange Öl wurde daraufhin in Dioxan (3.0 mL) aufgenommen, mit konzentrierter NH₄OH (1.0 mL) versetzt und über Nacht bei Raumtemperatur gerührt. Die Lösemittel wurden *in vacuo* entfernt und das Produkt **19** konnte über säulenchromatographische Aufreinigung (4% MeOH/DCM → 8% MeOH/DCM) isoliert werden (114 mg, 0.24 mmol, 96%).

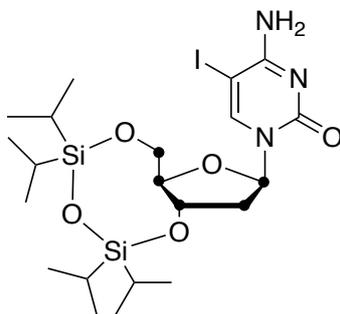
¹H-NMR (400 MHz, CDCl₃), δ/ppm = 7.95 (dd, ³J = 7.5 Hz, ⁴J = 2.4 Hz, 1H, 5-H), 6.29 (dd, ¹J_{C-H} = 177.4 Hz, ³J = 7.1 Hz, 1H, 1'-H), 5.71 (d, ³J = 7.4 Hz, 1H, 6-H), 4.61 - 3.52 (m, 4H, 3', 4', 5'-H), 2.77 - 2.08 (m, 2H, 2'-H), 1.09 - 0.97 (m, 28H, SiCH(CH₃)₂, Si-CH).

¹³C-NMR (101 MHz, CDCl₃), δ/ppm = 165.6 (C-4), 156.1 (2-C), 141.4 (C-5), 93.7 (C-6), 85.5 - 84.3 (4'-C, 1'-C), 66.5 (t, J = 40.7 Hz, 3'-C), 59.8 (d, J = 43.0 Hz, 5'-C), 39.9 (t, J = 36.8 Hz, 2'-C), 17.5 - 16.8 (Si-C(CH₃)₂), 13.4 - 12.4 (Si-C(CH₃)₂).

HRMS (ESI): berechnet für [C₁₆¹³C₅H₄₀N₃O₅Si₂]⁺ (M+H)⁺: 475.2669; gefunden: 475.2668.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 3344.0 (bw), 2942.0 (w), 2865.1 (w), 1621.8 (m), 1518.8 (w), 1480.0 (m), 1463.1 (m), 1400.3 (w), 1359.1 (w), 1281.9 (w), 1243.6 (w), 1190.8 (w), 1119.9 (m), 1095.1 (s), 1063.6 (w), 1029.9 (s), 966.7 (m), 938.7 (m), 918.8 (w), 882.9 (s), 853.8 (m), 785.1 (m), 767.3 (m), 691.0 (s).

Synthese von 3',5'-**(1,1,3,3,-Tetraisopropylidisiloxan-1,3-diyI)**-5-iodo-2'-desoxy-(1',2',3',4',5'-¹³C₅)-cytidin (**20**)



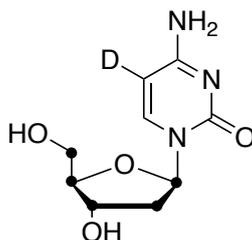
Zur Iodierung wurde **19** (105 mg, 0.22 mmol, 1.0 Äq) in MeCN (20 mL) gelöst und mit Cerammoniumnitrat (241 mg, 0.44 mmol, 2.0 Äq) versetzt. Iod (112 mg, 0.44 mmol, 2.0 Äq) wurde zugegeben und das dunkelbraune Gemisch für 90 Minuten bei 60 °C gerührt. Die Reaktion wurde durch Zugabe von gesättigter, wässriger NaHCO₃- (20 mL) und Na₂S₂O₃-Lösung (5.0 mL) gestoppt, die wässrige Phase mit DCM (3 x 20 mL) extrahiert und die vereinigten organischen Phasen über Na₂SO₄ getrocknet. Nach Entfernen des Lösemittels *in vacuo* konnte durch säulenchromatographische Aufreinigung (4% MeOH/DCM) **20** (86 mg, 0.14 mmol, 65%) in Form eines farblosen Glases isoliert werden.

¹H-NMR (400 MHz, CDCl₃), δ/ppm = 8.07 (s, 1H, 6-H), 5.97 (dd, ¹J_{C-H} = 175.8 Hz, ³J = 6.8 Hz, 1H, 1'-H), 4.64 - 3.48 (m, 4H, 3', 4', 5'-H), 2.82 - 2.06 (m, 2H, 2'-H), 1.26 - 0.86 (m, 28H, SiCH(CH₃)₂, Si-CH).

¹³C-NMR (101 MHz, CDCl₃), δ/ppm = 163.8 (C-4), 154.8 (2-C), 146.4 (C-5), 85.7 - 84.6 (4'-C, 1'-C), 66.5 (t, J = 41.2 Hz, 3'-C), 59.6 (d, J = 43.2 Hz, 5'-C), 55.5 (6-C), 39.9 (t, J = 37.1 Hz, 2'-C), 19.1 - 16.1 (Si-C(CH₃)₂), 14.0 - 11.3 (Si-C(CH₃)₂).

HRMS (ESI): berechnet für [C₁₆¹³C₅H₃₉N₃O₅ISi₂]⁺ (M+H)⁺: 601.1635; gefunden: 601.1632.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1}$ = 3432.8 (bw), 3326.7 (bw), 2942.9 (w), 2865.7 (m), 1641.7 (m), 1605.0 (m), 1569.5 (m), 1475.0 (m), 1387.4 (w), 1318.7 (w), 1299.0 (w), 1272.7 (m), 1238.0 (m), 1156.7 (m), 1115.1 (m), 1094.3 (s), 1044.6 (m), 1022.3 (s), 988.5 (m), 969.0 (m), 938.8 (m), 919.1 (w), 884.2 (m), 871.6 (m), 851.8 (m), 773.6 (m), 694.6 (s).

5-Deutero-2'-desoxy-(1',2',3',4',5'-¹³C₅)-cytidin (**13**)

Iodo-Derivat **20** (40 mg, 0.07 mmol, 1.0 Äq) wurde in EtOAc (1.5 mL) gelöst und mit HF·Pyridin (27 μ L, 70%ige Lösung) versetzt. Nachdem das Gemisch über Nacht gerührt wurde, wurde TMSOMe (200 μ L) hinzugefügt, wobei ein farbloser Feststoff ausfiel. Die Suspension wurde bei 6000 rpm abzentrifugiert, der Überstand verworfen und der dabei erhaltene, farblose Feststoff erneut mit EtOAc gewaschen (10 mL). Dieser Vorgang wurde zwei Mal wiederholt, der Rückstand schließlich *in vacuo* getrocknet und das dadurch entschlützte Iodo-Nukleosid ohne weitere Aufreinigung verwendet.

¹H-NMR (400 MHz, D₂O), δ /ppm = 8.15 (sd, ⁴J = 2.6 Hz, 1H, 6-H), 5.91 (dd, ¹J_{C-H} = 173.7 Hz, ³J = 6.2 Hz, 1H, 1'-H), 4.41 - 3.28 (m, 4H, 3', 4', 5'-H), 2.44 - 1.85 (m, 2H, 2'-H).

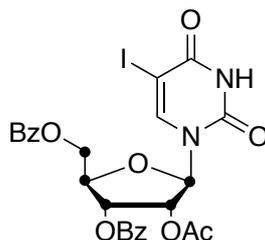
HRMS (ESI): berechnet für [C₄¹³C₅H₁₃N₃O₄I]⁺ (M+H)⁺: 359.0113; gefunden: 359.0113.

7.0 mg (0.02 mmol, 1.0 Äq) dieser Verbindung wurden dann mit NaHCO₃ (12 mg, 0.16 mmol, 8.0 Äq) und Pd/C (1.0 mg, 10%) in MeOD (1.0 mL) gelöst und unter D₂-Atmosphäre gesetzt. Das Gemisch wurde über Nacht bei Raumtemperatur verrührt, die Lösung über einen Spritzenfilter abfiltriert und das Lösemittel *in vacuo* entfernt. So konnte 2.0 mg (0.01 mmol, 50%) von **13** erhalten werden.

¹H-NMR (400 MHz, D₂O), δ /ppm = 7.70 (sd, ⁴J = 2.6 Hz, 1H, 6-H), 6.36 (dt, ¹J_{C-H} = 171.0 Hz, ³J = 7.3 Hz, 1H, 1'-H), 4.12 - 3.52 (m, 4H, 3', 4', 5'-H), 2.55 - 1.95 (m, 2H, 2'-H).

¹³C-NMR (101 MHz, D₂O), δ /ppm = 166.1 (C-4), 160.9 (C-2), 157.6 (C-6), 141.4 (C-5), 86.5 (dd, J = 41.9 Hz, 36.9 Hz, 4'-C), 86.0 (d, J = 36.0 Hz, 1'-C), 70.4 (t, J = 41.9 Hz, 3'-C), 61.2 (d, 41.7 Hz, 5'-C), 39.1 (t, J = 36.2 Hz, 2'-C).

HRMS (ESI): berechnet für [C₄¹³C₅H₁₃DN₃O₄]⁺ (M+H)⁺: 234.1209; gefunden: 234.1210.

Synthese von 2'-O-Acetyl-3',5'-di-O-benzoyl-5-iodo-(1',2',3',4',5'-¹³C₅)-uridin (**22**)

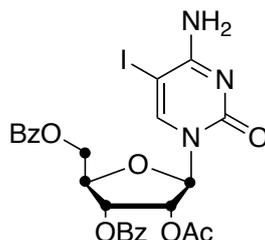
15 (110 mg, 0.22 mmol, 1.0 Äq) wurde in MeCN (4.0 mL) gelöst und mit CAN (61 mg, 0.11 mmol, 0.5 Äq) und Iod (34 mg, 0.13 mmol, 0.6 Äq) versetzt. Die dunkelbraune Lösung wurde daraufhin für 90 Minuten bei 60 °C gerührt und die Reaktion durch Zugabe einer gesättigten, wässrigen Na₂S₂O₃-Lösung (1.0 mL) gestoppt. Die wässrige Phase wurde dann mit DCM (2 x 20 mL) und EtOAc (1 x 20 mL) extrahiert. Vereinigte organische Phasen wurden über Na₂SO₄ getrocknet, das Lösemittel *in vacuo* entfernt und das Rohprodukt wurde über Säulenchromatographie (40% EtOAc/ⁱHex) aufgereinigt. Dadurch konnte **22** (105 mg, 017 mmol, 78%) in Form eines farblosen Schaums isoliert werden.

¹H-NMR (400 MHz, CDCl₃), δ/ppm = 8.60 (s, 1H, NH), 8.09 (ddd, *J* = 13.8 Hz, 8.1 Hz, 1.4 Hz, 4H, Bz-H), 7.79 (sd, ⁴*J* = 2.4 Hz, 1H, 6-H), 7.69 - 7.56 (m, 2H, Bz-H), 7.50 (td, ³*J* = 7.7 Hz, ⁴*J* = 1.5 Hz, 4H, Bz-H), 6.24 (d, ¹*J*_{C-H} = 169.7 Hz, 1H, 1'-H), 5.96 (d, ¹*J*_{C-H} = 155.8 Hz, 1H, 3'-H), 5.74 (d, ¹*J*_{C-H} = 152.4 Hz, 2'-H), 4.96 (d, ¹*J*_{C-H} = 154.1 Hz, 1H, 4'-H), 4.85 (d, ¹*J*_{C-H} = 148.4 Hz, 2H, 5'-H), 2.06 (s, 3H, Ac-CH₃).

¹³C-NMR (101 MHz, CDCl₃), δ/ppm = 169.8 (Ac-C=O), 166.0 (Bz-C=O), 165.3 (Bz-C=O), 162.3 (C-4), 150.2 (C-2), 143.4 (C-6), 139.1 (C-5), 133.9 (C_{ar}), 133.8 (C_{ar}), 129.8 (C_{ar}), 129.7 (C_{ar}), 129.0 (C_{ar}), 128.7 (C_{ar}), 128.5 (C-5), 86.8 (dd, *J* = 44.6 Hz, 3.8 Hz, 1'-C), 80.9 (dd, *J* = 42.7 Hz, 37.8 Hz, 4'-C), 73.0 (dd, *J* = 43.9, 39.1, 3'-C), 71.2 (td, *J* = 40.5 Hz, 3.7 Hz, 2'-C), 63.9 (d, *J* = 42.8 Hz, 5'-C), 20.4 (Ac-CH₃).

HRMS (ESI): berechnet für [C₂₀¹³C₅H₂₅N₃O₉I]⁺ (M+NH₄)⁺: 643.0798; gefunden: 643.0795.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 3066.9 (bw), 1713.9 (s), 1687.0 (s), 1600.9 (m), 1450.0 (m), 1438.9 (m), 1369.0 (m), 1314.4 (w), 1253.0 (s), 1221.8 (m), 1176.0 (m), 1095.7 (m), 1067.4 (m), 1024.4 (m), 907.2 (m), 707.4 (s).

Synthese von 2'-O-Acetyl-3',5'-di-O-benzoyl-5-iodo-(1',2',3',4',5'-¹³C₅)-cytidin (**23**)

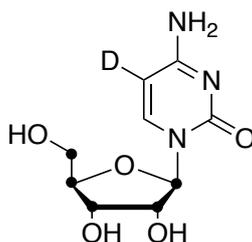
Zur Aminierung von **22** (84 mg, 0.13 mmol, 1.0 Äq) wurde die Verbindung in Pyridin (3.0 mL) gelöst und mit 1,2,4-Triazol (74 mg, 1.08 mmol, 8.0 Äq) versetzt. Dem Gemisch wurde dann bei Raumtemperatur POCl₃ (26 µL, 0.27 mmol, 2.0 Äq) zugefügt und für 90 Minuten gerührt. Die Reaktion wurde durch Zugabe von gesättigter, wässriger NH₄Cl-Lösung (25 mL) gestoppt und die wässrige Phase mit EtOAc (3 x 25 mL) extrahiert. Nach Trocknen der vereinigten, organischen Phasen mit Na₂SO₄ wurde das Lösemittel *in vacuo* entfernt und das Rohprodukt säulenchromatographisch (2% MeOH/DCM → 4% MeOH/DCM) aufgereinigt. So konnten 82 mg (0.13 mmol, quant.) von **23** in Form eines gelblichen Glases isoliert werden.

¹H-NMR (400 MHz, CDCl₃), δ/ppm = 8.43 (s, 1H, NH), 8.06 (dd, ³J = 12.6 Hz, ⁴J = 7.7 Hz, 4H, Bz-H), 7.75 (sd, ⁴J = 2.4 Hz, 1H, 6-H), 7.60 (q, ³J = 7.6 Hz, 2H, Bz-H), 7.47 (td, ³J = 7.5 Hz, ⁴J = 4.8 Hz, 4H, Bz-H), 6.52 (d, ¹J_{C-H} = 170.1 Hz, 1H, 1'-H), 5.95 - 5.38 (m, 3H, 2'-H, 3'-H), 5.74 (d, ¹J_{C-H} = 152.4 Hz, 2'-H), 4.96 (d, ¹J_{C-H} = 148.6 Hz, 1H, 4'-H), 4.82 (d, ¹J_{C-H} = 150.1 Hz, 2H, 5'-H), 2.06 (s, 3H, Ac-CH₃).

¹³C-NMR (101 MHz, CDCl₃), δ/ppm = 169.7 (Ac-C=O), 166.0 (Bz-C=O), 165.3 (Bz-C=O), 163.6 (C-4), 154.3 (C-2), 146.2 (C-6), 133.8 (C-5), 133.6 (C_{ar}), 129.9 (C_{ar}), 129.7 (C_{ar}), (C_{ar}), 128.9 (C_{ar}), 128.6 (C_{ar}), 87.6 (dd, J = 44.6 Hz, 3.4 Hz, 1'-C), 80.3 (dd, J = 42.7 Hz, 38.4 Hz, 4'-C), 73.8 (dd, J = 44.5, 39.8, 3'-C), 71.0 (td, J = 39.1 Hz, 3.5 Hz, 2'-C), 63.8 (d, J = 43.0 Hz, 5'-C), 20.5 (Ac-CH₃).

HRMS (ESI): berechnet für [C₂₀¹³C₅H₂₂N₃O₈INa]⁺ (M+Na)⁺: 647.0512; gefunden: 647.0514.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 3336.2 (bw), 1725.0 (s), 1661.1 (s), 1490.5 (m), 1451.8 (w), 1370.2 (w), 1267.6 (s), 1177.6 (w), 1110.5 (m), 1069.4 (m), 1026.1 (m), 711.0 (m).

Synthese von 5-Deutero-(1',2',3',4',5'-¹³C₅)-cytidin (**21**)

Zunächst wurde **23** (50 mg, 0.08 mmol, 1.0 Äq) durch Rühren in einer ethanolischen Methylamin-Lösung (30%, 10 mL) über Nacht entschützt. Das Lösemittel wurde *in vacuo* entfernt und der Rückstand in H₂O (10 mL) aufgenommen. Die wässrige Phase wurde mit DCM (5 x 20 mL) extrahiert und die wässrige Phase anschließend *in vacuo* eingengt. Das entschützte Rohprodukt wurde ohne weitere Aufreinigung weiter verwendet.

¹H-NMR (400 MHz, D₂O), δ /ppm = 8.15 (sd, ⁴J = 2.6 Hz, 1H, 6-H), 5.77 (d, ¹J_{C-H} = 173.1 Hz, 1H, 1'-H), 4.28 - 3.18 (m, 5H, 2'-H, 3'-H, 4'-H, 5'-H).

HRMS (ESI): berechnet für [C₄¹³C₅H₁₃N₃O₈I]⁺ (M+H)⁺: 375.0062; gefunden: 375.0066.

Zur Defunktionalisierung wurden 10 mg (0.03 mmol, 1.0 Äq) zusammen mit NaHCO₃ (18 mg, 0.21 mmol, 8.0 Äq) und Pd/C (2.0 mg, 10% wt) in MeOD (4.0 mL) gelöst. Die Lösung wurde daraufhin unter eine Deuterium-Atmosphäre gesetzt und über Nacht bei Raumtemperatur gerührt, anschließend über einen Spritzenfilter filtriert und das Lösemittel *in vacuo* entfernt. Dadurch konnten 3.0 mg (0.012 mmol, 40%) von **21** isoliert werden.

¹H-NMR (400 MHz, D₂O), δ /ppm = 7.72 (sd, ⁴J = 2.6 Hz, 1H, 6-H), 5.99 (d, ¹J_{C-H} = 170.2 Hz, 1H, 1'-H), 4.37 - 3.82 (m, 3H, 2'-H, 3'-H, 4'-H), 3.65 - 3.49 (m, 2H, 5'-H).

¹³C-NMR (101 MHz, D₂O), δ /ppm = 166.1 (C-4), 161.2 (C-2), 157.6 (C-6), 141.5 (C-5), 90.1 (dd, J = 42.9 Hz, 3.8 Hz, 1'-C), 83.7 (dd, J = 41.9 Hz, 39.1 Hz, 4'-C), 73.8 (dd, J = 42.8, 37.9, 3'-C), 69.1 (td, J = 38.4 Hz, 3.7 Hz, 2'-C), 60.6 (d, J = 42.0 Hz, 5'-C).

HRMS (ESI): berechnet für [C₄¹³C₅H₁₃DN₃O₅]⁺ (M+H)⁺: 250.1158; gefunden: 250.1158.

4.2 Synthese eines photoinduzierbaren Hydroxylamin-Testmoleküls zur *in cellulo*-Markierung von abasischen Stellen

Wie unter **Punkt 1.4.1** beschrieben, sind abasische Stellen eine der am häufigsten vorkommenden Intermediate im Metabolismus genomischer DNA. In der unter **Punkt 3.2** gezeigten Veröffentlichung wurde ein neues Reagenz **24** (**Abb. 4.13**) zur Derivatisierung dieser Intermediate in genomischer DNA vorgestellt. Da abasische Stellen kurzlebig sind und in der Zelle schnell prozessiert werden, sollte ein neues Derivat **25** dieses Reagenzes dargestellt werden (**Abb. 4.13**).

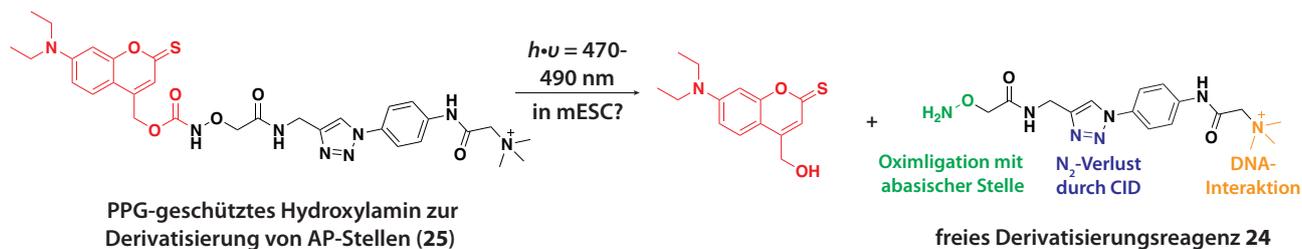
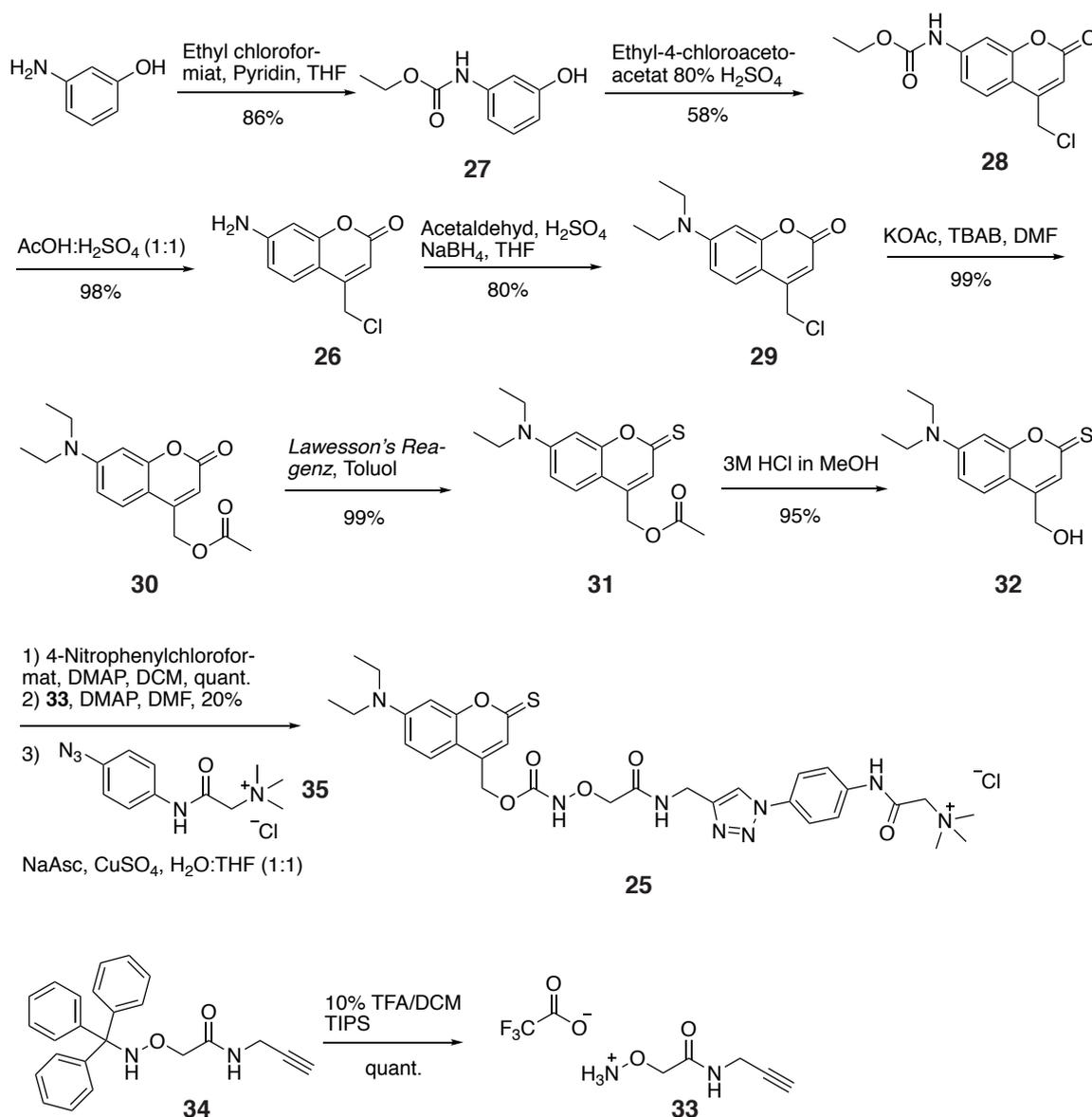


Abbildung 4.13: Struktur des Coumarin-geschützten Hydroxylamins **25**. Nach Bestrahlung mit Licht wird die photolabile Schutzgruppe entfernt und das reaktive Derivatisierungsreagenz **24** wird freigesetzt.

Das reaktive Hydroxylamin sollte über eine photolabile Schutzgruppe (PPG) geschützt werden, sodass zum einen die Zellmembran passiert werden kann und zum anderen eine zeitliche und räumliche Kontrolle der Freisetzung des Derivatisierungsreagenzes möglich ist. Außerdem sollte die PPG mit Licht abseits des UV-Spektrums abspaltbar sein. Durch UV-Licht können Riboflavin und Flavinmononukleotid angeregt werden, welche wiederum Singulett-Sauerstoff erzeugen und die DNA somit endogen oxidativ schädigen würden.^[211] Oxidative Schäden werden über die BER prozessiert, wobei abasische Stellen intermediär auftreten.^[141] Somit würde bei einer Entschützung mit UV-Licht der zu untersuchende Analyt (abasische Stelle) in größerem Maße artifiziell erzeugt werden und könnte dabei zu beobachtende Effekte überdecken (zum Beispiel abasische Stellen, welche im Rahmen der aktiven Demethylierung erzeugt werden).

Als potentielle PPG, welche nicht mit UV-Licht entschützt wird und deren Anwendung bereits *in cellulo* gezeigt werden konnte, wurde ein von *Fournier et al.* entwickeltes thioniertes Coumaringerüst identifiziert.^[212,213] Die Synthese des in **Abbildung 4.13** gezeigten Zielmoleküls **25** ist in **Schema 4.6** gezeigt. Zunächst wurde das 7-Amino-4-chloromethylcoumarin-Grundgerüst **26** ähnlich zu der in der Literatur beschriebenen Synthese dargestellt.^[214] 3-Aminophenol wurde anfangs mit Ethylchloroformat selektiv und in sehr guter Ausbeute an der Aminofunktion zu **27** geschützt und daraufhin mit Ethyl-4-chloroacetoacetat zum 4-Chloromethylcoumarin-Fragment **28** in 58%-iger Ausbeute zyklisiert. Darauf folgende Acetolyse von **28** lieferte das freie aromatische Amin **26** in nahezu quantitativer Ausbeute, welches dann mit Acetaldehyd unter reduktiven Bedingungen zum Bis-alkylierten



Schema 4.6: Synthese des Coumarin-geschützten Hydroxylamins **25**.

4-Chloromethylcoumarin **29** umgesetzt wurde. Das Chlorid in **29** wurde mit Kaliumacetat unter Zuhilfenahme von TBAB als Phasentransferkatalysator zu **30** substituiert, woraufhin das Thiolacton mit Hilfe des *Lawesson's Reagenz* eingeführt wurde. Thiocoumarin **31** wurde dann unter Reflux im Sauren zum freien Alkohol **32** entschützt. Dabei musste auf eine sehr hohe Verdünnung der methanolischen Salzsäure (60 mM) geachtet werden, da es ansonsten zu einer kompletten Hydrolyse des Thiolactons zum Lacton kam. Freies Thiolacton **32** wurde dann mit Hilfe von 4-Nitrophenylchloroformiat carbonyliert und das Intermediat mit Alkin **33** umgesetzt. **33** konnte ausgehend von der tritylgeschützten Vorläuferverbindung **34**^[215] selektiv mit 10%-iger Trifluoressigsäure in DCM und Zuhilfenahme des Kationen-*Scavengers* Triisopropylsilan quantitativ entschützt werden. In einem letzten Schritt wurde das dadurch erhaltene Coumarin-Alkin mit Azid **35**^[215] in einer kupferkatalysierten Click-Reaktion

zum gewünschten Endprodukt **25** umgesetzt. Mit Hilfe von LC-MS-Analytik konnte gezeigt werden, dass die Reaktion relativ sauber und vollständig abgelaufen war. Um das im Reaktionsgemisch enthaltene, zytotoxische Kupfer abzutrennen, sollte das gewünschte Produkt über HPLC noch zusätzlich aufgereinigt werden, dabei kam es jedoch zu erheblichen Problemen. Zunächst wurde eine analytische HPL-Chromatographie durchgeführt, wobei in der Methode die Absorption bei 260 nm sowie bei 475 nm (Absorptionsmaximum der photolabilen Coumarin-Gruppe) gemessen wurde. In **Abbildung 4.14 A** ist in dem dabei erhaltenen Spektrum zu sehen, dass die krude Reaktionsmischung schon relativ sauber war. Um sicher zu gehen, dass es sich bei dem bei 25 Minuten eluierten Peak um das gewünschte Produkt **25** handelte, wurden davon Fraktionen nach der HPLC-Auftrennung gesammelt und per LC-MS analysiert. Jedoch konnten bei der massenspektrometrischen Analyse lediglich die einzelnen Fragmente von Hydroxylamin **24**, dem freien Thiocoumarin **32** und sehr wenig Produkt **25** ermittelt werden (**Abb. 4.14 B**).

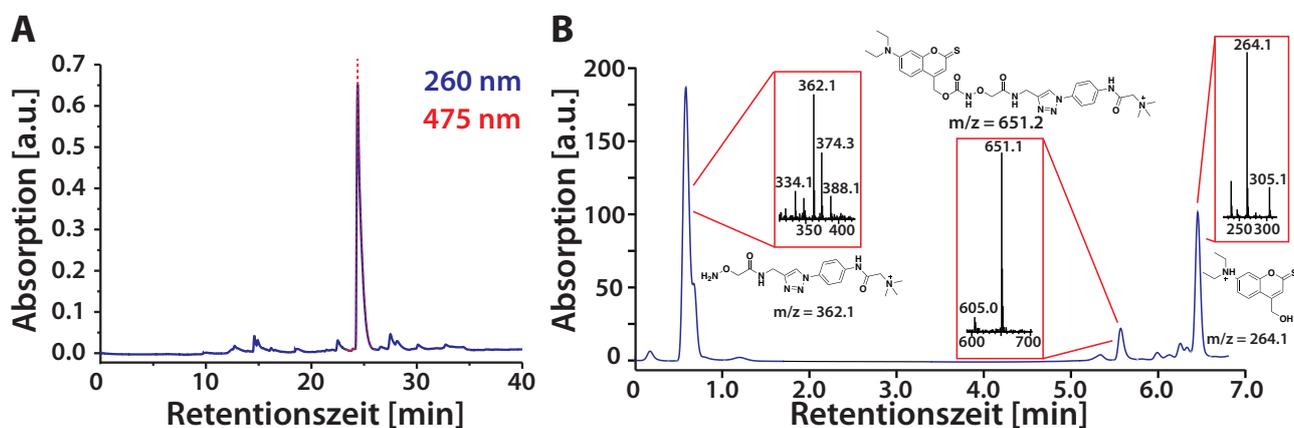
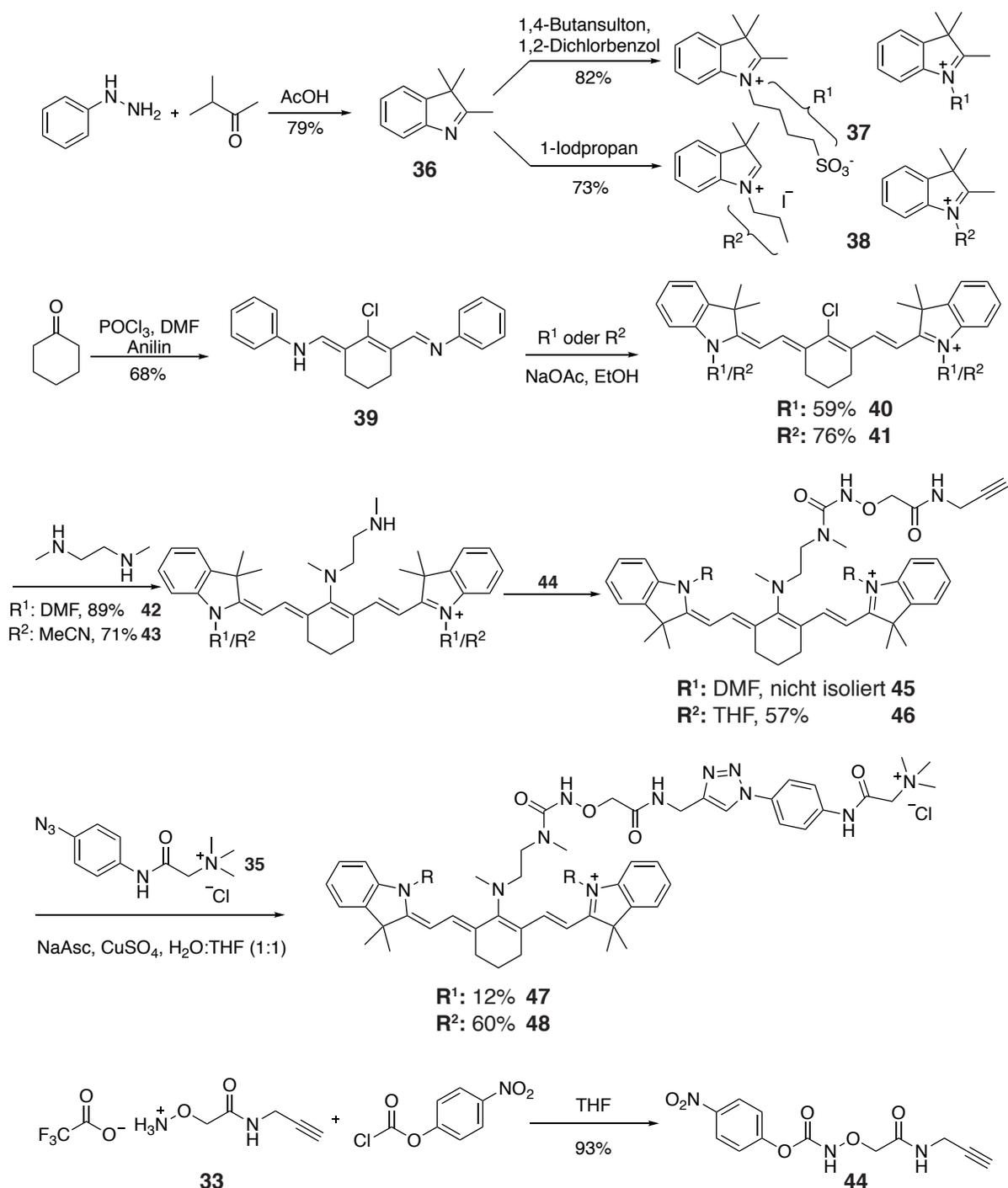


Abbildung 4.14: **A** HPL-Chromatogramm von **25**. **B** LC-MS-Analyse des in **A** bei 25 Minuten isolierten Peaks zeigt, dass **25** nach der Isolation sofort in seine Edukte **24** und **32** zerfällt.

In einer direkten Injektion der nicht aufgereinigten Reaktionsmischung in das UHPLC-MS-System wurde hingegen stets der Molekülpeak von **25** detektiert, somit erfolgte das *uncaging* von **25** höchst wahrscheinlich direkt nach der Isolation. Da bei der HPLC-Auftrennung mit dem UV/Vis-Detektor jeweils bei 260 nm und 475 nm gemessen wurde, liegt die Vermutung nahe, dass dadurch der Entschützungsprozess gestartet wurde. Daher wurde in einer erneuten HPLC-Auftrennung das UV/Vis-Spektrometer ausgeschaltet, sodass die Isolation von **25** unter komplettem Lichtausschluss geschah. Jedoch ergab sich bei der LC-MS-Analyse erneut das gleiche Bild. Offensichtlich ist das PPG-derivatisierte Hydroxylamin unter den Isolationsbedingungen nicht stabil, obwohl bei der Auftrennung ein scharfer Peak eluiert, der sich in der Retentionszeit deutlich von freiem Hydroxylamin **24** unterscheidet. Da der pH-Wert des Eluens zuvor nicht kontrolliert wurde und **25** im Reaktionsgemisch mit pH = 4.4 laut massenspektrometrischer Analyse stabil erscheint, wurde das Eluens im Folgenden mit Ammoniumformiat auf pH = 4.4 gepuffert.

Jedoch konnte auch hier kein intaktes Produkt **25** isoliert werden. Um sicherzustellen, dass hierfür nicht die Isolationsmethode über HPLC verantwortlich ist, wurden mehrere Versuche mit unterschiedlichen Elutionsmitteln an Umkehrphasen Silica Gel unternommen, jedoch blieben auch diese erfolglos. Folglich ist eine Coumarin-basierte PPG des Hydroxylamin-Testmoleküls nicht für die *in cellulo* Derivatisierung von abasischen Stellen geeignet, da es im isolierten Zustand zu instabil ist. Daher sollte eine neue Strategie für dieses Vorhaben entwickelt werden.



Schema 4.7: Synthese von Cyanin-geschütztem Hydroxylamin **47** und **48**.

Als Alternative wurde die Darstellung mit den von *Gorka et al.* entwickelten Cyanin-Derivaten forciert, deren *in cellulo* Anwendbarkeit ebenfalls schon gezeigt werden konnte.^[216] Diese können weiterhin im nahen Infrarotbereich entschützt werden und sind daher im Vergleich noch milder abspaltbar. Mit Licht im nahen IR-Bereich ist weiterhin eine erhöhte Durchdringung von Gewebe möglich, somit können diese PPG geschützten Derivate auch für Untersuchungen in Mäusen angewendet werden.^[217] Die dabei erfolgreich durchgeführte Synthese ist in **Schema 4.7** gezeigt. Das Ausgangsindol **36** wurde in einer Fischer-Indol-Synthese aus Phenylhydrazin und 3-Methylbutanon in refluxierendem Eisessig in guter Ausbeute dargestellt.^[218] Davon ausgehend konnten zwei alkylierte Derivate des Indols **36** durch Reaktion mit 1,4-Butansulton zu **37** und 1-Iodopropan zu **38** dargestellt werden.^[219,220] Benzamin **39** wurde in einer *Vilsmeier*-Formylierungs-Chlorierungs-Sequenz dargestellt und das dadurch entstandene Intermediat bei der Aufarbeitung mit Anilin abgefangen.^[221] **39** wurde daraufhin mit den Indolden **37** und **38** zu den CyaninGrundgerüsten **40** und **41** in guten Ausbeuten umgesetzt. Trotz der polaren Seitenketten in **40** und **41** konnten beide Cyanin-Derivate über Normalphasen-Säulenchromatographie aufgereinigt werden. Zur Vervollständigung der photolabilen Schutzgruppe wurden **40** und **41** analog zu *Gorka et al.* mit *N,N'*-Dimethylethylendiamin zu den aminierten Spezies **42** und **43** umgesetzt, wobei abweichend von der in der Literatur beschriebenen Synthese im Falle von **42** eine höhere Ausbeute durch Reaktion bei Raumtemperatur anstatt bei 80 °C erreicht werden konnte.^[216] An dieser Stelle wurde zunächst versucht, **42** und **43** mit *para*-Nitrophenolchloroformiat umzusetzen und das entstandene Intermediat mit dem in **3.2** vorgestellten Hydroxylamin **24** zu substituieren, so dass eine direkte Darstellung der gewünschten PPG-Addukte gewährleistet würde. Die *para*-Nitrophenol-Ester von **42** und **43** (hier nicht gezeigt) konnten zwar dargestellt werden, jedoch kam es bei der Umsetzung vom aktivierten Intermediat zu Problemen mit der Löslichkeit des Hydroxylamins **24**, welches lediglich in DMF oder polaren, protischen Lösemitteln löslich ist. Diese Strategie musste zunächst also verworfen werden. Daraufhin wurde das freie Alkin-Hydroxylamin-Fragment **33** (*vide supra*) über *para*-Nitrophenolchloroformiat zu **44** in sehr guter Ausbeute umgesetzt und die Cyanin-Gerüste **42** und **43** mit dem aktivierten Alkin zu **45** und **46** carbonyliert. Kupferkatalysierte Click-Reaktion von **45** mit Azid **35** lieferte das sulfonierte PPG-Derivat **47** in geringer Ausbeute, die vermutlich auf zweifache semipräparative HPLC-Aufreinigung zurückzuführen ist. Auf gleiche Art und Weise wurde **48** schließlich dargestellt, welches ebenfalls über HPLC aufgereinigt werden sollte. Überraschenderweise ist **48** jedoch so unpolar, so dass diese Methode der Aufreinigung nicht möglich war und nach Injektion des kruden Gemisches in ein semipräparatives HPLC-System kein Produkt eluierte. Reines Produkt **48** konnte jedoch über eine selbstgepackte Umkehrphasen-Säule bei leichtem Überdruck in hoher Reinheit und annehmbarer Ausbeute erhalten werden.

Zunächst wurden die verschiedenen Absorptionseigenschaften beider PPG-Derivate charakterisiert. In

Abbildung 4.15 ist gezeigt, dass **48** bei einem Maximum von $\lambda = 687$ nm (**A**) und **47** bei $\lambda = 686$ nm (**B**) und somit im nahen Infrarot (NIR) Bereich absorbieren. Beide PPGs zeigen weiterhin eine Fluoreszenz nach Anregung bei $\lambda = 690$ nm bei $\lambda_{Fluo} = 790$ nm (**48**) und bei $\lambda_{Fluo} = 786$ nm (**47**).

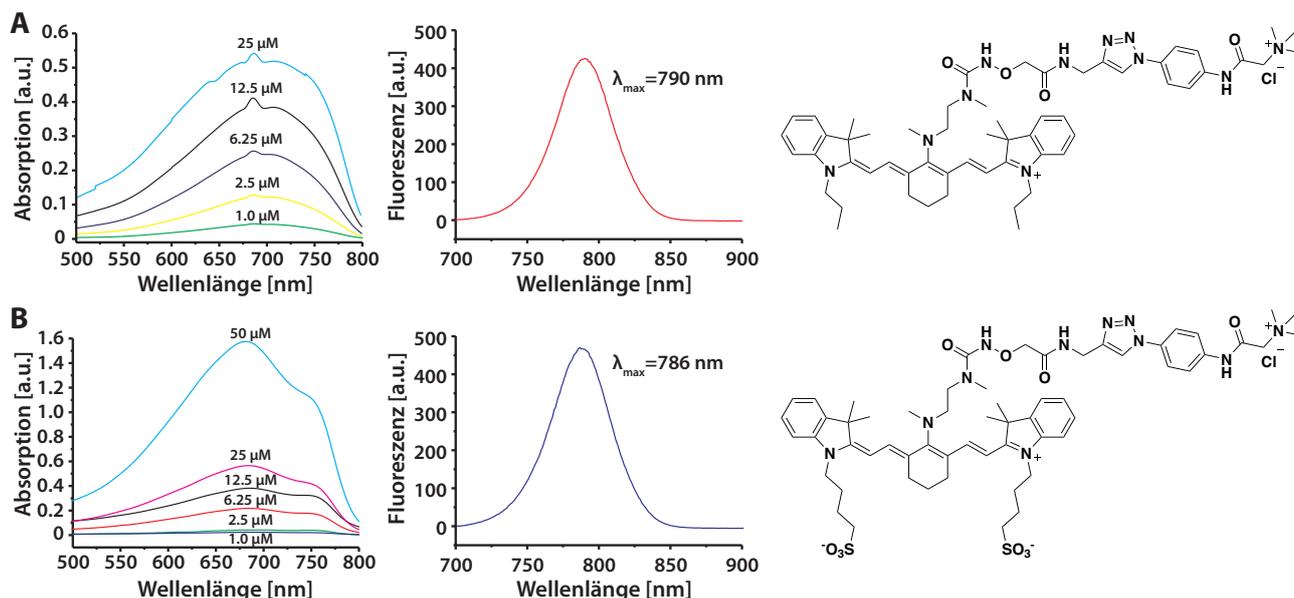


Abbildung 4.15: UV-Vis- und fluoreszenzspektroskopische Charakterisierung der PPG-Derivate. **A** und **B** Bestimmung der Absorption von **48** (**A**) und **47** (**B**) bei verschiedenen Konzentrationen in MeCN:H₂O (1:1). Rechts dazu die entsprechenden Fluoreszenzspektren bei einer Anregungswellenlänge von $\lambda = 690$ nm und einer Konzentration von jeweils $5.0 \mu\text{M}$.

Im nächsten Schritt sollte dann in einem *in vitro* Experiment untersucht werden, ob nach einer Entschützung des dargestellten PPG-Hydroxylaminderivats eine Derivatisierung abasischer Stellen möglich ist. Dafür wurde ein Oligodesoxynukleotid, welches eine einzelne, definierte abasische Stelle enthielt^[215] (siehe **Abb. 4.16 A**) mit **47** versetzt und daraufhin für 2 und 12 Stunden mit LED-Licht einer Wellenlänge von $\lambda = 700$ nm bestrahlt. Die Reaktion verlief außerdem unter physiologischen pH (7.4, 20 mM HEPES) und bei Raumtemperatur ab. Wie in **Abbildung 4.16 B** zu sehen ist, konnte nach zwei Stunden das ODN vollständig zum derivatisierten ODN umgesetzt werden. Die Absorption des derivatisierten ODNs ist weiterhin etwas höher als die in der unbestrahlten Kontrolle, was aufgrund des Einbaus von **24** als zusätzlichen Chromophor Sinn macht und zeigt, dass die Reaktion nach 2 Stunden beendet war. Bei der noch vor dem derivatisierten ODN eluierenden Spezies handelt es sich vermutlich um ein Fragment, welches aus der Entschützungreaktion entstammt. Dieses verschwindet weiterhin nach einer verlängerten Bestrahlung von 12 Stunden, wohingegen die Menge an derivatisiertem ODN konstant bleibt. Dies veranschaulicht, dass die DNA nicht durch eine Bestrahlung mit NIR-Licht geschädigt wird. Somit reagiert das freigesetzte Hydroxylamin **24** mit der im Doppelstrang eingebetteten, abasischen Stelle. Als Positivkontrolle wurde das ODN außerdem direkt mit dem freien Hydroxylaminreagenz **24**

umgesetzt, so dass ein direkter Vergleich anhand des HPL-Chromatogramms möglich war.

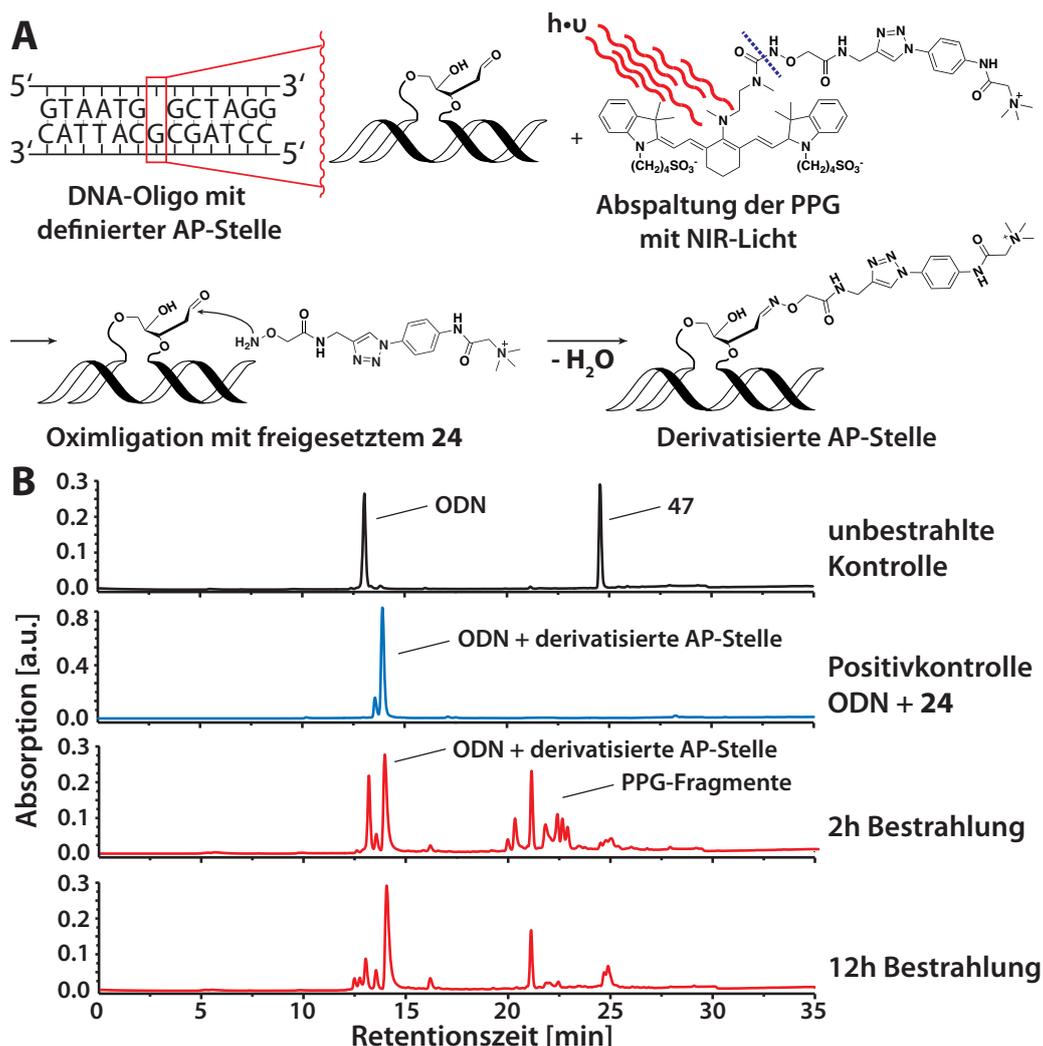


Abbildung 4.16: Derivatisierung von abasischen Stellen mit **47**. **A** Illustration des beschriebenen Experiments. Ein Oligodesoxynucleotid (ODN, 2.0 nmol) wird mit **47** (150 μ mol) unter physiologischen Bedingungen (20 mM HEPES, pH=7.4) versetzt und daraufhin mit Licht einer Wellenlänge von $\lambda = 700$ nm für mehrere Stunden bestrahlt und das erhaltene Produktgemisch mittels HPLC analysiert. **B** HPLC-Spuren der unterschiedlichen Injektionen. Als Positivkontrolle wurde das ODN (4.0 nmol) mit Hydroxylaminreagenz **24** (150 μ mol) versetzt und für 40 Minuten inkubiert.

Wie oben erwähnt konnte **48** aufgrund seiner überraschend niedrigen Polarität nicht von einer herkömmlichen C₁₈- oder C₈-Säule eluiert werden. In dem geplanten Experiment war allerdings auch eine gleichzeitige Beobachtung der DNA-Derivatisierung und der Vervollständigung des *uncaging*-Vorgangs beabsichtigt. Daher wurde dieses Experiment nur mit **47** durchgeführt. Der durch *Gorka et al.* beschriebene Entschützungsmechanismus der Cyanin-PPGs ist jedoch allgemein gültig und daher kann dieses Experiment stellvertretend für beide Derivate gesehen werden.^[216]

Nachdem die ersten *in vitro* Tests mit **47** sehr positiv waren, wurde in einem nächsten Schritt untersucht, ob die beiden Derivate **47** und **48** die Zellmembran durchdringen und somit eine *in cellulo*

Derivatisierung von abasischen Stellen potentiell möglich ist. Dazu wurden von *Angie Kirchner* zunächst CHOK1-Zellen vorbereitet und mit $100\ \mu\text{M}$ **47** für zwei Stunden inkubiert. Jedoch konnte dieses Derivat, anders als im vom *Nani et al.* gezeigten Fall,^[217] die Membran nicht durchdringen. Vermutlich war hier aufgrund der permanent doppelten negativen und doppelten positiven Ladungen die Ladungsdichte im Molekül zu groß, um in die Zellen zu gelangen. Daher wurde das zweite, unpolare Derivat **48** auf gleiche Art und Weise an CHOK1-Zellen verfüttert. Das mit $100\ \mu\text{M}$ **48** enthaltene Medium wurde nach 2 Stunden Inkubationszeit gegen frisches Medium ausgetauscht und die Zellen daraufhin mit einem Fluoreszenzmikroskop untersucht. **Abbildung 4.17 A** zeigt, dass **48** erfolgreich in CHOK1-Zellen aufgenommen wurde. Darüber hinaus wurden mESCs untersucht, die das interessantere System für die Untersuchung epigenetischer Vorgänge darstellen. Dabei wurde die Inkubationszeit auf 30 Minuten verkürzt und eine geringere Konzentration von $20\ \mu\text{M}$ **48** gefüttert.

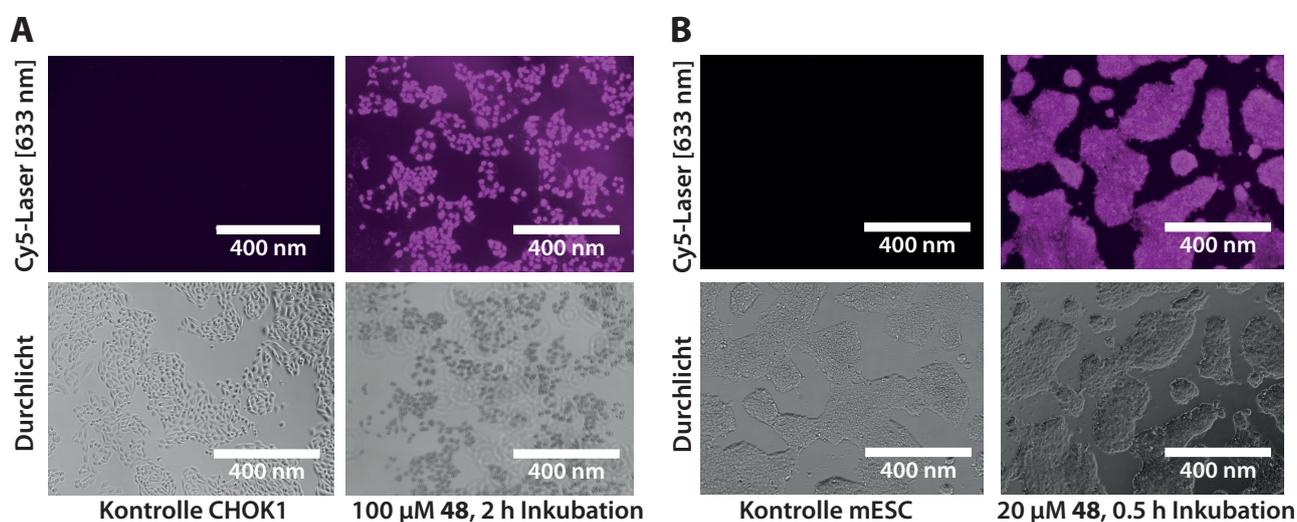


Abbildung 4.17: Fluoreszenzmikroskopische Aufnahmen von mit **48** behandelten Zellen. Als Anregungswellenlänge wurde die Cy5-Linie (633 nm) gewählt. **A** CHOK1-Zellen wurden mit $100\ \mu\text{M}$ **48** gefüttert und das Medium nach 2 Stunden Inkubationszeit gewechselt. **B** Inkubation von mESCs mit $20\ \mu\text{M}$ **48** für 30 Minuten zeigt, dass das geschützte Hydroxylamin-PPG-Derivat **48** die Zellmembran erfolgreich passieren kann.

Bei den in **Abbildung 4.27 B** gezeigten Bildern ist sichtbar, dass **48** im Zytosol der Zellen vorhanden ist und nicht in den Zellkern eindringt. Um zu sehen, ob eine Derivatisierung von abasischen Stellen möglich ist, wurden die mESC-Kulturen am Fluoreszenzmikroskop für zehn Minuten mit der Cy5-Laserlinie am Fluoreszenzmikroskop bestrahlt. Nach erfolgter Bestrahlung war keine Fluoreszenz nach Anregung bei 633 nm mehr sichtbar, was dafür sprach, dass **48** komplett entschützt war. Die Zellen wurden daraufhin von *Angie Kirchner* lysiert und die darin enthaltene genomische DNA isoliert. In enger Zusammenarbeit mit *Sarah Schiffers* wurde die genomische DNA schließlich, wie beschrieben, enzymatisch auf Nukleosidebene verdaut und am Triplequadrupol-Massenspektrometer analysiert.^[215] Jedoch konnten keine Oximligationsprodukte von abasischen Stellen oder β -eliminierten abasischen

Stellen detektiert werden. Für diese Beobachtung können nun einige Gründe angeführt werden. Zum einen war die Lokalisation von **48** bei den mESCs hauptsächlich im Zytosol zu sehen, die zur Untersuchung von abasischen Stellen benötigte genomische DNA befindet sich jedoch im Zellkern. Kommt es zu einer Freisetzung von Hydroxylamin **24** im Zytoplasma, so kann dieses sofort mit den dort befindlichen Aldehyden und Ketonen abreagieren. Da im Zytosol unter anderem die Glykolyse stattfindet, bei der neben D-Glukose auch Pyruvat und andere Aldehyde in relativ hohen Konzentrationen vorhanden sind, ist nicht ausgeschlossen, dass **24** mit diesen Metaboliten unter einer Oximligationsreaktion abreagiert und somit nicht mehr für die Derivatisierung von abasischen Stellen zur Verfügung steht. Zwar ist es prinzipiell möglich, dass das freigesetzte Hydroxylamin **24** schnell in den Zellkern ohne eine Reaktion mit den im Zytoplasma vorhandenen Aldehyden und Ketonen diffundieren könnte, da von *Anna-Skrollan Geiermann* allerdings schon eine schnelle Reaktion von **24** mit verschiedensten Aldosen ermittelt werden konnte, scheint dies unwahrscheinlich. Jedoch könnte in einem weiteren Versuch durch sukzessive Erhöhung der Konzentration von **48** versucht werden, genügend Hydroxylamin **24** in der Zelle freizusetzen, welches dann letztlich bis zur genomischen DNA vordringen und dort mit den abasischen Stellen reagieren kann. Jedoch sei erwähnt, dass bei einer Konzentration von $100\ \mu\text{M}$ **48** schon eine deutliche Lethalität der CHOK1-Zellen beobachtet wurde (siehe **Abbildung 4.17 A**). Eine Verkürzung der Inkubationszeit in Kombination mit der Fütterung einer höheren Konzentration an **48** könnte jedoch in einer geringeren Toxizität resultieren und zu einer erfolgreichen *in cellulo*-Derivatisierung von abasischen Stellen führen.

4.2.1 Projektbeiträge

In diesem Projektabschnitt war ich für die Synthese und Aufreinigung der drei gezeigten PPG-Hydroxylamin-Derivate verantwortlich. Außerdem führte ich erste Charakterisierungen und *in vitro* Tests durch. *Angie Kirchner* war für zellbiologische Arbeiten, wie zum Beispiel dem Verfüttern von **24** an CHOK1 sowie mESCs und für fluoreszenzmikroskopische Arbeiten verantwortlich. *Sarah Schiffers* führte erste LC-MS-Messungen der isolierten genomischen DNA nach Behandlung mit **48** durch.

4.2.2 Synthesevorschriften, Analytik und experimentelle Details

Absorptions- und Fluoreszenzmessungen

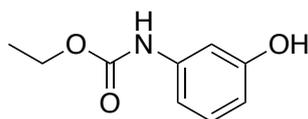
Zur Bestimmung der Absorption wurde eine $50\ \mu\text{M}$ - (**47**) bzw. $25\ \mu\text{M}$ -Lösung (**48**) in 50% MeCN/H₂O hergestellt. Nach jeder erfolgter Messung wurden die Lösungen sukzessive auf $12.5\ \mu\text{M}$, $6.25\ \mu\text{M}$, $2.5\ \mu\text{M}$ und letztendlich $1.0\ \mu\text{M}$ verdünnt. Zusätzlich wurde eine $25\ \mu\text{M}$ -Lösung von **48** für 5, 10, 15, 30 und 60 Minuten mit einer LED-Lampe von *ThorLabs* (700 nm) bestrahlt und dabei nach Ablauf der Zeit die Änderung der Absorption bestimmt (**Abb. 4.25 C**). Zur Bestimmung der Fluoreszenz wurde jeweils

eine 12.5 μM -Lösung von **47** und **48** bei $\lambda = 690$ nm bestrahlt und dabei mit Hilfe eines Fluorimeters das Fluoreszenzmaximum ermittelt.

***In vitro* Test zur Derivatisierung eines Oligodesoxynukleotid mit definierter abasischer Stelle**

Ein doppelsträngiges 13mer Oligodesoxynukleotid, welches eine definierte abasischer Stelle besaß (Darstellung siehe *Rahimoff et al.*^[215]) wurde mit bidestillierten H_2O und HEPES (pH = 7.5, Endkonzentration 20 mM) auf eine 0.2 nM Konzentration verdünnt. **47** (2.5 μM Endkonzentration) wurde hinzugegeben und die Reaktionslösung bei Raumtemperatur für 2 und 12 Stunden bestrahlt. Das Aliquot wurde nach Ablauf der Reaktionszeit mit 30 μL einer 1.0 mM HNEt_3OAc -Lösung verdünnt und dann mittels analytischer HPLC aufgetrennt. Dazu wurde ein Puffersystem aus 0.1 μM HNEt_3OAc in bidestilliertem H_2O (Puffer A) und 0.1 μM HNEt_3OAc in 80% MeCN/ H_2O verwendet. Die Flussrate betrug dabei 0.5 mL/min und das komplexe Gemisch konnte mit Hilfe einer C_{18} -Säule aufgetrennt werden. Dabei wurde zunächst ein Gradient von 0% \rightarrow 35% Puffer B in 20 Minuten und schließlich 35% \rightarrow 100% Puffer B von 20 auf 30 Minuten verwendet. Zur Visualisierung wurde die Absorption sowohl bei 260 nm als auch bei 700 nm beobachtet.

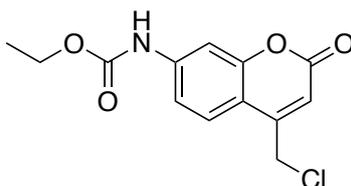
Synthese von Ethyl (3-hydroxyphenyl)carbammat (27)



3-Aminophenol (2.50 g, 22.5 mmol, 1.0 Äq) wurde in THF gelöst und mit Pyridin (2.30 mL, 28.2 mmol, 1.3 Äq) versetzt. Das Gemisch wurde mit einem Eisbad auf 0 °C gekühlt und daraufhin Ethylchloroformat (2.20 mL, 22.9 mmol, 1.0 Äq) langsam zugetropft. Nachdem das Gemisch zwei Tage bei Raumtemperatur rührte, wurde die Reaktion durch Zugabe von gesättigter wässriger NaHCO_3 -Lösung (50 mL) gestoppt und mit EtOAc (3 x 50 mL) extrahiert. Das Rohprodukt wurde über Na_2SO_4 getrocknet und das Lösemittel schließlich *in vacuo* entfernt. Dadurch wurden 3.58 g (19.8 mmol, 86%) von **27** als farbloser Feststoff erhalten, welcher ohne weitere Aufreinigung verwendet wurde.

Retentionsfaktor: $R_f = 0.25$ (20% EtOAc/ⁱHex).

Synthese von Ethyl-(4-(chloromethyl)-2-oxo-2H-chromen-7-yl)carbammat (28)



27 (2.83 g, 15.6 mmol, 1.0 Äq) und Ethyl-4-chloroacetoacetat (2.74 mL, 20.3 mmol, 1.3 Äq) wurden bei 0 °C in 80% H₂SO₄ (75 mL) gerührt. Die Lösung wurde daraufhin auf Raumtemperatur aufgewärmt und über Nacht gerührt. Das Gemisch wurde daraufhin in Eiswasser (75 mL) gegossen, das graue Präzipitat abfiltriert und mit Wasser (3 x 10 mL) gründlich gewaschen. Zur weiteren Aufreinigung wurde der Rückstand in Methanol (75 mL) aufgeschlämmt und bei 60 °C für 20 Minuten gerührt. Das noch warme Produktgemisch wurde erneut gefiltert, mit Methanol (3 x 10 mL) gewaschen und im Hochvakuum getrocknet. Dadurch wurde **28** als farbloser Feststoff (2.55 g, 9.05 mmol, 58%) erhalten.

Schmelzpunkt: T_m = 247 °C, Zersetzung.

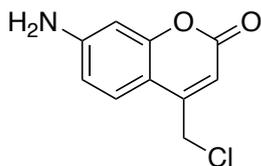
¹H-NMR (400 MHz, DMSO-d⁶), δ/ppm = 10.21 (s, 1H, OC(O)N-H), 7.76 (d, ³J = 8.7 Hz, 1H, C(5)-H), 7.59(dd, ⁴J = 2.0 Hz, 1H, C(8)-H), 7.42 (d, ³J = 8.8 Hz, 1H, C(6)-H), 6.51 (s, 1H, C(3)-H), 4.98 (s, 2H, C(4)CH₂Cl), 4.17 (q, ³J = 7.1 Hz, 2H, NHC(O)OCH₂CH₃), 1.26 (t, ³J = 7.1 Hz, 3H, NHC(O)OCH₂CH₃).

¹³C-NMR (101 MHz, DMSO-d⁶), δ/ppm = 159.9 (OC(2)(O)), 154.3 (NHC(O)OCH₂CH₃), 153.3 (C(4)CH₂Cl), 150.7 (C(8a)), 143.20 (C(7)), 125.9 (C(5)), 114.3 (C(6)-H), 112.7 (C(4a)), 111.6 (C3), 104.6 (C(8)), 60.80 (NHC(O)OCH₂CH₃), 41.2 (NHC(O)OCH₂CH₃), 14.4 (C(4)CH₂Cl).

HRMS (ESI): berechnet für [C₁₃H₁₃ClNO₄]⁺ (M+H)⁺: 282.0527; gefunden: 282.0529.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1}$ = 3282.9 (w), 3085.7 (w), 2992.7 (w), 2942.2 (w), 1705.5 (m), 1619.7 (m), 1589.9 (m), 1525.3 (m), 1405.2 (m), 1333.4 (m), 1222.0 (m), 1077.2 (m), 1015.7 (m), 982.6 (m), 862.0 (m), 847.0 (m), 763.7 (m), 703.5 (m).

Synthese von 7-Amino-4-(chloromethyl)-2H-chromen-2-on (26)



Zunächst wurde **28** (2.47 g, 8.77 mmol, 1.0 Äq) in einem Gemisch aus Eisessig und konzentrierter H_2SO_4 (1:1, à 9.0 mL) für zwei Stunden refluxiert. Das Gemisch wurde daraufhin in Eiswasser (75 mL) gegeben und mit festem K_2CO_3 neutralisiert. Die Extraktion der wässrigen Phase erfolgte mit EtOAc (3 x 150 mL) und vereinigte organische Phasen wurden mit gesättigter, wässriger NaCl- (150 mL) und NaHCO_3 -Lösung (150 mL) gewaschen. Nach Trocknen über Na_2SO_4 wurde das Lösemittel *in vacuo* entfernt und das Produkt **26** (1.80 g, 8.59 mmol, 98%) als farbloser Feststoff erhalten.

Schmelzpunkt: $T_m = 183 - 184$ °C.

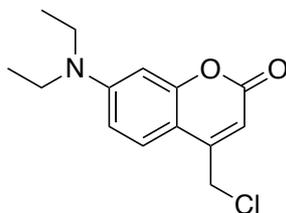
$^1\text{H-NMR}$ (400 MHz, DMSO-d_6), $\delta/\text{ppm} = 7.47$ (d, $^3J = 8.7$ Hz, 1H, C(5)-H), 6.58 (dd, $^3J = 8.7$ Hz, $^4J = 2.1$ Hz, 1H, C(6)-H), 6.43 (d, $^4J = 2.1$ Hz, 1H, C(8)-H), 6.24 (s, 2H, C(7)NH₂), 6.18 (s, 1H, C(3)-H), 4.87 (s, 2H, C(4)CH₂Cl).

$^{13}\text{C-NMR}$ (101 MHz, DMSO-d_6), $\delta/\text{ppm} = 160.7$ (OC(2)(O)), 156.0 (C(4)CH₂Cl), 153.4 (C(8a)), 151.3 (C(7)NH₂), 126.1 (C(5)-H), 111.3 (C(3)-H), 107.9 (C(7/8)), 106.1 (C(7/8)), 98.7 (C(8)-H), 41.5 (C(4)CH₂Cl).

HRMS (ESI): berechnet für $[\text{C}_{10}\text{H}_9\text{ClNO}_2]^+$ (M+H)⁺: 210.0316; gefunden: 210.0317.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1} = 3440.7$ (w), 3351.4 (w), 3239.7 (w), 1682.3 (m), 1602.1 (m), 1547.4 (m), 1450.0 (w), 1401.6 (m), 1337.2 (m), 1268.4 (w), 1214.3 (w), 1150.4 (w), 1051.6 (w), 999.4 (w), 932.6 (w), 891.8 (w), 848.1 (m), 821.8 (m), 733.7 (m), 695.7 (w).

Synthese von 4-(Chloromethyl)-7-(diethylamino)-2H-chromen-2-on (29)



Eine Mischung aus Acetaldehyd (1.57 g, 2.00 mL, 35.1 mmol, 9.2 Äq) und 3M wässriger H_2SO_4 wurde auf 0 °C gekühlt. Eine Suspension aus **26** (0.80 g, 3.82 mmol, 1.0 Äq), NaBH_4 (0.61 g, 16.0 mmol,

4.2 Äq) in THF (10 mL) wurde daraufhin mit einer Pasteurpipette zum Acetaldehyd-H₂SO₄-Gemisch gegeben und für 10 Minuten verrührt. Die Reaktion wurde durch Zugabe von gesättigter, wässriger NaHCO₃-Lösung (150 mL) gestoppt und mit festem K₂CO₃ neutralisiert. Die wässrige Phase wurde mit EtOAc (3 x 125 mL) extrahiert, die vereinigten organischen Phasen über Na₂SO₄ getrocknet und das Lösemittel *in vacuo* entfernt. Das Endprodukt konnte schließlich nach säulenchromatographischer Aufreinigung (30% EtOAc/ⁱHex) als gelber Feststoff (0.81 g, 3.06 mmol, 80%) in reiner Form erhalten werden.

Schmelzpunkt: T_m = 124 - 125 °C.

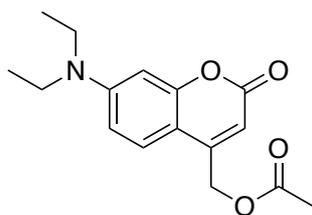
¹H-NMR (400 MHz, DMSO-d₆), δ/ppm = 7.57 (d, ³J = 9.0 Hz, 1H, C(5)-H), 6.73 (dd, ³J = 9.0 Hz, ⁴J = 2.5 Hz, 1H, C(6)-H), 6.54 (sd, ³J = 2.7 Hz, 1H, C(8)-H), 6.20 (s, 1H, C(3)-H), 4.90 (s, 2H, C(4)CH₂Cl), 3.43 (q, ³J = 7.0 Hz, 4H, C(7)N(CH₂CH₂)₂), 1.12 (t, ³J = 7.0 Hz, 6H, C(7)N(CH₂CH₃)₂).

¹³C-NMR (101 MHz, DMSO-d₆), δ/ppm = 160.7 (OC(2)(O)), 156.1 (C(8a)), 151.1 (C(4)CH₂Cl), 150.6 (C(7)N(CH₂CH₃)₂), 126.1 (C(5)-H), 108.8 (C(6)-H), 108.0 (C(3)-H), 105.5 (C(4a)), 96.9 (C(8)-H), 44.0 (C(7)N((CH₂CH₃)₂)), 41.4 (C(4)CH₂Cl), 12.3 (C(7)N((CH₂CH₃)₂)).

HRMS (ESI): berechnet für [C₁₄H₁₇ClNO₂]⁺ (M+H)⁺: 266.0942; gefunden: 266.0943.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1}$ = 3062.3 (w), 2946.8 (w), 2969.2 (w), 2928.7 (w), 1703.6 (m), 1603.3 (m), 1525.4 (m), 1441.7 (w), 1406.0 (m), 1350.7 (m), 1305.2 (m), 1277.8 (m), 1223.6 (m), 1196.8 (m), 1140.4 (m), 1092.6 (w), 1077.0 (m), 1043.4 (m), 1015.0 (w), 985.9 (w), 956.1 (w), 901.8 (w), 860.7 (w), 826.2 (m), 796.7 (w), 774.2 (m), 754.9 (w), 743.6 (w), 700.0 (w), 660.5 (w).

Synthese von (7-(Diethylamino)-2-oxo-2H-chromen-4-yl)methyl Acetat (**30**)



29 (497 mg, 1.87 mmol, 1.0 Äq) wurde in DMF (10 mL) gelöst und mit KOAc (220 mg, 2.24 mmol, 1.2 Äq) und Tetrabutylammoniumbromid (121 mg, 0.37 mmol, 0.2 Äq) versetzt. Die Reaktionsmischung wurde für 20 Minuten bei 50 °C gerührt und daraufhin mit gesättigter, wässriger NaCl-Lösung (150 mL) versetzt. Die wässrige Phase wurde dann mit EtOAc (2 x 75 mL) extrahiert und die vereinigten organischen Phasen über Na₂SO₄ getrocknet. Das somit erhaltenene Produkt **30** (537 mg, 1.86 mmol, 99%) wurde ohne weitere Aufreinigung für den nächsten Reaktionsschritt verwendet.

¹H-NMR (400 MHz, CDCl₃), δ/ppm = 7.24 (d, ³J = 9.0 Hz, 1H, C(5)-H), 6.54 (dd, ³J = 9.0 Hz,

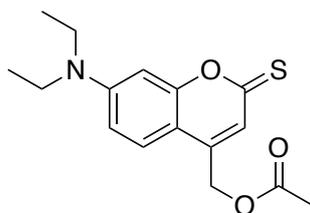
$^4J = 2.5$ Hz, 1H, C(6)-H), 6.45 (sd, $^3J = 2.7$ Hz, 1H, C(8)-H), 6.07 (s, 1H, C(3)-H), 5.17 (s, 2H, C(4)CH₂Cl), 3.37 (q, $^3J = 7.1$ Hz, 4H, C(7)N(CH₂CH₂)₂), 2.15 (s, 3H, Ac-CH₃), 1.16 (t, $^3J = 7.1$ Hz, 6H, C(7)N(CH₂CH₃)₂).

$^{13}\text{C-NMR}$ (101 MHz, CDCl₃), $\delta/\text{ppm} = 177.2$ (CH₃-C(O)), 161.9 (OC(2)(O)), 156.2 (C(8a)), 150.6 (C(4)CH₂O), 149.5 (C(7)N(CH₂CH₃)₂), 124.2 (C(5)-H), 108.6 (C(6)-H), 106.2 (C(3)-H), 105.9 (C(4a)), 97.7 (C(8)-H), 61.3 (C(4)CH₂O), 44.7 (C(7)N(CH₂CH₃)₂), 20.8 (AcCH₃), 12.4 (C(7)N((CH₂CH₃)₂)).

HRMS (ESI): berechnet für [C₁₆H₂₀NO₄]⁺ (M+H)⁺: 290.1387; gefunden: 290.1386.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1} = 2972.3$ (w), 1747.7 (m), 1706.2 (s), 1526.1 (m), 1848.7 (w), 1439.9 (m), 1414.4 (m), 1375.8 (m), 1355.1 (m), 1335.5 (m), 1271.4 (m), 1251.9 (m), 1239.8 (s), 1217.6 (m), 1195.8 (m), 1140.6 (m), 1073.8 (s), 1033.9 (m), 1012.1 (m), 989.0 (w), 932.2 (w), 898.1 (w), 858.8 (w), 840.2 (s), 822.7 (s), 810.8 (s), 793.1 (w), 744.5 (w), 704.0 (w), 664.9 (w).

Synthese von (7-(Diethylamino)-2-thioxo-2H-chromen-4-yl)methyl Acetat (31)



Zunächst wurde **30** (537 mg, 1.86 mmol, 1.0 Äq) in Toluol (30 mL) gelöst und Lawesson's Reagenz (489 mg, 1.21 mmol, 0.7 Äq) zugegeben. Das Reaktionsgemisch wurde daraufhin über Nacht refluxiert, wobei sich die Farbe der Lösung von orange zu dunkelgrün änderte. Nachdem das Lösemittel *in vacuo* entfernt wurde, konnte das Rohprodukt über Säulenchromatographie (25% EtOAc/ⁱHex) aufgereinigt werden, wobei **31** in quantitativer Ausbeute als dunkelorange Feststoff erhalten wurde (5.67 mg, 1.86 mmol).

Schmelzpunkt: $T_m = 118 - 120$ °C.

$^1\text{H-NMR}$ (400 MHz, CDCl₃), $\delta/\text{ppm} = 7.31$ (d, $^3J = 9.1$ Hz, 1H, C(5)-H), 7.01 (s, 1H, C(3)-H), 6.67 - 6.57 (m, 2H, C(6)-H + C(8)-H), 5.15 (s, 2H, C(4)-CH₂OH), 3.40 (q, $^3J = 7.1$ Hz, 4H, C(7)N(CH₂CH₃)₂), 2.17 (s, 3H, Ac-CH₃), 1.19 (t, $^3J = 7.1$ Hz, 6H, C(7)N(CH₂CH₃)₂).

$^{13}\text{C-NMR}$ (101 MHz, CDCl₃), $\delta/\text{ppm} = 197.0$ (OC(2)(S)), 170.1 (CH₃-C(O)), 158.9 (C(8a)), 150.9 (C(7)N(CH₂CH₃)₂), 141.9 (C(4)CH₂OAc), 124.4 (C(5)-H), 120.2 (C(3)-H), 110.2 (C(6)-H), 108.0 (C(4a)), 97.2 (C(8)-H), 60.8 (C(4)CH₂OAc), 44.8 (C(7)N(CH₂CH₃)₂), 20.7 (Ac-CH₃), 12.7 ((CH₂CH₃)₂).

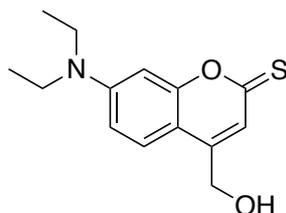
HRMS (ESI): berechnet für [C₁₆H₂₀NO₃S]⁺ (M+H)⁺: 306.1158; gefunden: 306.1156.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1} = 3074.4$ (w), 2969.4 (w), 2921.9 (w), 2893.9 (w), 1735.4 (m), 1626.4 (m), 1577.6 (m), 1515.9 (m), 1431.6 (m), 1380.3 (w), 1349.9 (m), 1292.8 (m), 1253.1 (m), 1236.5 (m), 1224.5 (m),

1198.9 (m), 1182.7 (m), 1163.0 (m), 1147.2 (m), 1120.4 (m), 1103.6 (m), 1076.2 (m), 1025.5 (m), 927.8 (m), 855.5 (m), 789.8 (m), 758.3 (m), 716.7 (w), 681.5 (w).

Elementaranalyse (CHNS): Berechnet für C₁₆H₁₉NO₃S: N: 4.59%, C: 62.93%, H: 6.27%, S: 10.50%; gefunden: N: 4.61%, C: 62.87%, H: 6.38%, S: 9.76%.

Synthese von 7-(Diethylamino)-4-(hydroxymethyl)-2H-chromen-2-thion (32)



Eine 3M Lösung von HCl in MeOH (1.55 mL, 4.65 mmol, 2.5 Äq) wurde mit trockenem MeOH (70 mL) verdünnt und **31** (568 mg, 1.86 mmol, 1.0 Äq) darin gelöst. Die Reaktionslösung wurde daraufhin bei 75 °C über Nacht gerührt und nach Beendigung der Reaktion das Lösemittel *in vacuo* entfernt. Das Produkt wurde schließlich über säulenchromatographische Aufreinigung (25% EtOAc/*i*Hex → 40% EtOAc/*i*Hex → 10% Aceton/DCM → Aceton) als dunkel-oranger Feststoff erhalten (510 mg, 1.76 mmol, 95%).

¹H-NMR (400 MHz, Aceton-d⁶), δ /ppm = 7.53 (d, ³*J* = 9.1 Hz, 1H, C(5)-**H**), 7.12 (s, 1H, C(3)-**H**), 6.81 (dd, ³*J* = 9.1 Hz, ⁴*J* = 2.6 Hz, 1H, C(6)-**H**), 6.66 (d, ⁴*J* = 2.5 Hz, 1H, C(8)-**H**), 4.81 (s, 2H, C(4)-**CH**₂OH), 4.64 (d, ³*J* = 6.4 Hz, 1H, C(4)CH₂**OH**), 3.54 (q, ³*J* = 7.1 Hz, 4H, C(7)N(**CH**₂CH₃)₂), 1.22 (t, ³*J* = 7.1 Hz, 6H, C(7)N(CH₂**CH**₃)₂).

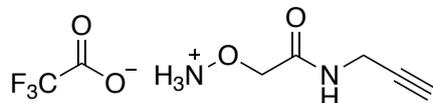
¹³C-NMR (101 MHz, Aceton-d⁶), δ /ppm = 198.3 (OC(2)(S)), 159.7 (C(8a)), 151.8 (C(7)N(CH₂CH₃)₂), 149.2 (C(4)CH₂OH), 125.8 (C(5)-H), 120.1 (C(3)-H), 111.1 (C(6)-H), 109.0 (C(4a)), 97.4 (C(8)-H), 60.2 (C(4)CH₂OH), 45.3 (C(7)N(CH₂CH₃)₂), 12.7 (C(7)N(CH₂CH₃)₂).

HRMS (ESI): berechnet für [C₁₄H₁₈NO₂S]⁺ (M+H)⁺: 264.1053; gefunden: 264.1053.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 3371.8 (w), 2960.9 (w), 2917.2 (w), 2849.9 (w), 2359.7 (w), 1626.2 (w), 1570.0 (w), 1424.6 (w), 1390.3 (w), 1376.7 (w), 1335.7 (w), 1295.5 (w), 1246.5 (w), 1231.0 (w), 1198.0 (w), 1177.8 (w), 1163.5 (w), 1145.5 (w), 1095.1 (w), 1074.6 (w), 1021.4 (w), 995.3 (w), 963.6 (w), 946.2 (w), 857.4 (w), 820.9 (w), 799.9 (w), 660.1 (w).

Elementaranalyse (CHNS): Berechnet für C₁₄H₁₇NO₂S: N: 5.32%, C: 63.85%, H: 6.51%, S: 12.17%; gefunden: N: 5.07%, C: 62.64%, H: 6.44%, S: 11.92%.

Synthese von *O*-(2-Oxo-2-(prop-2-yn-1-ylamino)ethyl)hydroxylammonium-2,2,2-trifluoroacetat (**33**)



Trityl-geschütztes Alkin **34** (100 mg, 0.27 mmol, 1.0 Äq) wurde in DCM (4.75 mL) suspendiert und TFA (0.25 mL) wurde tropfenweise unter Rühren hinzugefügt, wobei sich die Reaktionslösung sofort intensiv gelblich verfärbte. Triisopropylsilan (61 μ L, 0.30 mmol, 1.1 Äq) wurde als Kationen-*Scavenger* zugefügt, wobei sich die Lösung daraufhin innerhalb von 15 Minuten entfärbte. Wasser (25 mL) wurde zugegeben und die wässrige Phase mit DCM (3 x 10 mL) extrahiert. Die wässrige Phase wurde bei 60 °C mit Ethanol co-evaporiert und somit das entschützte Hydroxylamin **33** (57 mg, 0.24 mmol, 89%) als farbloses Öl isoliert.

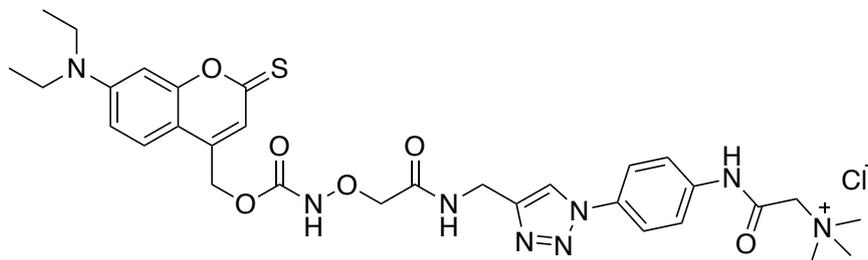
$^1\text{H-NMR}$ (400 MHz, D_2O), δ/ppm = 4.52 (s, 2H, $\text{H}_2\text{N-CH}_2\text{-C(O)}$), 3.99 (d, $^4J = 2.5$ Hz, 2H, $\text{NH-CH}_2\text{C}\equiv\text{CH}$), 2.58 (t, $^4J = 2.6$ Hz, 1H, $\text{NH-CH}_2\text{C}\equiv\text{CH}$).

$^{13}\text{C-NMR}$ (101 MHz, D_2O), δ/ppm = 169.3 ($\text{H}_2\text{N-CH}_2\text{-C(O)}$), 79.0 ($\text{NH-CH}_2\text{C}\equiv\text{CH}$), 72.1 ($\text{H}_2\text{N-CH}_2\text{-C(O)}$), 72.0 ($\text{NH-CH}_2\text{C}\equiv\text{CH}$), 28.4 ($\text{NH-CH}_2\text{C}\equiv\text{CH}$).

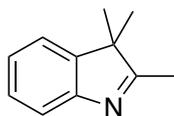
HRMS (ESI): berechnet für $[\text{C}_5\text{H}_9\text{N}_2\text{O}_2]^+$ ($\text{M}+\text{H}$) $^+$: 129.0659; gefunden: 129.0658.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1}$ = 3301.2 (w), 2946.7 (w, br), 2711.9 (w, br), 2126.2 (w), 1778.3 (w), 1658.9 (s), 1549.6 (m), 1428.7 (w), 1349.6 (w), 1273.6 (w), 1138.2 (s), 1083.1 (m), 1020.0 (w), 956.1 (w), 929.9 (w), 893.3 (w), 843.1 (w), 798.3 (m), 783.6 (w), 724.5 (m), 704.2 (w), 676.6 (w).

Synthese von 2-((4-(4-(1-(7-(Diethylamino)-2-thioxo-2H-chromen-4-yl)-3,7-dioxo-2,5-dioxo-4,8-diazanonan-9-yl)-1H-1,2,3-triazol-1-yl)phenyl)amino)-N,N,N-trimethyl-2-oxoethan-1-amoniumchlorid (25)



Zunächst wurde **32** (140 mg, 0.53 mmol, 1.0 Äq) in DCM (5 mL) gelöst und DMAP (64 mg, 0.53 mmol, 1.0 Äq) hinzugegeben. Die Reaktionslösung wurde daraufhin mit *para*-Nitrophenylchloroformat (107 mg, 0.53 mmol, 1.0 Äq) versetzt und bei Raumtemperatur für zwei Stunden verrührt. Das Gemisch wurde anschließend in gesättigte, wässrige NH₄Cl-Lösung (10 mL) gegeben und mit DCM (2 x 10 mL) extrahiert. Das Lösemittel wurde bei 30 °C *in vacuo* entfernt und das Rohprodukt für den nächsten Schritt ohne weitere Aufreinigung verwendet. Alkin **33** (128 mg, 0.53 mmol, 1.0 Äq) wurde in DMF (5 mL) gelöst, daraufhin wurde DMAP (129 mg, 1.06 mmol, 2.0 Äq) und eine Lösung aus dem zuvor carbonylierten Coumarin in DMF (5 mL) zugegeben. Nach Rühren des Gemisches für 30 Minuten bei Raumtemperatur wurde DCM (10 mL) zugegeben und das krude Gemisch mit gesättigter, wässriger NH₄Cl- (30 mL) und NaHCO₃-Lösung (30 mL) gewaschen. Die vereinigten organischen Phasen wurden über Na₂SO₄ getrocknet und das Lösemittel *in vacuo* entfernt. 33 mg (0.08 mmol, 1.0 Äq) des dabei erhaltenen, dunkel-orangen Prekursors wurden zusammen mit **35**^[215] (19 mg, 0.07 mmol, 0.9 Äq) in einem H₂O/THF-Gemisch (1:1, à 4 mL) gelöst und daraufhin mit Hilfe des *freeze-pump-thaw*-Verfahrens zwei Mal entgast. Dann wurde CuSO₄ (12 mg, 0.08 mmol, 1.0 Äq) hinzugegeben und ein weiteres Mal entgast. Schließlich wurde Natriumascorbat (6 mg, 0.032 mmol, 0.4 Äq) hinzugefügt und das Gemisch unter Lichtausschluss über Nacht bei Raumtemperatur verrührt. Die erfolgreiche Synthese des Produkts **25** konnte dabei über UHPLC-LRMS nachgewiesen werden, jedoch konnte das Produkt nicht isoliert werden (*vide supra*).

Synthese von 2,3,3-Trimethyl-3*H*-indol (36)

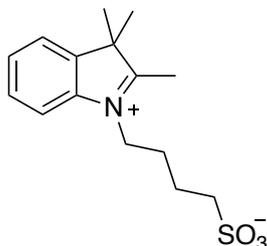
Phenylhydrazin (5.00 g, 46.2 mmol, 1.0 Äq) und 3-Methylbutenon (4.78 g, 5.9 mL, 55.4 mmol, 1.2 Äq) wurden in Eisessig (40 mL) gelöst und für eine Stunde bei 125 °C refluxiert. Das Lösemittel wurde bei 60 °C *in vacuo* entfernt, der Rückstand mit gesättigter, wässriger NaHCO₃-Lösung (50 mL) aufgeschlämmt und mit DCM (2 x 50 mL) extrahiert. Die vereinigten organischen Phasen wurden schließlich über Na₂SO₄ getrocknet und das Lösemittel unter vermindertem Druck abdestilliert. So wurde **36** als braunes Öl isoliert (5.76 g, 36.0 mmol, 79%).

¹H-NMR (400 MHz, CDCl₃), δ/ppm = 7.52 (dd, ³J = 7.6 Hz, ⁴J = 1.0 Hz, 1H, N-C-CH), 7.31 - 7.24 (m, 2H, CH-CH-CH-CH), 7.17 (td, ³J = 7.3 Hz, ⁴J = 1.0 Hz, 1H, CH-CH-CH-CH), 2.26 (s, 3H, N=C(CH₃)), 1.28 (s, 6H, C(CH₃)C(CH₃)₂).

¹³C-NMR (101 MHz, CDCl₃), δ/ppm = 188.0 (N=C-CH₃), 153.7 (CH-C-N), 145.6 (CH-C-C(CH₃)₂), 127.6 (C_{ar}), 125.1 (C_{ar}), 121.3 (C_{ar}), 119.9 (C_{ar}), 53.6 (C(CH₃)₂), 23.1 (CH-C-C(CH₃)₂), 15.4 (N=C-CH₃).

HRMS (ESI): berechnet für [C₁₁H₁₄N]⁺ (M+H)⁺: 160.1121; gefunden: 160.1121.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1}$ = 2961.9 (m), 2925.8 (w), 2865.1 (w), 1720.8 (m), 1618.8 (w), 1575.2 (m), 1510 (w), 1454.0 (s), 1428.2 (m), 1377.3 (m), 1360.1 (w), 1329.4 (w), 1300.0 (w), 1248.8 (w), 1201.0 (m), 1163.0 (w), 1152.5 (w), 1107.1 (w), 1013.9 (w), 935.4 (m), 861.0 (w), 791.2 (w), 771.9 (s), 749.9 (vs), 653.9 (m).

Synthese von 4-(2,3,3-Trimethyl-3*H*-indol-1-ium-1-yl)butan-1-sulfonat (37)

1,4-Butansulton (7.70 g, 56.4 mmol, 3.0 Äq) wurde in 1,2-Dichlorbenzol (5 mL) gelöst, mit **36** (3.0 g, 18.8 mmol, 1.0 Äq) versetzt und bei 120 °C für 3.5 Stunden verrührt. Das Lösemittel wurde bei 60 °C

in vacuo entfernt und der feste Rückstand in heißem Aceton umkristallisiert. Dadurch wurde **37** (4.55 g, 15.4 mmol, 82%) in Form eines gelblichen Feststoffs erhalten.

Schmelzpunkt: 242 °C (Zersetzung).

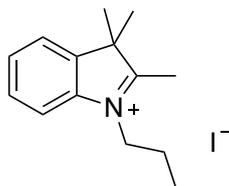
¹H-NMR (400 MHz, MeOD), δ /ppm = 8.02 (m, 1H, N-C-CH), 7.82 - 7.76 (m, 1H, C(CH₃)₂-C-CH), 7.70 - 7.62 (m, 2H), 4.89 (s, 3H, N=C(CH₃)), 4.59 (t, ³J = 7.8 Hz, 2H, CH₂SO₃), 2.92 (t, ³J = 7.2 Hz, 2H, N-CH₂), 2.18 (q, ³J = 7.8 Hz, 2H, CH₂CH₂-SO₃), 1.98 (q, ³J = 7.3 Hz, 2H, N-CH₂CH₂), 1.64 (s, 6H, C(CH₃)C(CH₃)₂).

¹³C-NMR (101 MHz, MeOD), δ /ppm = 196.7 (N=C-C-CH₃), 142.0 (CH-C-N), 141.2 (CH-C-C(CH₃)₂), 129.7 (C_{ar}), 129.1 (C_{ar}), 123.2 (C_{ar}), 115.3 (C_{ar}), 54.5 (C(CH₃)₂), 49.9 (N-CH₂), 47.5 (CH₂SO₃), 26.1 (CH₂CH₂-SO₃), 21.9 (N-CH₂CH₂), 21.4 (CH-C-C(CH₃)₂), 21.4 (N=C-CH₃).

HRMS (ESI): berechnet für [C₁₅H₂₂NO₃S]⁺ (M+H)⁺: 296.1315; gefunden: 296.1313.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 3423.2 (bw), 2938.8 (w), 1625.2 (w), 1477.0 (m), 1461.2 (m), 1396.9 (w), 1296.2 (w), 1257.5 (w), 1188.1 (vs), 1034.5 (s), 934.1 (w), 912.7 (w), 789.9 (m), 779.0 (s), 753.4 (m), 726.9 (m), 666.8 (w).

Synthese von 2,3,3-Trimethyl-1-propyl-3*H*-indol-1-iumiodid (**38**)



36 (2.00 g, 12.5 mmol, 1.0 Äq) wurde mit 1-Iodpropan (5.34 g, 3.10 mL, 31.4 mmol, 2.5 Äq) für zwei Stunden auf 120 °C erhitzt. Daraufhin wurde EtOAc (20 mL) hinzugefügt, die Lösemittel *in vacuo* entfernt und der Rückstand säulenchromatographisch (5% MeOH/DCM → 10% MeOH/DCM) aufgereinigt, wobei 2.60 g (7.9 mmol, 63%) von **38** in Form eines violetten Feststoffes erhalten wurden.

Schmelzpunkt: 145 - 146 °C.

¹H-NMR (400 MHz, CDCl₃), δ /ppm = 7.69 - 7.67 (m, 1H, N-C-CH), 7.59 - 7.48 (m, 3H, C(CH₃)₂-C-CH, CH-CH-), 4.59 (t, ³J = 7.6 Hz, 2H, N-CH₂), 3.06 (s, 3H, N=C(CH₃)), 1.98 (s, ³J = 7.5 Hz, 2H, N-CH₂CH₂CH₃), 1.60 (s, 6H, C(CH₃)C(CH₃)₂), 1.03 (s, 3H, CH₂CH₃).

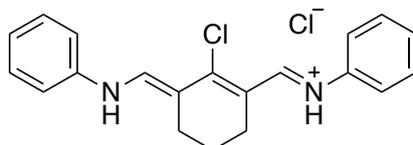
¹³C-NMR (101 MHz, CDCl₃), δ /ppm = 195.8 (N=C-C-CH₃), 141.6 (CH-C-N), 140.9 (CH-C-C(CH₃)₂), 130.3 (C_{ar}), 129.5 (C_{ar}), 123.5 (C_{ar}), 115.5 (C_{ar}), 54.7 (C(CH₃)₂), 51.3 (N-CH₂), 23.3 (CH-C-C(CH₃)₂), 21.6 (N-CH₂CH₂), 17.2 (N=C-CH₃), 11.4 /CH₂CH₃).

HRMS (ESI): berechnet für [C₁₄H₂₀N]⁺ (M)⁺: 202.1590; gefunden: 202.1589.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 3022.2 (w), 2977.2 (w), 2959.4 (w), 2929.0 (w), 2869.8 (w), 1623.4 (w), 1608.9

(m), 1594.4 (w), 1477.3 (m), 1469.0 (m), 1458.6 (m), 1432.3 (m), 1389.0 (w), 1374.0 (w), 1363.4 (m), 1292.6 (w), 1121.4 (m), 1030.8 (w), 1019.9 (w), 992.2 (w), 936.0 (m), 907.0 (w), 781.0 (s), 751.5 (m), 956.2 (m).

Synthese von *N*-((*E*)-((*E*)-2-chloro-3-((phenylamino)methylen)cyclohex-1-en-1-yl)methylen)benzenaminium Chlorid (39)



Zunächst wurde DMF (6.50 mL, 85.0 mmol, 3.2 Äq) vorgelegt und unter Kühlung im Eisbad Phosphorylchlorid (5.50 mL, 60 mmol, 2.3 Äq) tropfenweise über eine Spritze zugegeben. Nachdem die klare Lösung 30 Minuten bei Raumtemperatur gerührt hatte, wurde Cyclohexanon (2.25 mL, 26.5 mmol, 1.0 Äq) hinzugefügt und das Gemisch für drei Stunden unter Reflux erhitzt. Daraufhin wurden 9.0 mL einer Lösung aus Anilin und Ethanol (1:1) zugegeben und die Reaktion für weitere 30 Minuten bei Raumtemperatur verrührt. Zur Kristallisation wurde 1.5 M wässrige HCl (10 mL) zur dunkelvioletten Lösung gegeben und das Gemisch über zwei Tage bei 0 °C gelagert. Die dadurch entstandenen violetten Kristallnadeln wurden abgefiltert und mit wenig kalter, 1.5 M wässriger HCl und Diethylether gewaschen. Dadurch wurden 5.50 g (15.3 mmol, 58%) von **39** in Form dunkelvioletter Kristallnadeln erhalten.

Schmelzpunkt: 222 - 226 °C.

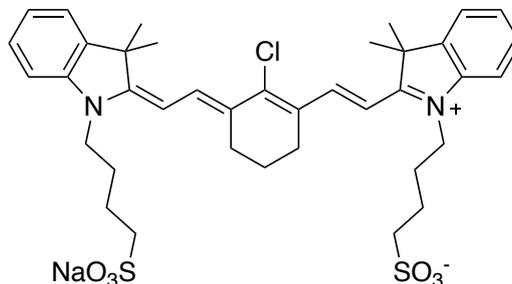
¹H-NMR (400 MHz, DMSO-d⁶), δ /ppm = 11.4 (bs, 2H, NH), 8.53 (s, 2H, NH-CH=C), 7.64 (d, ³*J* = 7.9 Hz, 4H, CH_{ar}), 7.46 (d, ³*J* = 7.8 Hz, 4H, CH_{ar}), 7.27 (t, ³*J* = 7.3 Hz, 2H, CH_{ar}), 2.77 (t, ³*J* = 6.2 Hz, 4H, CH₂CH₂CH₂), 1.84 (p, ³*J* = 6.2 Hz, 2H, CH₂CH₂CH₂).

¹³C-NMR (101 MHz, DMSO-d⁶), δ /ppm = 149.1 (NH-CH=C), 140.1 (C_{ar}), 130.1 (C_{ar}), 126.7 (C_{ar}), 123.4 (C-Cl), 119.4 (C_{ar}), 115.5 (C=CH), 25.4 (CH₂CH₂CH₂), 19.9 (CH₂CH₂CH₂).

HRMS (ESI): berechnet für [C₂₀H₂₀ClN₂]⁺ (M+H)⁺: 323.1310; gefunden: 323.1309.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 2865.4 (bw), 1619.8 (m), 1609.2 (m), 1565.9 (s), 1473.7 (s), 1461.0 (s), 1406.5 (m), 1274.5 (vs), 1196.5 (vs), 946.0 (m), 909.6 (m), 853.4 (m), 752.6 (s), 727.1 (m), 684.0 (m), 658.1 (m).

Synthese von Natrium 4-((*E*)-2-((*E*)-2-(2-Chloro-3-((*E*)-2-(3,3-dimethyl-1-(4-sulfonatobutyl)-3*H*-indol-1-ium-2-yl)vinyl)cyclohex-2-en-1-yliden)ethyliden)-3,3-dimethylindolin-1-yl)-butan-1-sulfonat (40)



Indol-Sulfonat **37** (1.64 g, 5.57 mmol, 2.0 Äq) wurde zusammen mit **39** (1.00 g, 2.78 mmol, 1.0 Äq) und NaOAc (0.46 g, 5.57 mmol, 2.0 Äq) für zwei Stunden in EtOH (150 mL) unter Reflux erhitzt. Das Lösemittel wurde daraufhin *in vacuo* entfernt und der dabei erhaltene Rückstand säulenchromatographisch (25% MeOH/DCM → 30% MeOH/DCM) aufgereinigt. **40** (1.20 g, 1.65 mmol) wurde dabei in 59% Ausbeute in Form eines hellgrün schimmernden Feststoffes isoliert.

Schmelzpunkt: 265 - 267 °C.

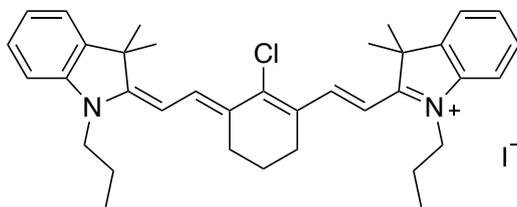
¹H-NMR (400 MHz, DMSO-d⁶), δ /ppm = 8.25 (d, ³*J* = 13.9 Hz, 2H, C-CH=CH), 7.62 (d, ³*J* = 7.3 Hz, 2H, CH_{ar}), 7.50 (d, ³*J* = 8.0 Hz, 2H, CH_{ar}), 7.43 (t, ³*J* = 7.7 Hz, 2H, CH_{ar}), 7.28 (t, ³*J* = 7.4 Hz, 2H, CH_{ar}), 6.38 (d, 14.2 Hz, 2H, C-CH=CH), 4.39 (t, ³*J* = 7.3 Hz, 4H, CH₂-SO₃), 3.71 ((t, ³*J* = 6.2 Hz, 4H, N-CH₂), 1.91 - 1.69 (m, 12 H, CH₂CH₂CH₂), 1.67 (s, 12H, C-CH₃).

¹³C-NMR (101 MHz, DMSO-d⁶), δ /ppm = 172.5 (N-C=C), 143.5 (C-CH=CH), 143.4 (CH=C) 142.5 (C_{ar}), 141.5 (C_{ar}), 129.1 (C_{ar}), 126.7 (C_{ar}), 125.6 (C_{ar}), 122.9 (C_{ar}), 112.1 (C_{ar}), 102.1 (CH=C), 51.1 (N-CH₂), 49.4 (CH₂SO₃), 28.0 (C-CH₃), 26.6 (CH₂), 26.3 (CH₂), 22.9 (CH₂).

HRMS (ESI): berechnet für [C₃₈H₄₈ClN₂O₆S₂]⁺ (M)⁺: 727.2637; gefunden: 727.2638.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 3385.9 (bw), 2920.1 (bw), 1545.7 (m), 1507.8 (w), 1482.5 (w), 1450.8 (w), 1428.2 (w), 1390.8 (m), 1350.4 (m), 1241.1 (m), 1157.4 (m), 1127.1 (m), 1100.0 (s), 1084.8 (s), 1027.1 (s), 989.0 (s), 893.0 (s), 747.9 (s), 709.2 (s).

Synthese von 2-((*E*)-2-((*E*)-2-Chloro-3-(2-((*E*)-3,3-dimethyl-1-propylindolin-2-yliden)ethy-
liden)cyclohex-1-en-1-yl)vinyl)-3,3-dimethyl-1-propyl-3*H*-indol-1-iumiodid (**41**)



39 (1.00 g, 2.78 mmol, 1.0 Äq) wurde zusammen mit **38** (1.83 g, 5.56 mmol, 2.0 Äq) und Natriumacetat (0.46 g, 5.57 mmol, 2.0 Äq) in EtOH (150 mL) gelöst. Nachdem das Gemisch für zwei Stunden bei 50 °C gerührt wurde, wurde das Lösemittel *in vacuo* entfernt. Der Rückstand konnte in 2% MeOH/DCM angelöst und mit selbiger Mischung säulenchromatographisch aufgereinigt werden. Dadurch wurden 1.40 g (2.10 mmol, 76%) **41** als schimmernd grüner Feststoff isoliert.

Schmelzpunkt: 248 - 250 °C.

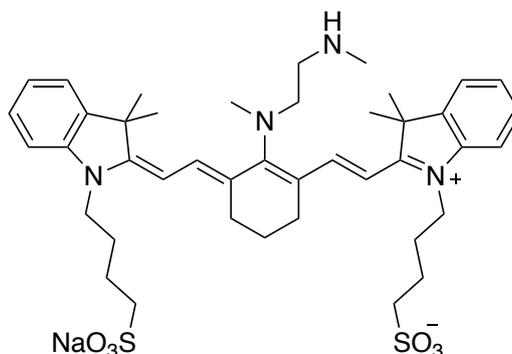
¹H-NMR (400 MHz, CDCl₃), δ/ppm = 8.47 (d, ³J = 14.1 Hz, 2H, C-CH=CH), 7.56 (d, ³J = 7.5 Hz, 2H, CH_{ar}), 7.46 (dd, ³J = 8.5 Hz, ⁴J = 1.2 Hz, 2H, CH_{ar}), 7.37 (d, ³J = 7.9 Hz, 2H, CH_{ar}), 7.32 (t, ³J = 7.5 Hz, 2H, CH_{ar}), 6.33 (d, 14.1 Hz, 2H, C-CH=CH), 4.19 (t, ³J = 7.3 Hz, 4H, CH₂CH₂CH₂), 3.33 (p, ³J = 1.6 Hz, 4H, CH₂CH₂CH₃), 2.76 (t, ³J = 6.2 Hz, 4H, N-CH₂), 1.92 (q, ³J = 7.4 Hz, 2H, CH₂CH₂CH₂), 1.76 (s, 12H, C-CH₃), 1.08 (t, ³J = 7.4 Hz, 6H, CH₂CH₂CH₃).

¹³C-NMR (101 MHz, CDCl₃), δ/ppm = 173.0 (N-C=C), 149.6 (CH=C), 144.1 (C-CH=CH), 142.3 (C_{ar}), 141.2 (C_{ar}), 128.5 (C_{ar}), 126.5 (C_{ar}), 125.1 (C_{ar}), 122.1 (C_{ar}), 111.0 (C_{ar}), 101.0 (CH=C), 49.3 (CH₂CH₂CH₂), 45.2 (CCl), 26.9 (C-CH₃), 25.9 (N-CH₂), 20.7 (CH₂CH₂CH₃), 20.5 (CH₂CH₂CH₂), 10.3 (CH₂CH₃).

HRMS (ESI): berechnet für [C₃₆H₄₄ClN₂]⁺ (M)⁺: 539.3188; gefunden: 539.3187.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 2962.8 (w), 2925.9 (w), 2867.3 (w), 1581.2 (w), 1551.1 (s), 1515.6 (m), 1480.6 (w), 1451.7 (m), 1427.7 (s), 1415.2 (s), 1399.7 (s), 1366.7 (s), 1352.7 (m), 1336.1 (m), 1249.0 (s), 1235.9 (s), 1168.9 (s), 1156.6 (s), 1113.7 (s), 1040.3 (s), 1018.7 (s), 934.2 (s), 857.5 (s), 816.5 (s), 757.9 (s), 701.5 (s).

Synthese von Natrium 4-((*E*)-2-((*E*)-2-(3-((*E*)-2-(3,3-Dimethyl-1-(4-sulfonatobutyl)-3*H*-indol-1-ium-2-yl)vinyl)-2-(methyl(2-(methylamino)ethyl)amino)cyclohex-2-en-1-ylidene)ethyliden)-3,3-dimethylindolin-1-yl)butan-1-sulfonat (42)



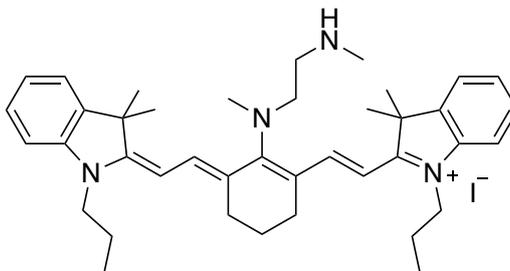
Cyaninderivat **40** (60 mg, 0.08 mmol, 1.0 Äq), *N,N'*-Dimethylethylendiamin (25 μ L, 0.20 mmol, 2.5 Äq) und DIPEA (14 μ L, 0.08 mmol, 1.0 Äq) wurden zunächst in DMF (3.0 mL) gelöst. Daraufhin wurde die Lösung bei Raumtemperatur für insgesamt 45 Minuten gerührt, wobei sich die dunkelgrüne Lösung in eine tiefblaue verfärbte. Das Gemisch wurde dann in Et₂O (40 mL) gegeben und der dabei ausgefallene, blaue Feststoff abfiltriert. Der Rückstand wurde in wenig MeOH aufgenommen und das Lösemittel *in vacuo* entfernt. Das reine Produkt konnte mittels säulenchromatographischer Aufreinigung (30% MeOH/DCM \rightarrow 35% MeOH/DCM) in einer Ausbeute von 89% (57 mg, 0.07 mmol) isoliert werden.

¹H-NMR (400 MHz, MeOD), δ /ppm = 7.61 (d, ³*J* = 13.6 Hz, 2H, C-CH=CH), 7.35 (d, ³*J* = 7.5 Hz, 2H, CH_{ar}), 7.28 (t, ³*J* = 7.8 Hz, 2H, CH_{ar}), 7.16 (d, ³*J* = 7.9 Hz, 2H, CH_{ar}) 7.10 (t, ³*J* = 7.4 Hz, 2H, CH_{ar}), 6.02 (d, 13.7 Hz, 2H, C-CH=CH), 4.04 (t, ³*J* = 7.0 Hz, 4H, CH₂-SO₃), 3.81 - 3.74 (m, 2H, N-CH₂CH₂-N), 3.26 (s, N-CH₃), 3.23 - 3.17 (m, 2H, N-CH₂CH₂-N), 2.76 (t, ³*J* = 7.4 Hz, 4H, N-CH₂), 2.65 (s, 3H, N-CH₃) 2.46 (m, 4H, CH₂CH₂CH₂CH₂-SO₃), 1.98 - 1.71 (m, 10 H, CH₂), 1.59 (s, 12H, C-CH₃).

¹³C-NMR (101 MHz, MeOD), δ /ppm = 170.7 (N-C=C), 142.6 (C-CH=CH), 142.4 (C-N(CH₃)), 140.5 (CH=C), 128.4 (C_{ar}), 125.5 (C_{ar}), 124.0 (C_{ar}), 121.9 (C_{ar}), 110.2 (C_{ar}), 98.2 (CH=C), 53.0 (N(CH₃)-CH₂CH₂), 50.5 (N-CH₂), 48.6 (CH₂), 44.3 (CH₂CH₂NH(CH₃)), 43.0 (CH₂SO₃), 32.6 (NH(CH₃)), 27.9 (C-CH₃), 25.8 (CH₂), 24.6 (CH₂), 22.1 (CH₂), 21.6 (CH₂).

HRMS (ESI): berechnet für [C₄₂H₅₉N₄O₆S₂]⁺ (M)⁺: 779.3871; gefunden: 779.3869.

Synthese von 2-((*E*)-2-((*E*)-3-(2-((*E*)-3,3-Dimethyl-1-propylindolin-2-yliden)ethyliden)-2-(methyl(2-(methylamino)ethyl)amino)cyclohex-1-en-1-yl)vinyl)-3,3-dimethyl-1-propyl-3*H*-indol-1-iumiodid (43)



41 (100 mg, 0.14 mmol, 1.0 Äq) und *N,N'*-Dimethylethylendiamin (75 μ L, 0.70 mmol, 5.0 Äq) wurden in MeCN (5.0 mL) bei 70 °C für zehn Minuten verrührt. Dabei konnte bei der Reaktionslösung ein Farbumschlag von dunkelgrün zu dunkelblau beobachtet werden. Nach Beendigung der Reaktion wurde das Lösemittel *in vacuo* entfernt und der ölige Rückstand in 5% MeOH/DCM aufgenommen. Durch säulenchromatographische Aufreinigung des Rohprodukts (5% MeOH/DCM) konnte **43** in reiner Form isoliert werden (72 mg, 0.10 mmol, 71%).

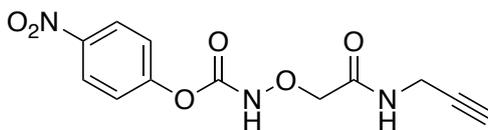
¹H-NMR (400 MHz, CDCl₃), δ /ppm = 7.59 (d, 3J = 13.4 Hz, 2H, C-CH=CH), 7.27 - 7.21 (m, 4H, CH_{ar}), 7.05 (t, 3J = 7.4 Hz, 2H, CH_{ar}), 6.91 (d, 3J = 8.0 Hz, 2H, CH_{ar}), 5.72 (d, 13.5 Hz, 2H, C-CH=CH), 3.93 (t, 3J = 6.4 Hz, 2H, N(CH₃)CH₂CH₂), 3.81 (t, 3J = 7.4 Hz, 4H, CH₂CH₂CH₂), 3.39 (s, 3H, N(CH₃)CH₂), 3.01 (t, 3J = 6.3 Hz, 2H, N(CH₃)CH₂CH₂), 2.41 - 2.38 (m, 7H, NH-CH₃, CH₂), 1.83 - 1.74 (m, 6H, CH₂), 1.60 (s, 12H, C-CH₃), 0.98 (t, 3J = 7.4 Hz, 6H, CH₂CH₂CH₃).

¹³C-NMR (101 MHz, CDCl₃), δ /ppm = 169.0 (N-C=C), 142.9 (CH=C), 140.3 (C-CH=CH), 128.4 (C_{ar}), 123.9 (C_{ar}), 123.5 (C_{ar}), 122.2 (C_{ar}), 109.3 (C_{ar}), 96.0 (CH=C), 53.5 (N-CH₂), 49.7 CH₂CH₂NH(CH₃), 45.2 (CH₂CH₂CH₂), 45.2 (N-CH₃), 35.7 (N-CH₃), 29.3 (C-CH₃), 24.6 (N-CH₂), 20.4 (CH₂), 11.7 (CH₂CH₃).

HRMS (ESI): berechnet für [C₄₀H₅₅N₄]⁺ (M)⁺: 591.4421; gefunden: 591.4415.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 2961.5 (m), 2928.0 (m), 2871.9 (w), 1547.9 (m), 1510.5 (m), 1485.0 (m), 1463.9 (m), 1451.9 (m), 1371.7 (s), 1346.3 (m), 1314.8 (m), 1274.0 (m), 1252.2 (m), 1228.6 (m), 1194.0 (s), 1164.4 (s), 1154.4 (s), 1124.8 (s), 1098.9 (s), 1060.3 (m), 1047.0 (s), 1018.7 (m), 1003.4 (m), 934.6 (m), 865.1 (w), 827.9 (w), 799.1 (m), 748.5 (w), 714.2 (m).

Synthese von 4-Nitrophenyl-(2-oxo-2-(prop-2-yn-1-ylamino)ethoxy)carbamat (44)



33 (160 mg, 0.66 mmol, 1.0 Äq) wurde zunächst in THF (5.0 mL) gelöst und auf 0 °C gekühlt. Der Lösung wurde dann NaH (60% in Mineralöl, 42 mg, 0.79 mmol, 1.2 Äq) zugegeben und die Reaktionsmischung bis zur Beendigung der Gasentwicklung verrührt. Daraufhin wurde *para*-Nitrophenylchloroformat (133 mg, 0.66 mmol, 1.0 Äq) bei 0 °C hinzugefügt und das Gemisch auf 10 °C erwärmt. Nach Rühren bei 10 °C für zehn Minuten war die Reaktion abgeschlossen und die Reaktionslösung wurde direkt auf eine Silica-Säule geladen. Das Produkt wurde mit 60% EtOAc/*i*Hex eluiert und dadurch konnten 180 mg von **44** (0.61 mmol, 93%) isoliert werden.

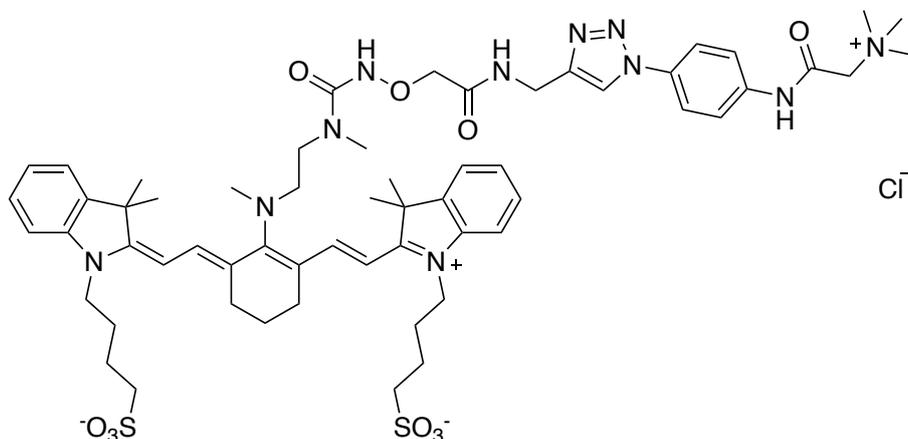
¹H-NMR (400 MHz, DMSO-d⁶), δ /ppm = 10.40 (s, 1H, C(O)-NH-O), 8.67 (t, ³J = 5.6 Hz, 1H, NH-CH₂), 8.13 (d, ³J = 9.1 Hz, 2H, CH_{ar}), 6.94 (d, ³J = 9.1 Hz, 2H, CH_{ar}), 4.25 (s, 2H, C(O)CH₂), 3.94 (dd, ³J = 5.6 Hz, ⁴J = 2.5 Hz, 2H, NH-CH₂-C), 3.16 (t, ⁴J = 2.5 Hz, 1H, CH_{alkin}).

¹³C-NMR (101 MHz, DMSO-d⁶), δ /ppm = 168.7 (OC(O)-NH), 164.4 (NH-C(O)CH₂), 161.6 (C_{ar}OC(O)), 140.0 (C_{ar}NO₂), 126.7 (C_{ar}), 116.3 (C_{ar}), 81.1 (CCH₂-NH), 75.0 (O-CH₂C(O)), 73.8 (C_{alkin}), 28.2 (NH-CH₂C).

HRMS (ESI): berechnet für [C₁₂H₁₂N₃O₆]⁺ (M)⁺: 294.0721; gefunden: 294.0721.

IR (ATR): $\tilde{\nu}$ /cm⁻¹ = 3408.7 (w), 3251.8 (m), 2936.9 (w), 2526.3 (w), 2303.4 (w), 1755.3 (s), 1651.1 (s), 1613.2 (m), 1590.9 (m), 1517.4 (s), 1489.5 (m), 1475.0 (m), 1426.8 (w), 1408.1 (m), 1335.7 (s), 1291.1 (m), 1266.0 (m), 1217.1 (s), 1163.9 (m), 1109.1 (m), 1022.6 (m), 975.9 (m), 954.4 (m), 931.4 (m), 862.5 (s), 813.6 (m), 765.5 (m), 746.8 (m), 722.0 (m), 680.5 (s), 660.8 (m).

Synthese von 4-((*E*)-2-((*E*)-2-(3-((*E*)-2-(3,3-Dimethyl-1-(4-sulfonatobutyl)-3*H*-indol-1-ium-2-yl)vinyl)-2-(methyl(8-methyl-3,7-dioxo-1-(1-(4-(2-(trimethylammonio)acetamido)-phenyl)-1*H*-1,2,3-triazol-4-yl)-5-oxa-2,6,8-triazadecan-10-yl)amino)cyclohex-2-en-1-yliden)ethyliden)-3,3-dimethylindolin-1-yl)butan-1-sulfonat (47)

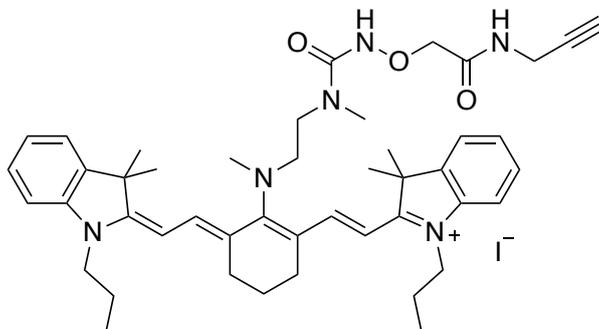


Das aminierte Cyaninderivat **42** (48 mg, 0.06 mmol, 1.0 Äq) wurde in DMF (3.0 mL) gelöst, mit **44** (18 mg, 0.06 mmol, 1.0 Äq) versetzt und bei Raumtemperatur über Nacht gerührt. Nachdem die Reaktion zum Alkinderivat **45** beendet war (ermittelt durch LC-MS), wurde DMF am Hochvakuum entfernt und das Rohprodukt in einer Mischung aus THF:H₂O (1:1, à 3.0 mL) aufgenommen. **35** (16 mg, 0.06 mmol, 1.0 Äq) wurde zugefügt und das Reaktionsgemisch zwei Mal über den *freeze-pump-thaw*-Zyklus entgast. Schließlich wurde durch Zugabe von CuSO₄ (10 mg, 0.06 mmol, 1.0 Äq), Natriumascorbat (5.0 mg, 0.02 mmol, 0.4 Äq) und erneutem Entgasen die Cycloadditionsreaktion gestartet. Nachdem das Gemisch über Nacht gerührt hatte, wurde das Lösemittel *in vacuo* entfernt und der Rückstand in wenig 20% MeCN/H₂O aufgenommen. Das komplexe Gemisch wurde schließlich über zweifache, semipräparative HPLC aufgereinigt, wobei 12% (8 mg, 0.007 mmol) an **47** in Form eines dunkelblauen Glases erhalten werden konnten.

¹H-NMR (800 MHz, MeOD), δ/ppm = 8.45 (s, 1H, CH_{triazol}), 7.83 (d, ³J = 8.4 Hz, 2H, CH_{ar}), 7.74 (d, ³J = 8.5 Hz, 2H, CH_{ar}), 7.59 (d, ³J = 13.4 Hz, 2H, C-CH=CH), 7.42 - 7.30 (m, 4H, CH_{ar}), 7.24 - 7.09 (m, 4H, CH_{ar}), 5.95 (d, 13.3 Hz, 2H, C-CH=CH), 4.62 (s, 2H, C(O)CH₂), 4.61 (s, 2H, C(O)CH₂N), 4.36 (s, 2H, NH-CH₂C), 4.04 (s, 3H, N(CH₃)C(O)), 4.21 (s, 2H, CH₂N(CH₃)₃), 3.61 (t, ³J = 7.3 Hz, 4H, N-CH₂-C), 3.35 (s, 3H, N(CH₃)CH₂), 3.41 (s, 9H, N(CH₃)₃), 3.39 (s, 3H, N(CH₃)CH₂), 2.91 (t, ³J = 7.0 Hz, 4H, CH₂-SO₃), 2.81 (s, 4H, CH₂), 2.46 (s, 4H, CH₂), 1.94 - 1.90 (m, 8 H, CH₂), 1.77 - 1.69 (m, 2H, CH₂), 1.52 (s, 12H, C-CH₃).

HRMS (ESI): berechnet für [C₅₉H₇₈N₁₁O₁₀S₂]⁻ (M-H)⁻: 1164.5380; gefunden: 1164.5389.

Synthese von 2-((*E*)-2-((*E*)-3-(2-((*E*)-3,3-dimethyl-1-propylindolin-2-ylidene)ethylidene)-2-(methyl(3-methyl-4,8-dioxo-6-oxa-3,5,9-triazadodec-11-yn-1-yl)amino)cyclohex-1-en-1-yl)vinyl)-3,3-dimethyl-1-propyl-3*H*-indol-1-iumiodid (**46**)



44 (9.0 mg, 0.028 mmol, 2.0 Äq) und Diamin **43** (10 mg, 0.014 mmol, 1.0 Äq) wurden in THF (2.0 mL) gelöst und über Nacht bei Raumtemperatur gerührt. Das Lösemittel wurde dann *in vacuo* entfernt und das dadurch erhaltene Rohprodukt säulenchromatographisch aufgereinigt (EtOAc → 5% MeOH/DCM). Dadurch wurden 7 mg (0.008 mmol, 57%) von **46** als tiefblaues Glas isoliert.

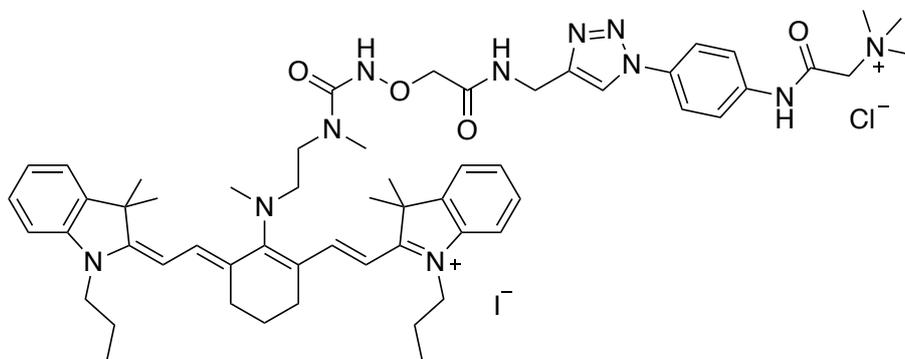
¹H-NMR (400 MHz, CDCl₃), δ/ppm = 9.75 (bs, 1H, C(O)-NH-O), 8.83 (t, ³J = 5.7 Hz, 1H, NH-CH₂), 7.51 (d, ³J = 13.4 Hz, 2H, C-CH=CH), 7.33 - 7.18 (m, 4H, CH_{ar}), 7.07 (t, ³J = 7.4 Hz, 2H, CH_{ar}), 6.90 (d, ³J = 8.0 Hz, 2H, CH_{ar}), 5.72 (d, 13.3 Hz, 2H, C-CH=CH), 4.31 (s, 2H, C(O)CH₂), 4.01 (dd, ³J = 5.7 Hz, ⁴J = 2.5 Hz, 2H, NH-CH₂-C), 3.88 - 3.73 (m, 6H, N(CH₃)CH₂CH₂, CH₂CH₂CH₂), 3.63 (t, ³J = 7.2 Hz, 2H, N(CH₃)CH₂CH₂), 3.46 (s, 3H, N(CH₃)CH₂), 3.35 (s, 1H, CH_{alkin}), 2.92 (s, 3H, N-CH₃), 2.41 (t, ³J = 6.6 Hz, 4H, N-CH₂), 1.83 - 1.74 (m, 6H, CH₂), 1.60 (s, 12H, C-CH₃), 0.99 (t, ³J = 7.4 Hz, 6H, CH₂CH₂CH₃).

¹³C-NMR (101 MHz, CDCl₃), δ/ppm = 173.7 (NC(O)-NH), 169.0 (N-C=C), 160.1 (NH-C(O)CH₂), 142.8 (CH=C), 140.4 (C-CH=CH), 128.4 (C_{ar}), 124.0 (C_{ar}), 123.7 (C_{ar}), 122.3 (C_{ar}), 109.3 (C_{ar}), 96.2 (CH=C), 80.3 (CCH₂-NH), 76.0 (O-CH₂C(O)), 70.4 (C_{alkin}), 55.8 (N-CH₂), 53.3 CH₂CH₂NH(CH₃), 48.3 (CH₂CH₂CH₂), 45.3 (N-CH₃), 35.4 (N-CH₃), 29.3 (C-CH₃), 28.3 (NH-CH₂C), 24.8 (N-CH₂), 20.4 (CH₂), 11.8 (CH₂CH₃).

HRMS (ESI): berechnet für [C₄₆H₆₁N₆O₃]⁺ (M)⁺: 745.4800; gefunden: 745.4798.

IR (ATR): $\tilde{\nu}/\text{cm}^{-1}$ = 2963.1 (w), 1671.3 (w), 1547.1 (m), 1510.0 (m), 1484.8 (m), 1452.2 (m), 1372.7 (s), 1347.4 (m), 1315.2 (m), 1274.3 (m), 1252.3 (m), 1229.1 (m), 1195.4 (m), 1165.9 (s), 1154.3 (s), 1125.5 (s), 1099.9 (vs), 1060.0 (m), 1047.4 (m), 1018.2 (m), 1003.0 (w), 933.3 (m), 800.1 (m), 714.1 (m).

Synthese von 2-((*E*)-2-((*E*)-3-(2-((*E*)-3,3-dimethyl-1-propylindolin-2-ylidene)ethylidene)-2-(methyl(8-methyl-3,7-dioxo-1-(1-(4-(2-(trimethylammonio)acetamido)phenyl)-1*H*-1,2,3-triazol-4-yl)-5-oxa-2,6,8-triazadecan-10-yl)amino)cyclohex-1-en-1-yl)vinyl)-3,3-dimethyl-1-propyl-3*H*-indol-1-iumiodid (48)



Alkin **46** (8.0 mg, 0.009 mmol, 1.0 Äq) und Azid **35** (3.0 mg, 0.009 mmol, 1.0 Äq) wurden in einem THF:H₂O-Gemisch (1:1, à 1.0 mL) gelöst und zwei Mal über die *freeze-pump-thaw*-Methode entgast. Daraufhin wurde dem Gemisch CuSO₄ (2.0 mg, 0.009 mmol, 1.0 Äq) zugegeben, die Mischung erneut *freeze-pump-thaw*-entgast und die Reaktion letztendlich durch die Zugabe von Natriumascorbat (1.5 mg, 0.007 mmol, 0.4 Äq) gestartet. Die Reaktion wurde daraufhin für zwei Tage bei Raumtemperatur gerührt und die Lösemittel schließlich *in vacuo* entfernt. Zur Aufreinigung wurde eine selbstgepackte Umkehrphasen-Silica-Säule verwendet, wobei diese zunächst mit 80% MeCN/H₂O + 0.05% Ameisensäure equilibriert, mit 30% MeCN/H₂O + 0.05% Ameisensäure gewaschen und das Produkt schließlich mit 40% MeCN/H₂O + 0.05% Ameisensäure eluiert wurde. Dadurch konnten 60% (6.0 mg, 0.0054 mmol) von **48** in Form eines dunkelblauen Feststoffes isoliert werden.

¹H-NMR (400 MHz, MeOD), δ/ppm = 8.31 (s, 1H, CH_{triazol}), 7.85 (d, ³J = 8.4 Hz, 2H, CH_{ar}), 7.63 (d, ³J = 8.5 Hz, 2H, CH_{ar}), 7.53 (d, ³J = 13.4 Hz, 2H, C-CH=CH), 7.31 - 7.18 (m, 4H, CH_{ar}), 7.03 (d, ³J = 7.9 Hz, 4H, CH_{ar}), 5.83 (d, 13.3 Hz, 2H, C-CH=CH), 4.48 (s, 2H, C(O)CH₂), 4.41 (s, 3H, N(CH₃)C(O)), 4.21 (s, 2H, CH₂N(CH₃)₃), 3.87 (t, ³J = 7.3 Hz, 4H, N-CH₂-C), 3.35 (s, 3H, N(CH₃)CH₂), 3.30 (s, 12H, N(CH₃)₃, CH₂), 2.70 (s, 2H, NH-CH₂), 2.37 (t, ³J = 6.7 Hz, 4H, N-CH₂), 1.75 - 1.65 (m, 8H, CH₂), 1.52 (s, 12H, C-CH₃), 0.92 (t, ³J = 7.4 Hz, 6H, CH₂CH₂CH₃).

¹³C-NMR (101 MHz, MeOD), δ/ppm = 173.2 (NC(O)-NH), 170.4 (N-C=C), 162.5 (NH-C(O)CH₂), 160.6 (NH-C(O)-CH₂), 153.1 (C_{ar}), 145.2 (C_{ar}), 142.9 (CH=C), 140.4 (C-CH=CH), 138.6 (C_{ar}), 133.2 (C_{ar}), 128.1 (C_{ar}), 123.0 (C_{ar}), 121.8 (C_{ar}), 121.1 (C_{ar}), 120.1 (CH_{triazol}), 120.4 (C_{triazol}), 109.4 (C_{ar}), 96.0 (CH=C), 75.2 (CCH₂-NH), 64.9 (O-CH₂C(O)), 55.3 (N-CH₂), 54.5 (CH₂CH₂NH(CH₃),

4 Unveröffentlichte Arbeiten

53.4 (N(CH₃)₃), 52.5 (CH₂), 46.6 (CH₂CH₂CH₂), 46.4 (N-CH₃), 46.2 (N(CH₃)₃), 33.8 (NH-CH₂C(O)), 31.7 (N-CH₃), 28.0 (C-CH₃), 28.3 (NH-CH₂C), 23.5 (N-CH₂), 21.6 (CH₂), 20.0 (CH₂), 10.4 (CH₂CH₃).
HRMS (ESI): berechnet für [C₅₇H₇₇N₁₁O₄]⁺ (M)²⁺: 489.8075; gefunden: 489.8071.

5 Material und Methoden

Chemische Synthesen und verwendete Chemikalien

Sofern dies explizit nicht anders angegeben wurde, wurden alle Reaktionen unter Verwendung der Standard-Schlenktechnik durchgeführt. Glasgeräte wurden vor ihrer Verwendung unter Hochvakuum mit einer Heißluftpistole bei 650 °C ausgeheizt und unter Vakuum ausgekühlt. Für die Synthesen wurden generell über Molsieb getrocknete Lösemittel der Firma *Sigma Aldrich* (< 50 ppm H₂O) verwendet, welche nur unter trockenen Bedingungen und in einem leichten Stickstoff-Gegenstrom entnommen wurden. Darstellungen, welche keiner Schutzgasatmosphäre bedurften, wurden in Lösemittel mit HPLC-Güte durchgeführt. Für Extraktionen und säulenchromatographische Aufreinigungen der Rohprodukte wurden Lösemittel technischer Güte verwendet, welche nicht destilliert waren. Chemikalien wurden von den Firmen *Fluka*, *Acros*, *Sigma-Aldrich*, *Carbolution*, *Carbosynth* und *TCI* in den gängigen Qualitätsstufen puriss, p.a. und purum bezogen.

CO-Insertionen mit einem Druckautoklav

Zur Synthese der verschiedenen fdC-Derivate wurde in einem Druckautoklav eine CO-Atmosphäre von 3.5 bar erzeugt. Der dabei verwendete Reaktor war ein *Miniclav*-Glas-Druckreaktor mit einem Borosilikat-Glas 3.3 der Firma *Büchi*. Zur Zugabe des Tri-*n*-butylzinnhydrids wurde eine Spritzenpumpe der Firma *KD Scientific* verwendet.

Dünnschichtchromatographie (DC)

Zur Reaktions- und Trennkontrolle bei säulenchromatographischen Aufreinigungen wurden DC-Platten F254 (Kieselgel auf Aluminium) der Firma *Merck* verwendet. Amine konnten dabei durch Anfärben der Substanz mit einer Lösung aus Ninhydrin (0.3 g Ninhydrin in 3.0 mL Eisessig und 100 mL *n*-Butanol) sichtbar gemacht werden. Aromatische sowie UV- und fluoreszenzaktive Substanzen wurden unter einer UV-Lampe ($\lambda = 254$ nm, 366 nm) beobachtet. Des Weiteren wurden Substanzen mit CAN-Lösung (Cer(IV)-ammoniummolybdat-Lösung, 120 g (NH₄)₆Mo₇O₂₄ · 24H₂O und 5.0 g Ce(NH₄)₄(NO₃)₆ in 800 mL H₂SO₄, 10%-ig) angefärbt. Die verwendeten Lösungen wurden alle unter Lichtausschluss aufbewahrt.

Normalphasen-Säulenchromatographie

Rohprodukte wurden standardmäßig über Normalphasen-Säulenchromatographie unter Verwendung von Kieselgel 60 (Korngröße: 0.063 - 0.200 mm) der Firma *Merck* aufgereinigt. Der Säule wurde dabei ein leichter Überdruck mit Hilfe von Stickstoff angelegt (*Flash*-Methode). Fraktionen wurden in Reagenzgläsern gesammelt und über DC beziehungsweise LCMS analysiert.

Rotationsverdampfer

Zur *in vacuo* Destillation der Lösemittel wurde ein *Laborota 4000* der Firma *Heidolph* verwendet. Sofern nicht abweichend angegeben wurde das Wasserbad auf 40 °C geheizt.

Lyophilisation

Gefriertrocknung wurde mit Hilfe einer *Alpha 2-4 LD plus*-Lyophille der Firma *Christ* realisiert. Die zu entfernenden Lösemittel wurden dabei zuvor mit flüssigem Stickstoff eingefroren.

Bidestilliertes Wasser

Für mobile Phasen der HPLC-Analysen und für biochemische Zwecke wurde das Wasser mit einer MilliQ-Anlage der Firma *Sartorius* verwendet.

Schmelzpunktanalyse

Schmelzpunkte kristalliner Feststoffe wurden an einem *Büchi Melting Point B-540* aufgenommen.

High Performance Liquid Chromatography (HPLC & UHPLC)

Semipräparative HPLC-Aufreinigungen wurden an einem *Breeze*-System von *Waters* (2487 *Dual Array Detector*, 1525 *Binary HPLC Pump*) durchgeführt. Dabei wurde eine C₁₈-Säule (VP 250/10 *Nucleosil 100/7 C₁₈*) der Firma *Macherey & Nagel* bei einem Fluss von 5.0 mL/min verwendet. Analytische HPLC wurde an einer *Waters Alliance 2695* mit *Photodiode Array Detector 2996* durchgeführt. Zur Auftrennung wurde eine VP 250/10 *Nucleodur 100-5 C₁₈ eC-* beziehungsweise eine CC 250/4 *Nucleosil N 120-3 C₁₈*-Säule von *Machery & Nagel* verwendet, wobei ein Fluss von 0.5 mL/min angewendet wurde. Als flüssige Phasen wurden bidestilliertes Wasser und Acetonitril bzw. davon gepufferte Lösungen verwendet. Die Puffersysteme sind in den jeweiligen Synthesvorschriften konkretisiert (*vide supra*).

pH-Meter

Zur Einstellung des pH-Wertes der verwendeten Puffer wurde ein *MP 220* von *Mettler Toledo* verwendet, welches vor jeder Messung mit einer pH = 4.0 und pH = 7.0 Standardlösung kalibriert wurde.

Kernresonanzspektroskopie (NMR)

¹H-NMR, ¹³C- und 2D-Spektren (COSY, HSQC, HMBC) wurden größtenteils selbstständig an einem *Bruker Ascend 400* oder von der Analytikabteilung für Kernresonanzspektroskopie des Departments Chemie der Ludwig-Maximilians-Universität München von *David Stephenson*, *Claudia Dubler* und *Petra Keilholz* aufgenommen. Dabei wurde ein *Varian Inova 400*, *Bruker ARX 600* oder *Bruker Avance III HD* (800 MHz) verwendet. Die chemischen Verschiebungen sind in ppm (*parts per million*) angegeben, welche sich auf die Mitte des jeweiligen Kopplungsmusters beziehungsweise auf die Ausdehnung des Multipletts beziehen. Als interner Standard relativ zu den chemischen Verschiebungen dienen die verwendeten deuterierten Lösemittel. Die Multiplizitäten der Signale werden als s (Singulet), d (Duplet), t (Triplet), q (Quartet), p (Quintet), s (Sechstet), m (Multiplet) bzw. deren Kombinationen angegeben. Außerdem sind die Angaben der Multiplizitäten phänomenologisch, das heißt, dass die Daten entsprechend ihrer Erscheinung und nicht ihrer theoretisch zu erwartenden Multiplizität angegeben wurden. Kopplungskonstanten sind in Hertz (Hz) angegeben und beziehen sich, sofern nicht anders angegeben, auf H-H-Nah- und Fernkopplungen.

Massenspektrometrie (LC-MS, HR-ESI- und EI-MS)

HR-ESI- und EI-MS-Analysen wurden von der Analytikabteilung für Massenspektrometrie des Departments Chemie der Ludwig-Maximilians-Universität München von *Werner Spahl*, *Sonja Kosak* und *Carola Draxler* angefertigt. Es wurde dabei ein LTQ-FTI-CR (ESI) der Firma *Finnigan* verwendet. EI Spektren wurden an eine *QExactive* der Firma *Thermo Fischer* angefertigt. Zur Reaktionskontrolle wurden selbstständig LC-MS-Analysen an einem *Dionex micro UHPLC*-System mit gekoppeltem *MSQ* Single-Quadrupol durchgeführt. Als mobile Phasen wurde dabei bidestilliertes Wasser und MeCN mit jeweils 0.05% Ameisensäure verwendet.

Infrarotspektroskopie (IR)

Zur Aufnahme von IR-Spektren wurde ein *BX FT-IR* System von *Perkin-Elmer* mit einer Diamant-ATR- (*attenuated total reflection*) Einheit verwendet. Feststoffe wurden dabei entweder direkt auf die ATR-Einheit gegeben und mit einem Stempel auf der ATR-Einheit verdichtet oder zuvor in wenig DCM gelöst, ein Tropfen der Lösung auf die ATR-Einheit gegeben und das Lösemittel verdampft. Öle bzw. flüssige Stoffe wurden als Filme aufgetragen. Die Wellenzahlen sind in cm^{-1} angegeben und die Intensitäten der erhaltenen Signale mit vs (sehr stark), s (stark), m (mittel), w (schwach) oder b (breit) beschrieben.

UV/Vis-Spektroskopie

UV/Vis-Spektren wurden an einem *JASCO V650* aufgenommen. Dabei wurden 100 Scans/min bei einem Wellenlängenintervall von 0.5 nm verwendet. Die für die zu analysierenden Verbindungen verwendeten Lösemittel wurden dabei als Blindproben verwendet.

Fluoreszenzspektroskopie

Zur Messung der Fluoreszenzaktivität von Verbindungen wurde ein *Cary Eclipse* der Firma *Varian* verwendet.

6 Abkürzungsverzeichnis

Aid	<i>Activation induced cytidine deaminase</i>
Ape1	<i>Apurinic/apyrimidinic endonuclease 1</i>
Apobec	<i>Apolipoprotein B mRNA-editing enzyme catalytic polypeptide</i>
Äq	Äquivalente
ARP	<i>Aldehyde reactive probe</i>
ATR	<i>Attenuated total reflection</i>
au	Willkürliche Einheit bzw. <i>arbitrary unit</i>
BER	Basenexzisionsreparatur bzw. <i>base excision repair</i>
cadC	5-Carboxy-2'-desoxycytidin
CAN	Cer(IV)-ammoniumnitrat
CpG	dC-dG Dinukleotid
dA	2'-Desoxyadenosin
dC	2'-Desoxycytidin
dG	2'-Desoxyguanosin
DCM	Dichlormethan
DIPEA	Ethyl-di(<i>isopropyl</i>)amin
DMAP	<i>N,N</i> -Dimethylaminopyridin
DMF	<i>N,N</i> -Dimethylformamid
DMSO	Dimethylsulfoxid
DNA	Desoxyribonukleinsäure bzw. <i>desoxyribonucleic acid</i>
Dnmt	DNA Methyltransferase
dT	2'-Desoxythymidin
EI	Elektronstoß-Ionisation

6 Abkürzungsverzeichnis

ESI	Elektronenspray-Ionisation
EtOAc	Ethylacetat
EtOH	Ethanol
fdC	5-Formyl-2'-desoxycytidin
gef.	gefunden
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure
hmdC	5-Hydroxymethyl-2'-desoxycytidin
HPLC	<i>High performance liquid chromatography</i>
HRMS	<i>High resolution mass spectrometry</i>
IDCase	Isoorotat-Decarboxylase
IR	Infrarotspektroskopie
KO	<i>Knockout</i>
LC	<i>Liquid chromatography</i>
Mbd	<i>Methyl-CpG binding domain</i>
mdC	5-Methyl-2'-desoxycytidin
MeCN	Acetonitril
MeOH	Methanol
mESC	<i>Murine embryonic stem cell</i>
MS	Massenspektrometrie
NaOAc	Natriumacetat
NBHA	<i>O</i> -4-Nitrobenzylhydroxylamin
NCP	<i>Nucleosome core particle</i>
Neil	<i>Nei endonuclease VIII-like</i>
NEt₃	Triethylamin

6 Abkürzungsverzeichnis

NMR	<i>Nuclear magnetic resonance</i>
ODN	Oligodesoxynukleotid
ppm	<i>parts per million</i>
quant.	Quantitativer Umsatz
ROS	<i>Reactive oxygen species</i>
RNA	Ribonukleinsäure bzw. <i>ribonucleic acid</i>
rpm	<i>rounds per minute</i>
SAM	<i>S</i> -Adenosylmethionin
Smug	<i>Single strand selective monofunctional uracil-DNA glycosylase</i>
TBS	<i>tert</i> -Butyldimethylsilyl
Tdg	Thymin-DNA Glykosylase
Tet	<i>Ten-eleven translocation</i>
TFA	Trifluoressigsäure
TMS	Trimethylsilyl
THF	Tetrahydrofuran
UHPLC	<i>Ultra high performance liquid chromatography</i>
UV/Vis	Ultraviolettes und sichtbares Licht

Literatur

- [1] M. Cohn, *Protein Sci.* **1995**, *4*, 2444–2447.
- [2] S. Ruben, M. D. Kamen, *Phys. Rev.* **1941**, *59*, 349–354.
- [3] D. Rittenberg, R. Schoenheimer, *J. Biol. Chem.* **1937**, *121*, 235–253.
- [4] S. M. Houten, R. J. A. Wanders, *J. Inherit. Metab. Dis.* **2010**, *33*, 469–477.
- [5] P. Singh, R. Saxena, G. Srinivas, G. Pande, A. Chattopadhyay, *PLoS One* **2013**, *8*, e58833.
- [6] H. C. Urey, L. J. Greiff, *J. Am. Chem. Soc.* **1935**, *57*, 321–327.
- [7] G. N. Lewis, R. E. Cornish, *J. Am. Chem. Soc.* **1933**, *55*, 2616–2617.
- [8] M. H. Wahl, H. C. Urey, *J. Chem. Phys.* **1935**, *3*, 411–414.
- [9] H. C. Urey, G. B. Pegram, J. R. Huffman, *J. Chem. Phys.* **1936**, *4*, 623–623.
- [10] J. R. Huffman, H. C. Urey, *J. Ind. Eng. Chem.* **1937**, *29*, 531–535.
- [11] H. C. Urey, J. R. Huffman, H. G. Thode, M. Fox, *J. Chem. Phys.* **1937**, *5*, 856–868.
- [12] H. G. Thode, J. E. Gorham, H. C. Urey, *J. Chem. Phys.* **1938**, *6*, 296–296.
- [13] I. Roberts, H. G. Thode, H. C. Urey, *J. Chem. Phys.* **1939**, *7*, 137–138.
- [14] R. Schoenheimer, S. Ratner, D. Rittenberg, *J. Biol. Chem.* **1939**, *127*, 333–344.
- [15] R. Schoenheimer, D. Rittenberg, A. S. Keston, *J. Biol. Chem.* **1939**, *127*, 385–389.
- [16] R. Schoenheimer, S. Ratner, D. Rittenberg, *J. Biol. Chem.* **1939**, *130*, 703–732.
- [17] O. Beckonert, H. C. Keun, T. M. D. Ebbels, J. Bundy, E. Holmes, J. C. Lindon, J. K. Nicholson, *Nat. Protoc.* **2007**, *2*, 2692–2703.
- [18] I. D. Wilson, J. K. Nicholson, J. Castro-Perez, J. H. Granger, K. A. Johnson, B. W. Smith, R. S. Plumb, *J. Proteome Res.* **2005**, *4*, 591–598.
- [19] O. Fiehn, *Plant Mol. Biol.* **2002**, *48*, 155–171.
- [20] C. Jang, L. Chen, J. D. Rabinowitz, *Cell* **2018**, *173*, 822–837.
- [21] G. Mendel, *Verhandlungen des Naturforschenden Vereines in Brünn* **1866**, *4*, 3–47.
- [22] F. Miescher, *Medizinal-Chemische Untersuchungen* **1871**, *4*, 441–460.
- [23] R. Dahm, *Dev. Biol.* **2005**, *278*, 274–288.
- [24] W. Sutton, *Biol. Bull.* **1902**, *4*, 24–39.
- [25] T. Boveri, *Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns*, G. Fischer, **1904**.

- [26] H. Satzinger, *Nat. Rev. Genet.* **2008**, *9*, 231–238.
- [27] E. W. Crow, J. F. Crow, *Genetics* **2002**, *160*, 1–4.
- [28] O. T. Avery, C. M. MacLeod, M. McCarty, *J. Exp. Med.* **1944**, *79*, 137–158.
- [29] C. Tamm, H. S. Shapiro, R. Lipshitz, E. Chargaff, *J. Biol. Chem.* **1953**, *203*, 673–688.
- [30] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
- [31] M. H. F. Wilkins, A. R. Stokes, H. R. Wilson, *Nature* **1953**, *171*, 738–740.
- [32] R. E. Franklin, R. G. Gosling, *Nature* **1953**, *171*, 740–741.
- [33] R. E. Franklin, R. G. Gosling, *Nature* **1953**, *172*, 156–157.
- [34] J. Watson, F. Crick, *Nature* **1953**, *171*, 964–967.
- [35] M. Meselson, F. W. Stahl, *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 671–682.
- [36] M. W. Nirenberg, J. H. Matthaei, *Proc. Natl. Acad. Sci. U.S.A.* **1961**, *47*, 1588–1602.
- [37] F. H. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, *Nature* **1961**, *192*, 1227–1232.
- [38] F. H. Crick, *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–163.
- [39] F. H. Crick, *Nature* **1970**, *227*, 561–563.
- [40] J. P. Evans, H. M. Florman, *Nat. Cell Biol.* **2002**, *4*, s57–s63.
- [41] R. Jaenisch, R. Young, *Cell* **2008**, *132*, 567–582.
- [42] M. J. Evans, M. H. Kaufman, *Nature* **1981**, *292*, 154–156.
- [43] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, J. M. Jones, *Science* **1998**, *282*, 1145–1147.
- [44] G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 7634–7638.
- [45] A. Bradley, M. Evans, M. H. Kaufman, E. Robertson, *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *309*, 255–256.
- [46] J. B. Gurdon, *J. Embryol. Exp. Morphol.* **1962**, *10*, 622–640.
- [47] K. Takahashi, S. Yamanaka, *Cell* **2006**, *126*, 663–676.
- [48] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, *Cell* **2007**, *131*, 861–872.
- [49] J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, I. I. Slukvin, J. A. Thomson, *Science* **2007**, *318*, 1917–1920.
- [50] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607.

- [51] C. H. Waddington, *The Strategy of the Genes*, Allen & Unwin, **1957**.
- [52] R. Holliday, *Biol. Rev. Cambridge Philos. Soc.* **1990**, *65*, 431–471.
- [53] A. Bird, *Nature* **2007**, *447*, 396–398.
- [54] R. D. Kornberg, *Science* **1974**, *184*, 868–871.
- [55] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251–260.
- [56] T. Schalch, S. Duda, D. F. Sargent, T. J. Richmond, *Nature* **2005**, *436*, 138–141.
- [57] G. Felsenfeld, M. Groudine, *Nature* **2003**, *421*, 448–453.
- [58] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, G. Li, *Science* **2014**, *344*, 376–380.
- [59] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, *Molecular Cell Biology*, W. H. Freeman, **2000**.
- [60] D. Y. Lee, J. J. Hayes, D. Pruss, A. P. Wolffe, *Cell* **1993**, *72*, 73–84.
- [61] Y. Shii, R. N. Eisenman, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13225–13230.
- [62] H. Wang, L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones, Y. Zhang, *Nature* **2004**, *431*, 873–878.
- [63] A. Kuzmichev, T. Jenuwein, P. Tempst, D. Reinberg, *Mol. Cell* **2004**, *14*, 183–193.
- [64] T. Kouzarides, *Cell* **2007**, *128*, 693–705.
- [65] C. A. Musselman, M.-E. Lalonde, J. Cote, T. G. Kutateladze, *Nat. Struct. Mol. Biol.* **2012**, *19*, 1218–1227.
- [66] F. Ling, Y. Tang, M. Li, Q.-S. Li, X. Li, L. Yang, W. Zhao, C.-C. Jin, Z. Zeng, C. Liu, C.-F. Wu, W.-W. Chen, X. Lin, Y.-L. Wang, M. D. Threadgill, *Oncotarget* **2017**, *8*, 72773–72787.
- [67] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41–45.
- [68] T. Jenuwein, C. D. Allis, *Science* **2001**, *293*, 1074–1080.
- [69] 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.
- [70] 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.
- [71] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *ChemBioChem* **2011**, *12*, 206–222.

- [72] S. Saxonov, P. Berg, D. L. Brutlag, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1412–1417.
- [73] R. D. Hotchkiss, *J. Biol. Chem.* **1948**, *175*, 315–332.
- [74] R. Holliday, J. E. Pugh, *Science* **1975**, *187*, 226–232.
- [75] A. D. Riggs, *Cytogenet. Cell Genet.* **1975**, *14*, 9–25.
- [76] P. A. Jones, *Trends Genet.* **1999**, *15*, 34–37.
- [77] M. L. Gonzalgo, T. Hayashida, C. M. Bender, M. M. Pao, Y. C. Tsai, F. A. Gonzales, H. D. Nguyen, T. T. Nguyen, P. A. Jones, *Cancer Res.* **1998**, *58*, 1245–1252.
- [78] P. L. Jones, G. J. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis, A. P. Wolffe, *Nat. Genet.* **1998**, *19*, 187–191.
- [79] X. Nan, H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, A. Bird, *Nature* **1998**, *393*, 386–389.
- [80] K. Hata, M. Okano, H. Lei, E. Li, *Development* **2002**, *129*, 1983–1993.
- [81] U. Aapola, I. Liiv, P. Peterson, *Nucleic Acids Res.* **2002**, *30*, 3602–3608.
- [82] R. Deplus, C. Brenner, W. A. Burgers, P. Putmans, T. Kouzarides, Y. de Launoit, F. Fuks, *Nucleic Acids Res.* **2002**, *30*, 3831–3838.
- [83] L. S. Chuang, H. I. Ian, T. W. Koh, H. H. Ng, G. Xu, B. F. Li, *Science* **1997**, *277*, 1996–2000.
- [84] A. Hermann, R. Goyal, A. Jeltsch, *J. Biol. Chem.* **2004**, *279*, 48350–48359.
- [85] T. Iida, I. Suetake, S. Tajima, H. Morioka, S. Ohta, C. Obuse, T. Tsurimoto, *Genes Cells* **2002**, *7*, 997–1007.
- [86] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247–257.
- [87] A. Jeltsch, *ChemBioChem* **2002**, *3*, 274–293.
- [88] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
- [89] 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.
- [90] G. R. Wyatt, S. S. Cohen, *Nature* **1952**, *170*, 1072–1073.
- [91] G. R. Wyatt, S. S. Cohen, *Biochem. J.* **1953**, *55*, 774–782.
- [92] T. Carell, M. Q. Kurz, M. Müller, M. Rossa, F. Spada, *Angew. Chem. Int. Ed.* **2018**, *57*, 4296–4312.
- [93] N. W. Penn, R. Suwalski, C. O’Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781–790.
- [94] S. Zuo, R. J. Boorstein, G. W. Teebor, *Nucleic Acids Res.* **1995**, *23*, 3239–3243.

- [95] V. Valinluck, H.-H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird, L. C. Sowers, *Nucleic Acids Res.* **2004**, *32*, 4100–4108.
- [96] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, *Cell Cycle* **2009**, *8*, 1698–1710.
- [97] C. Loenarz, C. J. Schofield, *Trends Biochem. Sci.* **2011**, *36*, 7–18.
- [98] P. Borst, R. Sabatini, *Annu. Rev. Microbiol.* **2008**, *62*, 235–251.
- [99] 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.
- [100] 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.
- [101] H. Hashimoto, J. E. Pais, N. Dai, I. R. Correa, X. Zhang, Y. Zheng, X. Cheng, *Nucleic Acids Res.* **2015**, *43*, 10713–10721.
- [102] F. P. Guengerich, *J. Biol. Chem.* **2015**, *290*, 20700–20701.
- [103] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS One* **2010**, *5*, e15367.
- [104] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460–6468.
- [105] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmler, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377.
- [106] S. Dzitoyeva, H. Chen, H. Manev, *Neurobiol. Aging* **2012**, *33*, 2881–2891.
- [107] H. Chen, S. Dzitoyeva, H. Manev, *Restor. Neurol. Neurosci.* **2012**, *30*, 237–245.
- [108] M. Wagner, J. Steinbacher, T. F. J. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511–12514.
- [109] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 1049–1055.
- [110] S. M. Kinney, H. G. Chin, R. Vaisvila, J. Bitinaite, Y. Zheng, P.-O. Esteve, S. Feng, H. Stroud, S. E. Jacobsen, S. Pradhan, *J. Biol. Chem.* **2011**, *286*, 24685–24693.
- [111] S. R. Kornberg, S. B. Zimmerman, A. Kornberg, *J. Biol. Chem.* **1961**, *236*, 1487–1493.
- [112] J. Terragni, J. Bitinaite, Y. Zheng, S. Pradhan, *Biochemistry* **2012**, *51*, 1009–1019.
- [113] B. Khulan, R. F. Thompson, K. Ye, M. J. Fazzari, M. Suzuki, E. Stasiak, M. E. Figueroa, J. L. Glass, Q. Chen, C. Montagna, E. Hatchwell, R. R. Selzer, T. A. Richmond, R. D. Green, A. Melnick, J. M. Greally, *Genome Res.* **2006**, *16*, 1046–1055.

- [114] A. L. Brunner, D. S. Johnson, S. W. Kim, A. Valouev, T. E. Reddy, N. F. Neff, E. Anton, C. Medina, L. Nguyen, E. Chiao, C. B. Oyolu, G. P. Schroth, D. M. Absher, J. C. Baker, R. M. Myers, *Genome Res.* **2009**, *19*, 1044–1056.
- [115] 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.
- [116] S. Kriaucionis, M. Tahiliani, *Cold Spring Harbor Perspect. Biol.* **2014**, *6*, a018630–a018630.
- [117] L. Shen, Y. Zhang, *Curr. Opin. Cell Biol.* **2013**, *25*, 289–296.
- [118] K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. C. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343–348.
- [119] H. Zhang, X. Zhang, E. Clark, M. Mulcahey, S. Huang, Y. G. Shi, *Cell Res.* **2010**, *20*, 1390–1393.
- [120] N. Kitsera, J. Allgayer, E. Parsa, N. Geier, M. Rossa, T. Carell, A. Khobta, *Nucleic Acids Res.* **2017**, *45*, 11033–11042.
- [121] H. Cao, Y. Wang, *Nucleic Acids Res.* **2007**, *35*, 4833–4844.
- [122] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeyer, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.
- [123] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
- [124] 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.
- [125] 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**, *41*, 6421–6429.
- [126] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555–557.
- [127] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angew. Chem. Int. Ed.* **2016**, *55*, 11797–11800.
- [128] M. W. Kellinger, C.-X. Song, J. Chong, X.-Y. Lu, C. He, D. Wang, *Nat. Struct. Mol. Biol.* **2012**, *19*, 831–833.
- [129] M. de la Mata, C. R. Alonso, S. Kadener, J. P. Fededa, M. Blaustein, F. Pelisch, P. Cramer, D. Bentley, A. R. Kornblihtt, *Mol. Cell* **2003**, *12*, 525–532.
- [130] S. Shukla, E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, S. Oberdoerffer, *Nature* **2011**, *479*, 74–79.

- [131] L. J. Core, J. T. Lis, *Science* **2008**, *319*, 1791–1792.
- [132] 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.
- [133] M. Iurlaro, G. Ficiz, D. Oxley, E.-A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian, W. Reik, *Genome Biol.* **2013**, *14*, R119.
- [134] C.-X. Song, K. 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.
- [135] E.-A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, *Nat. Struct. Mol. Biol.* **2015**, *22*, 44–49.
- [136] J. S. Hardwick, A. N. Lane, T. Brown, *BioEssays* **2018**, *40*, 1700199.
- [137] J. S. Hardwick, D. Ptchelkine, A. H. El-Sagheer, I. Tear, D. Singleton, S. E. V. Phillips, A. N. Lane, T. Brown, *Nat. Struct. Mol. Biol.* **2017**, *24*, 544–552.
- [138] T. T. M. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, T. Ha, *Nat. Commun.* **2016**, *7*, 10813.
- [139] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501–502.
- [140] 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.
- [141] A. L. Jacobs, P. Schär, *Chromosoma* **2012**, *121*, 1–20.
- [142] T. Lindahl, B. Nyberg, *Biochemistry* **1972**, *11*, 3610–3618.
- [143] T. Lindahl, *Nature* **1993**, *362*, 709–715.
- [144] T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 15330–15333.
- [145] T. Lindahl, *Angew. Chem. Int. Ed.* **2016**, *55*, 8528–8534.
- [146] A. C. Drohat, A. Maiti, *Org. Biomol. Chem.* **2014**, *12*, 8367–8378.
- [147] P. Fortini, E. Dogliotti, *DNA Repair* **2007**, *6*, 398–409.
- [148] A. J. Podlutzky, I. I. Dianova, S. H. Wilson, V. A. Bohr, G. L. Dianov, *Biochemistry* **2001**, *40*, 809–813.
- [149] K. L. Limpose, K. S. Trego, Z. Li, S. W. Leung, A. H. Sarker, J. A. Shah, S. S. Ramalingam, E. M. Werner, W. S. Dynan, P. K. Cooper, A. H. Corbett, P. W. Doetsch, *Nucleic Acids Res.* **2018**, *46*, 4515–4532.

- [150] J. P. Jost, M. Siegmann, L. Sun, R. Leung, *J. Biol. Chem.* **1995**, *270*, 9734–9739.
- [151] B. Zhu, Y. Zheng, D. Hess, H. Angliker, S. Schwarz, M. Siegmann, S. Thiry, J.-P. Jost, *Nucleic Acids Res.* **2000**, *28*, 5135–5139.
- [152] D. Cortazar, C. Kunz, Y. Saito, R. Steinacher, P. Schär, *DNA Repair* **2007**, *6*, 489–504.
- [153] H. D. Morgan, W. Dean, H. A. Coker, W. Reik, S. K. Petersen-Mahrt, *J. Biol. Chem.* **2004**, *279*, 52353–52360.
- [154] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101–1105.
- [155] N. Bhutani, J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel, H. M. Blau, *Nature* **2010**, *463*, 1042–1047.
- [156] N. Bhutani, M. N. Decker, J. J. Brady, R. T. Bussat, D. M. Burns, S. Y. Corbel, H. M. Blau, *FASEB J.* **2013**, *27*, 1107–1113.
- [157] R. Kumar, L. DiMenna, N. Schrode, T.-C. Liu, P. Franck, S. Munoz-Descalzo, A.-K. Hadjantonakis, A. A. Zarrin, J. Chaudhuri, O. Elemento, T. Evans, *Nature* **2013**, *500*, 89–92.
- [158] 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.
- [159] K. Iqbal, S.-G. Jin, G. P. Pfeifer, P. E. Szabo, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3642–3647.
- [160] 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.
- [161] D. Ji, K. Lin, J. Song, Y. Wang, *Mol. Biosyst.* **2014**, *10*, 1749–1752.
- [162] A. Inoue, Y. Zhang, *Science* **2011**, *334*, 194.
- [163] G. Rangam, K.-M. Schmitz, A. J. A. Cobb, S. K. Petersen-Mahrt, *PLoS One* **2012**, *7*, e43279.
- [164] 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.
- [165] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. 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.

- [166] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338.
- [167] L. S. Pidugu, J. W. Flowers, C. T. Coey, E. Pozharski, M. M. Greenberg, A. C. Drohat, *Biochemistry* **2016**, *55*, 6205–6208.
- [168] L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H.-L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692–706.
- [169] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, *Cell Rep.* **2015**, 674–683.
- [170] A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kusnierczyk, C. B. Vagbo, D. Schuermann, A. Klungland, P. Schär, *Nat. Commun.* **2016**, *7*, 10806.
- [171] I. Morland, V. Rolseth, L. Luna, T. Rognes, M. Bjoras, E. Seeberg, *Nucleic Acids Res.* **2002**, *30*, 4926–4936.
- [172] T. K. Hazra, Y. W. Kow, Z. Hatahet, B. Imhoff, I. Boldogh, S. K. Mokkalapati, S. Mitra, T. Izumi, *J. Biol. Chem.* **2002**, *277*, 30417–30420.
- [173] L. Schomacher, D. Han, M. U. Musheev, K. Arab, S. Kienhöfer, A. Von Seggern, C. Niehrs, *Nat. Struct. Mol. Biol.* **2016**, *23*, 116.
- [174] J. C. Fromme, G. L. Verdine, Base Excision Repair, **2004**.
- [175] A. Slyvka, K. Mierzejewska, M. Bochtler, *Sci. Rep.* **2017**, *7*, 9001.
- [176] 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*, 447–459.
- [177] S. K. Bhattacharya, S. Ramchandani, N. Cervoni, M. Szyf, *Nature* **1999**, *397*, 579–583.
- [178] H. Cedar, G. L. Verdine, *Nature* **1999**, *397*, 568–569.
- [179] H. H. Ng, Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg, A. Bird, *Nat. Genet.* **1999**, *23*, 58–61.
- [180] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, *J. Biol. Chem.* **2013**, *288*, 9084–9091.
- [181] C.-C. Chen, K.-Y. Wang, C.-K. J. Shen, *J. Biol. Chem.* **2012**, *287*, 33116–33121.
- [182] Z. Liutkeviciute, G. Lukinavicius, V. Masevicius, D. Daujotyte, S. Klimasauskas, *Nat. Chem. Biol.* **2009**, *5*, 400–402.
- [183] J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes, A. A. Axhemi, *Biochim. Biophys. Acta* **2005**, *1723*, 256–264.
- [184] C. K. Liu, C. A. Hsu, M. T. Abbott, *Arch. Biochem. Biophys.* **1973**, *159*, 180–187.
- [185] S. Xu, W. Li, J. Zhu, R. Wang, Z. Li, G.-L. Xu, J. Ding, *Cell Res.* **2013**, *23*, 1296–1309.

- [186] R. J. Klose, K. Yamane, Y. Bae, D. Zhang, H. Erdjument-Bromage, P. Tempst, J. Wong, Y. Zhang, *Nature* **2006**, *442*, 312–316.
- [187] S. M. Kooistra, K. Helin, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 297–311.
- [188] K. M. Arcipowski, C. A. Martinez, P. Ntziachristos, *Curr. Opin. Genet. Dev.* **2016**, *36*, 59–67.
- [189] S. Markolovic, T. M. Leissing, R. Chowdhury, S. E. Wilkins, X. Lu, C. J. Schofield, *Curr. Opin. Struct. Biol.* **2016**, *41*, 62–72.
- [190] S. C. Trewick, T. F. Henshaw, R. P. Hausinger, T. Lindahl, B. Sedgwick, *Nature* **2002**, *419*, 174–178.
- [191] P. A. Aas, M. Otterlei, P. O. Falnes, C. B. Vagbo, F. Skorpen, M. Akbari, O. Sundheim, M. Bjoras, G. Slupphaug, E. Seeberg, H. E. Krokan, *Nature* **2003**, *421*, 859–863.
- [192] L. Shen, C.-X. Song, C. He, Y. Zhang, *Annu. Rev. Biochem.* **2014**, *83*, 585–614.
- [193] T. Y. Hargrove, Z. Wawrzak, J. Liu, W. D. Nes, M. R. Waterman, G. I. Lepesheva, *J. Biol. Chem.* **2011**, *286*, 26838–26848.
- [194] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeyer, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516–6520.
- [195] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599.
- [196] Z. Liutkeviciute, E. Kriukiene, J. Licyte, M. Rudyte, G. Urbanaviciute, S. Klimasauskas, *J. Am. Chem. Soc.* **2014**, *136*, 5884–5887.
- [197] K. Iwan, R. Rahimoff, A. Kirchner, F. Spada, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell, *Nat. Chem. Biol.* **2018**, *14*, 72.
- [198] K. A. P. Payne, M. D. White, K. Fisher, B. Khara, S. S. Bailey, D. Parker, N. J. W. Rattray, D. K. Trivedi, R. Goodacre, R. Beveridge, P. Barran, S. E. J. Rigby, N. S. Scrutton, S. Hay, D. Leys, *Nature* **2015**, *522*, 497–501.
- [199] J. Thiem, H. Mohn, A. Heesing, *Synthesis* **1985**, 775–778.
- [200] A. S. Schröder, Diss., Ludwig-Maximilians-Universität München, **2016**.
- [201] R. K. Hill, N. D. Ledford, L. A. Renbaum, *J. Labelled Compd. Radiopharm.* **1985**, *22*, 143–148.
- [202] J. Atzrodt, V. Derdau, T. Fey, J. Zimmermann, *Angew. Chem. Int. Ed.* **2007**, *46*, 7744–7765.
- [203] M. Maeda, Y. Kawazoe, *Tetrahedron Lett.* **1975**, *16*, 1643–1646.
- [204] H. Sajiki, H. Esaki, F. Aoki, T. Maegawa, K. Hirota, *Synlett* **2005**, 1385–1388.

- [205] M. Münzel, Diss., Ludwig-Maximilians-Universität München, **2011**.
- [206] C. H. Hwang, J. S. Park, J. H. Won, J. N. Kim, E. K. Ryu, *Arch. Pharmacol Res.* **1992**, *15*, 69–72.
- [207] A. Ramanathan, L. S. Jimenez, *Synthesis* **2010**, 217–220.
- [208] V. Busskamp, N. E. Lewis, P. Guye, A. H. Ng, S. L. Shipman, S. M. Byrne, N. E. Sanjana, J. Murn, Y. Li, S. Li, M. Stadler, R. Weiss, M. G. Church, *Mol. Syst. Biol.* **2014**, *10*, 760.
- [209] J. Steinbacher, Diss., Ludwig-Maximilians-Universität München, **2016**.
- [210] A. Schön, *Masterarbeit*, **2016**.
- [211] J. Baier, T. Maisch, M. Maier, E. Engel, M. Landthaler, W. Bäuml, *Biophys. J.* **2006**, *91*, 1452–1459.
- [212] L. Fournier, I. Aujard, T. Le Saux, S. Maurin, S. Beaupierre, J.-B. Baudin, L. Jullien, *Chem. Eur. J.* **2013**, *19*, 17494–17507.
- [213] L. Fournier, C. Gauron, L. Xu, I. Aujard, T. L. Saux, N. Gagey-Eilstein, S. Maurin, S. Dubruille, J.-B. Baudin, D. Bensimon, M. Volovitch, S. Vrız, L. Jullien, *ACS Chem. Biol.* **2013**, *8*, 1528–1536.
- [214] B. Cürten, P. H. M. Kullmann, M. E. Bier, K. Kandler, B. F. Schmidt, *Photochem. Photobiol.* **2005**, *81*, 641.
- [215] R. Rahimoff, O. Kosmatchev, A. Kirchner, T. Pfaffeneder, F. Spada, V. Brantl, M. Müller, T. Carell, *J. Am. Chem. Soc.* **2017**, *139*, 10359–10364.
- [216] A. P. Gorka, R. R. Nani, J. Zhu, S. Mackem, M. J. Schnermann, *J. Am. Chem. Soc.* **2014**, *136*, 14153–14159.
- [217] R. R. Nani, A. P. Gorka, T. Nagaya, H. Kobayashi, M. J. Schnermann, *Angew. Chem. Int. Ed.* **2015**, *54*, 13635–13638.
- [218] J. Gu, U. R. Anumala, F. Lo Monte, T. Kramer, R. Heyny von Haußen, J. Hölzer, V. Goetschy-Meyer, G. Mall, I. Hilger, C. Czech, B. Schmidt, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 7667–7671.
- [219] S. C. Alexander, A. Schepartz, *Org. Lett.* **2014**, *16*, 3824–3827.
- [220] A. R. Tyler, A. O. Okoh, C. L. Lawrence, V. C. Jones, C. Moffatt, R. B. Smith, *Eur. J. Med. Chem.* **2013**, *64*, 222–227.
- [221] O. A. Okoh, R. H. Bisby, C. L. Lawrence, C. E. Rolph, R. B. Smith, *J. Sulfur Chem.* **2014**, *35*, 42–56.

„I Jumped in the river, what did I see?“

Radiohead, Pyramid Song (2001)