

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

**Massenspektrometrische Analyse der chemischen
Prozesse an nicht-kanonischen RNA- und DNA-
Nukleosiden**

Katharina Iwan
aus
Starnberg, Deutschland

2018

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 28.10.2018

Katharina Iwan

Dissertation eingereicht am: 08.11.2018

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachter: Dr. Stefanie Kellner

Mündliche Prüfung am: 07.12.2018

Teile dieser Arbeit wurden in folgenden Journalen oder auf Fachkonferenzen publiziert.

Publikationen während der Doktorarbeit

F. R. Traube*, S. Schiffers*, K. Iwan*, S. Kellner, F. Spada, M. Müller and T. Carell; Isotope-dilution mass spectrometry for exact quantification of non-canonical DNA nucleosides. *Nat. Protoc.* **2019**, 14, 283-312 DOI: 10.1038/s41596-018-0094-6.

T. T. Ensfelder*, M. Q. Kurz*, K. Iwan*, S. Geiger, S. Matheisl, M. Müller, R. Beckmann, T. Carell; ALKBH5-Induced Demethylation of Mono- and Dimethylated Adenosine. *Chem. Commun.* **2018**, 54, 8591-8593 DOI: 10.1039/C8CC03980A.

K. Iwan*, R. Rahimoff*, A. Kirchner*, F. Spada*, A. S. Schröder, O. Kosmatchev, S. Ferizaj, J. Steinbacher, E. Parsa, M. Müller, T. Carell; 5-Formylcytosine to cytosine conversion by C-C bond cleavage *in vivo*, *Nat. Chem. Biol.* **2018**, 14, 72–78 DOI: 10.1038/nchembio.2531.

A. S. Schröder, E. Parsa, K. Iwan, F. R. Traube, M. Wallner, S. Serdjukow, T. Carell; 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes, *Chem. Commun.* **2016**, 52, 14361-14364 DOI: 10.1039/C6CC07517G.

A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui, T. Carell; Synthesis of (R)-Configured 2'-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides, *Org. Lett.* **2016**, 18 (17), 4368-4371 DOI:10.1021/acs.orglett.6b02110.

* Beitrag der Autoren zu gleichen Teilen.

Weitere Publikationen

K. H. DuBay, K. Iwan, L. Osorio-Planes, P. L. Geissler, M. Groll, D. Trauner, J. Broichhagen; A Predictive Approach for the Optical Control of Carbonic Anhydrase II Activity, *ACS Chem. Biol.* **2018**, 13 (3), 793-800 DOI: 10.1021/acscchembio.7b00862.

J. Broichhagen, I. Jurastow, K. Iwan, W. Kummerer, D. Trauner; Optical Control of Acetylcholinesterase with a Tacrine Switch, *Angew. Chem. Int. Ed.* **2014**, 53 (29), 7657-7660 DOI: 10.1002/anie.201403666.

Konferenzbeiträge

SFB 1309 Meeting, München (2017) – Posterpräsentation “*Investigation of enzymes involved in the C-C bond cleavage of 5-formyldeoxycytidine*”.

Frontiers in Epigenetic Chemical Biology, London, Royal Society (2017) – Posterpräsentation: “*Mass spectrometric investigation of formylcytidine deformylation activity in living cells*”.

XXII IRT 2016 Conference, Paris, Institut Pasteur (2016) – Posterpräsentation “*Quantification of Abasic Sites and β -Elimination Products Using Tandem Mass Spectrometry*”.

Danksagung

An dieser Stelle möchte ich mich bei zahlreichen Personen für ihre Unterstützung während meiner Promotion bedanken.

Mein erster Dank geht an meinen Doktorvater *Prof. Dr. Thomas Carell* für die spannende Themenauswahl, die Unterstützung und die hervorragenden Arbeitsbedingungen.

Dr. Stefanie Kellner danke ich für die Übernahme des Koreferats.

Auch möchte ich mich bei den Mitgliedern der Prüfungskommission herzlich bedanken.

Frau *Slava Gärtner* danke ich für ihre Hilfe bei allen organisatorischen Belangen. Auch möchte ich mich ganz herzlich bei Frau *Kerstin Kurz* bedanken, die mir stets eine sehr große Hilfe bei der Bewältigung riesiger Probenmengen für die massenspektrometrische Analyse war.

Herrn *Markus Müller* möchte ich für die Hilfe bei biochemischen Fragestellungen und Experimenten danken, wie auch der stetigen wissenschaftlichen Diskussion die zum besseren Verständnis der neu gewonnenen Daten geführt haben. Und natürlich für die Hilfe bei allen technischen Geräten – ohne ihn hätte ich noch immer keinen Zugang zum VPN Client.

Frau *Stefanie Kellner* möchte ich herzlichst für ihre Hilfe und Beratung in allen wissenschaftlichen Belangen danken.

Mein besonderer Dank gilt meinen Laborkollegen mit denen ich an den verschiedensten Themen gearbeitet habe. Ich danke der kompletten Triple Quad-Gang (*Jessica Steinbacher, Olesea Kosmatchev, Sarah Schiffers und Kerstin Kurz*). Auch wenn wir an unterschiedlichen Themen gearbeitet haben, so hat uns das Triple Quad zusammengeführt und zu vielen spannenden Diskussionen angeregt. Bedanken möchte ich mich bei dem kompletten C-C bond cleavage Team (*Arne Schröder, Edris Parsa, Olga Kamerer, Jessica Steinbacher, Eva Korytiakovà, Fabio Spada, René Rahimoff und Angie Kirchner*). Besonders hervorheben möchte ich die Zusammenarbeit mit *René Rahimoff* und *Angie Kirchner*. Ich glaube die Veröffentlichung unseres Manuskripts hat uns alle viel Geduld und Durchhaltevermögen gelehrt und trotzdem war es eine gute Zeit mit euch!

Des Weiteren möchte ich den Mitgliedern des RNA-Teams danken (*Simon Geiger, Jessica Steinbacher, Angie Kirchner, Sarah Schiffers, Olga Kamerer, Martin Rossa und Matthias Kurz*). Leider haben wir dieses Projekt nicht bis ins Ziel gebracht, doch wir haben alle viel

dazugelernt. Bis zuletzt war es euer unerschütterlicher Wille und Glaube an das Projekt, die mich motiviert haben.

Angie Kirchner, wir hatten das große Glück quasi an allen Projekten im Team zu arbeiten. Es war eine tolle Zeit mit dir. *Franziska Traube*, deine Arbeitsweise hat mich unglaublich motiviert und unsere fachlichen Diskussionen haben mich immer einen Schritt weitergebracht. Vielen Dank euch beiden.

Ein großer Dank geht auch an alle Korrekturleser, *Angie Kirchner*, *René Rahimoff*, *Franziska Traube*, *Martin Rossa* und *Markus Müller*.

Auch meinen Laborkollegen möchte ich danken. Durch euch sind die 4 Jahre Promotion wie im Flug vergangen. Mit interessanten Diskussionen und guter Musik hat mir das Arbeiten mit euch gut gefallen.

Mara Gnädig, *David Schmidl*, *Peyton Randolph* und *Sarah Schiffers*, vielen Dank, dass ich eure Praktika bzw. Masterarbeit betreuen durfte. Ihr habt tolle Arbeit geleistet und ich konnte viel von euch lernen.

Die „wichtigste Frage des Tages“ habe ich immer mit euch geklärt, *Charlotte Ebert*, *Angie Kirchner* und *Franziska Traube*. Diese kleine Auszeit vom Laborstress mit euch habe ich immer sehr genossen.

Danken möchte ich besonders meinen Freunden; ihr habt mich immer unterstützt und mir mehr Selbstvertrauen geschenkt. Nie hätte ich gedacht, dass man zu seinen besten Freunden aus der Schulzeit so viele wunderbare Freundschaften während der letzten 10 Uni-Jahre dazugewinnen kann. Vielen Dank euch allen!

Mein größter Dank gilt meiner Familie. Ihr habt mich immer unterstützt, wo ihr konntet, sowohl finanziell als auch emotional. Dass ihr alle so interessiert und stolz seid, bedeutet mir alles. Zusammen mit meiner „quasi-Familie“, meinen zwei Mitbewohnerinnen *Christine Hösl* und *Gundula Hölzel*, habt ihr für mein Wohlfühlklima gesorgt und mich diese besondere Zeit der Promotion meistern lassen.

Inhaltsverzeichnis

Zusammenfassung	I
Summary	VI
1 Einleitung	1
1.1 Grundlagen	1
1.2 Modifikationen der Ribonukleinsäuren	5
1.2.1 Hypermodifikationen: i ⁶ A und ms ² i ⁶ A	8
1.2.1.1 Einfluss auf die Translation	9
1.2.1.2 Bedeutung in Krankheitsbildern	11
1.2.2 Modifikationen in mRNA	13
1.2.2.1 Inosin	13
1.2.2.2 N ⁶ -Methyladenosin - ein potentieller epigenetischer Marker in mRNA	14
2 Aufgabenstellung	22
3 Veröffentlichte Arbeiten	23
3.1 Isotope-dilution mass spectrometry for exact quantification of non-canonical DNA nucleosides	23
3.2 Synthesis of (<i>R</i>)-Configured 2'-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides	54
3.3 2'-(<i>R</i>)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes	60
3.4 5-Formylcytosine to cytosine conversion by C–C bond cleavage <i>in vivo</i>	65
3.5 ALKBH5-Induced Demethylation of Mono- and Dimethylated Adenosine	75
4 Unveröffentlichte Arbeiten	79
4.1 Studie zur biologischen Funktion der Hypermodifikation ms ² i ⁶ A	79
4.1.1 Prolog	79
4.1.2 Ergebnis und Diskussion	81
4.1.3 Zusammenfassung der Projektbeiträge	91
4.1.4 Material und Methoden	92
4.1.4.1 Zelllinien	92
4.1.4.2 RNA Isolation	93
4.1.4.3 Agarose Gel-Elektrophorese	94
4.1.4.4 Enzymatischer Verdau	95
4.1.4.5 UHPLC-MS/MS Analyse	95
4.1.4.6 Quantitative Echt-Zeit PCR Analyse	98
4.1.4.7 Visualisierung der Mitochondrien durch den Mito Tracker [®] Deep Red	99

4.2	Studie zur aktiven Demethylierung mit verschiedenen dC- und C-Derivaten.....	100
4.2.1	Prolog	100
4.2.2	Ergebnis und Diskussion	103
4.2.3	Zusammenfassung der Projektbeiträge.....	127
4.2.4	Material und Methoden	128
4.2.4.1	Zelllinien	128
4.2.4.2	Enzymatischer Verdau	128
4.2.4.3	UHPLC-MS/MS Analyse	128
4.2.4.4	Auswertung	130
4.3	Studien zur Identität des Enzyms mit Deformylierungsaktivität.....	131
4.3.1	Prolog	131
4.3.2	Ergebnis und Diskussion	133
4.3.3	Zusammenfassung der Projektbeiträge.....	141
4.3.4	Material und Methoden	141
4.3.4.1	Zelllinien	141
4.3.4.2	Enzymatischer Verdau	141
4.3.4.3	UHPLC-MS/MS Analyse	141
4.3.4.4	Auswertung	141
5	Abkürzungsverzeichnis	142
6	Literaturverzeichnis	147

Zusammenfassung

Seit einiger Zeit ist bekannt, dass die Desoxyribonukleinsäure (DNA) nicht nur kanonische Nucleoside enthält, sondern auch eine Vielzahl an modifizierten Nucleobasen. Die Bestimmung der genauen Menge dieser nicht-kanonischen Basen kann zur Charakterisierung des Zustands eines biologischen Systems, beispielsweise zur Bestimmung von Krankheiten oder Entwicklungsprozessen, beitragen und ist daher von großer Bedeutung. Da die meisten dieser Nucleoside nur in sehr kleinen Mengen vorkommen benötigt man zur Vermessung eine präzise Probenvorbereitung sowie extrem sensitive Messmethoden. Die Parameter der massenspektrometrischen Analyse sowie deren Validierung müssen präzise bestimmt und eingehalten werden.

Im Zuge dieser Doktorarbeit wurde ein Protokoll ausgearbeitet, in dem zunächst die Extraktion von genomischer DNA über eine optimierte *Spin Column* Methode mit anschließendem enzymatischen Verdau der DNA durchgeführt wird. In der Arbeit wird die Verwendung von isotope-markierten Standards erläutert, welche eine exakte Quantifizierung mittels der Isotopenverdünnungstechnik erlaubt. Diese Messmethode liefert sehr präzise und robuste Werte. Die Theorie hinter dieser beschriebenen UHPLC-MS/MS (Triple Quadrupol) Massenspektrometrie-Methode ist in Abbildung 1 schematisch dargestellt. Sie stellt eine methodische Einleitung für alle weiteren Projekte im Verlauf dieser Doktorarbeit dar.

Neben der epigenetischen Modifikation 5-Methyl-2'-desoxycytidin (mdC) wurde bereits 2009 ein weiteres nicht-kanonisches Nucleosid, 5-Hydroxymethyl-2'-desoxycytidin (hmdC), im Genom von Säugetieren entdeckt. Kurze Zeit darauf konnten noch zwei weitere neuartige DNA-Modifikationen gefunden werden. Es handelte sich dabei um die höher oxidierten Spezies 5-Formyl-2'-desoxycytidin (fdC) und 5-Carboxyl-2'-desoxycytidin (cadC). Während mittlerweile bekannt ist, dass mdC and hmdC eine wichtige Rolle in der Genexpression spielen, wurden den beiden höher oxidierten Spezies noch keine genaue Funktion zugeschrieben. Diese Modifikationen können auch als Intermediate eines aktiven Demethylierungsweges angesehen werden. Die Erforschung der Funktionen dieser neuen epigenetischen Basen ist eine große Herausforderung, vor allem da die oxidierten Derivate, fdC und cadC durch die sogenannte Basenexzisionsreparatur (BER) ausgeschnitten werden können. Beide befinden sich daher in nur sehr geringen Mengen in der DNA. In der Arbeit werden Nucleosid-Analoga entwickelt, die eine präzise Vermessung unabhängig von BER Prozessen zulassen. Eine 2'-F Substitution der Nucleoside verhindert die Aktivität der Glykosylasen, die für die BER verantwortlich sind. Der Vergleich eines F-fdC-Oligonucleotids mit einem fdC-haltigen Oligonucleotid in einer Aktivitätsstudie mit der humanen Thymin DNA

Glykosylasen (hTDG) zeigte, dass der fdC-haltige Doppelstrang gespalten wurde, während der F-fdC-haltige Doppelstrang intakt blieb.

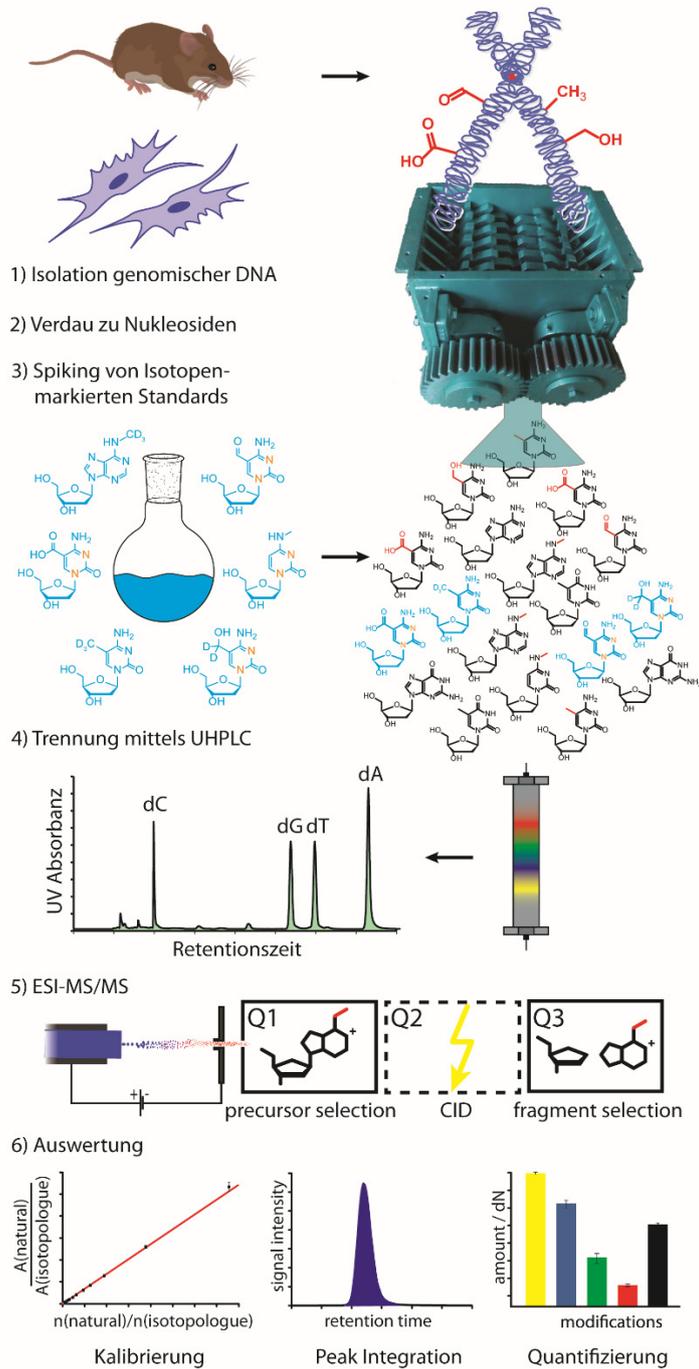


Abbildung 1: Schematische Darstellung des Ablaufs der DNA Isolation, des DNA Verdau, der anschließenden Messung und Auswertung der DNA Proben mittels der Isotopenverdünnungsmethode.

Der 2'-F Substituent schützt das Nucleosid-Analogon demnach vor der BER. Die Methylierung des F-dC mit der Methyltransferase M.Sss1 zeigte, dass die 2'-F Substitutionen keinen Effekt

auf die Methylierung hat. Somit sind 2'-F Nukleosid Analoga das ideale Werkzeug, um den epigenetischen Metabolismus ohne den Einfluss der BER zu untersuchen.

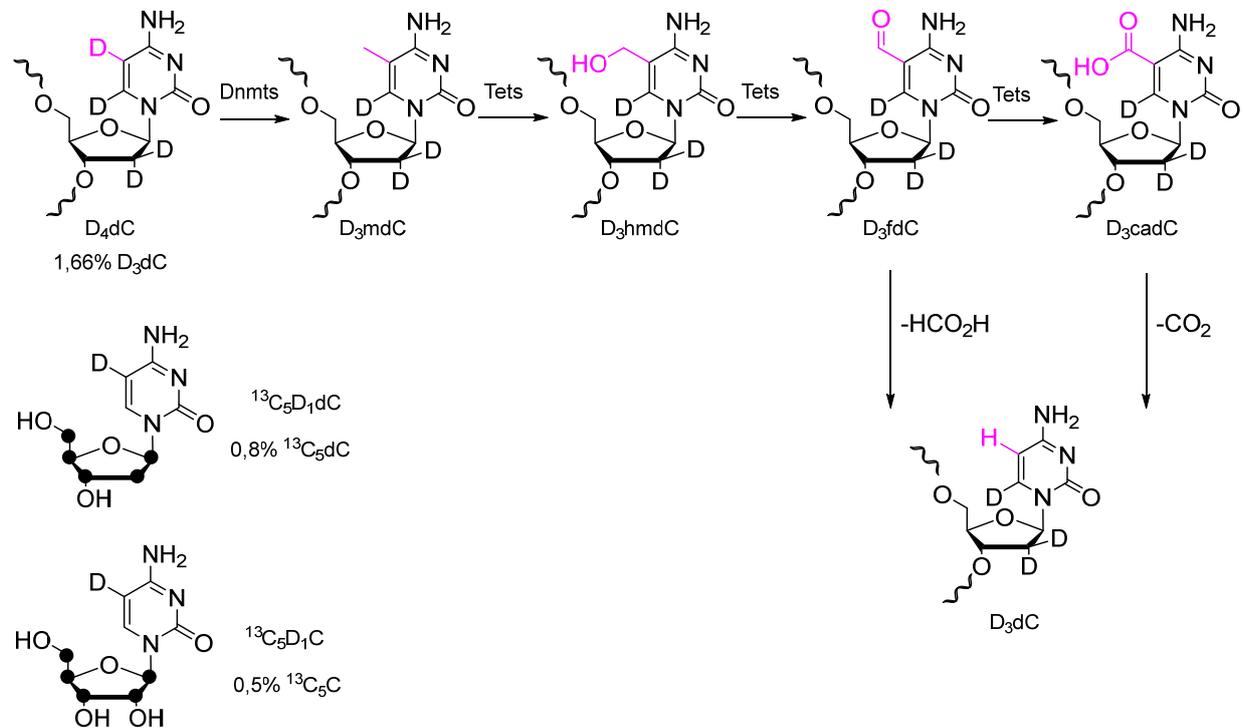
Im weiteren Verlauf wurde überprüft, ob die 2'-F Nukleosid Analoga auch als Substrate für DNA Polymerasen und *Ten-eleven Translocation* (TET)-Enzyme fungieren können. Hierfür wurden zunächst die 2'-(*R*)-fluorierten mdC, hmdC, fdC und cadC Triphosphate synthetisiert. Als mögliches Templat wurde ein Fragment der *octamer-binding transcription factor 4* (*Oct4*)-Promotorsequenz gewählt. Diese Sequenz ist bekannt für eine hohe Dichte an epigenetisch relevanten Nukleosiden, da es viele CpG-Dinukleotide enthält, was sie aber oft zu einem für die Polymerasekettenreaktion (PCR) schwer zugänglichen Templat macht. Mit allen 2'-(*R*)-F-xdCTPs konnte ein Amplifikat der *Oct4*-Promotorsequenz mittels PCR erzielt werden. Allerdings wurde beobachtet, dass die Aktivität der TET-Enzyme durch den F-Substituenten etwas verringert ist. Dennoch eignen sich diese fluorierten Nukleosid-Analoga hervorragend für weitere Studien des epigenetischen Metabolismus.

Im Folgenden wurden diese fluorierten Nukleosid-Analoga eingesetzt, um die BER-unabhängige aktive Demethylierung zu erforschen. In dieser Arbeit wurde gezeigt, dass es möglich ist, isopenmarkiertes fdC, wie auch 2'-(*R*)-F-fdC, in das Genom von Säugetierzellen zu inkorporieren. Der Metabolismus dieser fdC-Derivate in genomischer DNA kann durch UHPLC-MS/MS Messungen nachverfolgt werden. Diese Quantifizierungsexperimente deuten darauf hin, dass eine Deformylierungsreaktion vermutlich, nach der Inkorporation in das Genom, stattfindet. Diese C-C Bindungsbruchreaktion führt zu einer direkten Umwandlung von fdC zu einem unmodifizierten dC ohne die Entstehung einer abasischen Stelle bzw. eines Strangbruches.

Auf der Suche nach dem Enzym mit dieser Deformylase-Aktivität wurden verschiedene Enzymklassen untersucht. Bisher konnte kein Enzym gefunden werden, durch dessen Inhibition die Deformylierungsaktivität auf null gesenkt wurde. Neben einer wahrscheinlich indirekten Wirkung von der DNA Methyltransferase 3b (DNMT3b) auf die Deformylierung schien vor allem die Enzymklasse der Sirtuine einen Einfluss auf die Deformylierung aufzuweisen. Durch die Verfütterung von F-fdC und der gleichzeitigen Inkubation mit Nicotinamid, einem Breitbandinhibitor der Sirtuine, konnte gezeigt werden, dass dadurch die Deformylierungsaktivität um ca. 24 % gesenkt wird. Auch die Zugabe eines spezifischen Sirtuin 2 Inhibitors führte zu einer verminderten Deformylierung. Es stellte sich dabei heraus, dass es sehr schwierig ist, gleichbleibende Effekte durch die Verwendung von Inhibitoren zu erhalten. Die nachfolgende Untersuchung einer Sirtuin 1 Knockout-Zelllinie, im Vergleich zum korrespondierenden Wildtyp, zeigte eine konstante Abnahme der Deformylierungsreaktion um ca. 60 %. Die Enzymklasse der Sirtuine scheint die Deformylierungsaktivität folglich zu beeinflussen. Da die Inhibition der Sirtuine aber zu keiner kompletten Hemmung der

Deformylierungsreaktion führte, ist anzunehmen, dass es weitere Prozesse gibt die zu einer Deformylierung führen können.

Da die bisherigen Studien mit den 2'-fluorierten Nucleosiden vorgenommen wurden und diese nicht natürlich vorkommen, wurden in dem folgenden Projekt verschiedene isotoopenmarkierte dC- und C- Derivate verwendet, um die Methylierung, Oxidation und die anschließende aktive Demethylierung in Zellen nachverfolgen zu können (Schema 1).



Schema 1: Der mögliche Weg der aktiven Demethylierung über einen C-C Bindungsbruch kann entweder über die Deformylierung ($-HCO_2H$) oder über die Decarboxylierung ($-CO_2$) verlaufen. Durch die D-Markierung können die Nucleoside nach dem Totalverdau von den natürlichen Nucleosiden unterschieden werden.

Das erste Testmolekül, D_4dC , führte zu einer erschwerten massenspektrometrischen Auswertung der Ergebnisse aufgrund einer Verunreinigung von 1,66 % D_3dC . Die Weiterentwicklung des Testmoleküls zum $^{13}C_5D_1dC$ führte dann zu einer verbesserten Isotopenreinheit mit nur 0,8 % $^{13}C_5dC$ Verunreinigung. Um die Inkorporation und die Isotopenreinheit zu verbessern, wurde das Testmolekül trotzdem noch einmal weiterentwickelt. Das neueste Testmolekül, das Ribonucleosid $^{13}C_5D_1C$, zeigte eine wesentlich bessere Inkorporation, über den Ribonucleinsäure (RNA) *Salvage Pathway*, in die Zellen. Auch die Isotopenreinheit wurde nochmals verbessert und beziffert zum Schluss nur noch 0,5 % $^{13}C_5C$. Die drei beschriebenen Testmoleküle wurden an verschiedene Knockout-Zelllinien und deren korrespondierende Wildtypen verfüttert und anschließend massenspektrometrisch analysiert. Aus den Daten konnten keine weiteren Hinweise auf eine aktive Demethylierungsreaktion in Form einer Deformylierungs- bzw. Decarboxylierungsreaktion erhalten werden. Dennoch gibt es messbare Unterschiede zwischen den jeweiligen

Knockout-Zelllinien und deren korrespondierenden Wildtypen. Vor allem bei der Verwendung von Dnmt Triple Knockout Zellen im Vergleich zum Wildtyp ist deutlich mehr potentiell C-C Bindungsbruchprodukt detektierbar. Das ist allerdings noch kein eindeutiger Beweis, da dieses Produkt auch durch einen D/H-Austausch entstanden sein könnte.

Das zweite große Thema der Doktorarbeit bestand in der Erforschung des RNA Metabolismus und der dafür notwendigen Entwicklung einer Quantifizierungsmethode für verschiedene RNA-Nukleoside.

Die RNA enthält eine große Anzahl von modifizierten Nukleosiden, wie zum Beispiel die methylierten Derivate *N*⁶-Methyladenosin (*m*⁶A) und das dimethylierte Adenosin *m*⁶₂A. Für diese beiden Modifikationen wurde eine UHPLC-MS/MS Quantifizierungsmethode entwickelt. Es sollte überprüft werden, ob auch die Modifikation *m*⁶₂A durch die α -Ketoglutarat-abhängige Oxidase ALKBH5 demethyliert werden kann. Bisher wurde angenommen, dass nur *m*⁶A hauptsächlich durch ALKBH5 demethyliert wird. In dieser Arbeit wurde gezeigt, dass ALKBH5 auch *m*⁶₂A demethylieren kann.

In dem zweiten RNA Projekt wurde die Hypermodifikation *N*⁶-Isopentenyl-2-thiomethyladenosin (*ms*²ⁱ⁶A) und deren Vorkommen in *messenger RNA* (mRNA) erforscht. Für die Modifikation *ms*²ⁱ⁶A und deren biologische Vorläufermodifikation *N*⁶-Isopentenyl-adenosin (*i*⁶A) wurde eine neue sensitivere UHPLC-MS/MS Quantifizierungsmethode entwickelt um auch kleinere Mengen an PolyA-RNA verwenden zu können. Mit Hilfe dieser Methode konnten die Mengen an *ms*²ⁱ⁶A auch nach einer zweifachen PolyA-RNA Anreicherung und darauffolgenden rRNA Abreicherung quantifiziert werden. Da es zu einer Anreicherung von *ms*²ⁱ⁶A im Vergleich zur Gesamt-RNA kam, kann man daraus schließen, dass diese Modifikation in PolyA-RNA vorkommt. Zusätzlich dazu wurde vermutet, dass *ms*²ⁱ⁶A eine Rolle im Splicing spielt. Die Anwendung des Splicing Inhibitors Isoginkgetin und die nachfolgende Quantifizierung des Gehalts an *ms*²ⁱ⁶A und *i*⁶A mittels UHPLC-MS/MS zeigte, dass die Inhibition des Splicings zu einer kleinen, aber signifikanten Erhöhung der Menge an *ms*²ⁱ⁶A und *i*⁶A führte. Auch wenn die Messung der relativen Expression eine Erhöhung der prä-mRNA zeigte, ist es schwierig einen direkten Zusammenhang zwischen der Modifikation *ms*²ⁱ⁶A und Splicing herzustellen. Da die vorangegangenen Experimente noch keine endgültigen Schlüsse zuließen, wurde der Phänotyp von *ms*²ⁱ⁶A Knockout Zellen genauer untersucht, um Hinweise auf die Funktion dieser Hypermodifikation zu erzielen. Nach der Evaluation aller Ergebnisse und der neuesten Forschung auf dem Gebiet ist es allerdings sehr unwahrscheinlich, dass die Modifikation *ms*²ⁱ⁶A in mRNA vorkommt. Trotzdem konnte eine Phänotypisierung dieser *ms*²ⁱ⁶A-defizienten Zelllinie erreicht werden, die neue Forschungsergebnisse liefert und somit zum Verständnis dieser Hypermodifikation beiträgt.

Summary

It has been known for some time that deoxyribonucleic acid (DNA) contains not only canonical nucleosides but also a large number of modified nucleobases. The determination of the exact amount of these non-canonical bases can contribute to the characterization of the condition of a biological system, for example to determine diseases or development processes, and is therefore of great importance. Since most of these nucleosides only occur in very small quantities, precise sample preparation and extremely sensitive measuring methods are required for the measurement. The parameters of mass spectrometric analysis and its validation must be precisely determined and observed. In the course of this doctoral thesis, a protocol was developed in which the extraction of genomic DNA is first performed using an optimized spin column method followed by enzymatic digestion of the DNA. In this thesis the use of isotope-labelled standards is explained, which allows an exact quantification using the isotope dilution technique. This measuring method provides very precise and robust values. The theory behind this described UHPLC-MS/MS (Triple Quadrupole) mass spectrometry method is shown schematically in Figure 1. It provides a methodological introduction for all further projects in the course of this doctoral thesis.

In addition to the epigenetic modification 5-methyl-2'-deoxycytidine (mdC), another non-canonical nucleoside, 5-hydroxymethyl-2'-deoxycytidine (hmdC), was already discovered in the genome of mammals in 2009. Shortly afterwards, two further novel DNA modifications were found. These are the higher oxidized species 5-formyl-2'-deoxycytidine (fdC) and 5-carboxy-2'-deoxycytidine (cadC). While it is now known that mdC and hmdC play an important role in gene expression, no clear function has yet been assigned for the two higher oxidized species. These modifications can also be considered as intermediates of an active demethylation pathway. Research towards the functions of these new epigenetic bases is a major challenge, especially since the oxidized derivatives, fdC and cadC can be cut out by the so-called base excision repair (BER). Both are present in the DNA in very small amounts. In this thesis nucleoside analogues are developed which allow a precise measurement independent of BER processes. A 2'-F substitution of nucleosides prevents the activity of the glycosylases responsible for BER. Comparison of an F-fdC- with an fdC-containing oligonucleotide in a study of human thymine DNA glycosylase (hTDG) activity showed that the fdC-containing double strand was cleaved while the F-fdC-containing double strand remained intact. The 2'-F substituent protects the nucleoside analog from BER. Methylation of F-dC with methyltransferase M.Sss1 showed that the 2'-F substitutions have no effect on methylation. Thus, 2'-F-nucleoside analogues are the ideal tool to study epigenetic metabolism without the influence of BER.

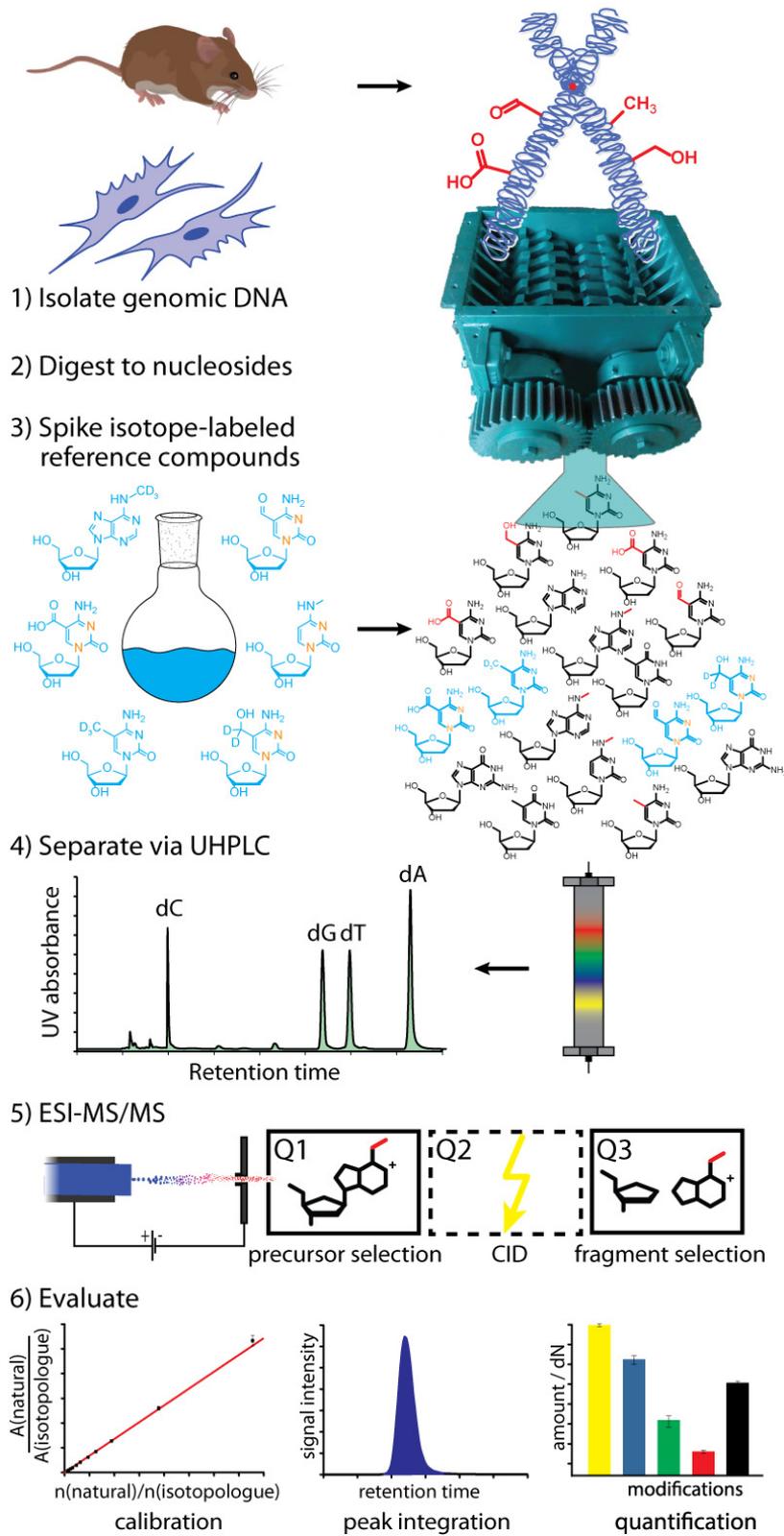


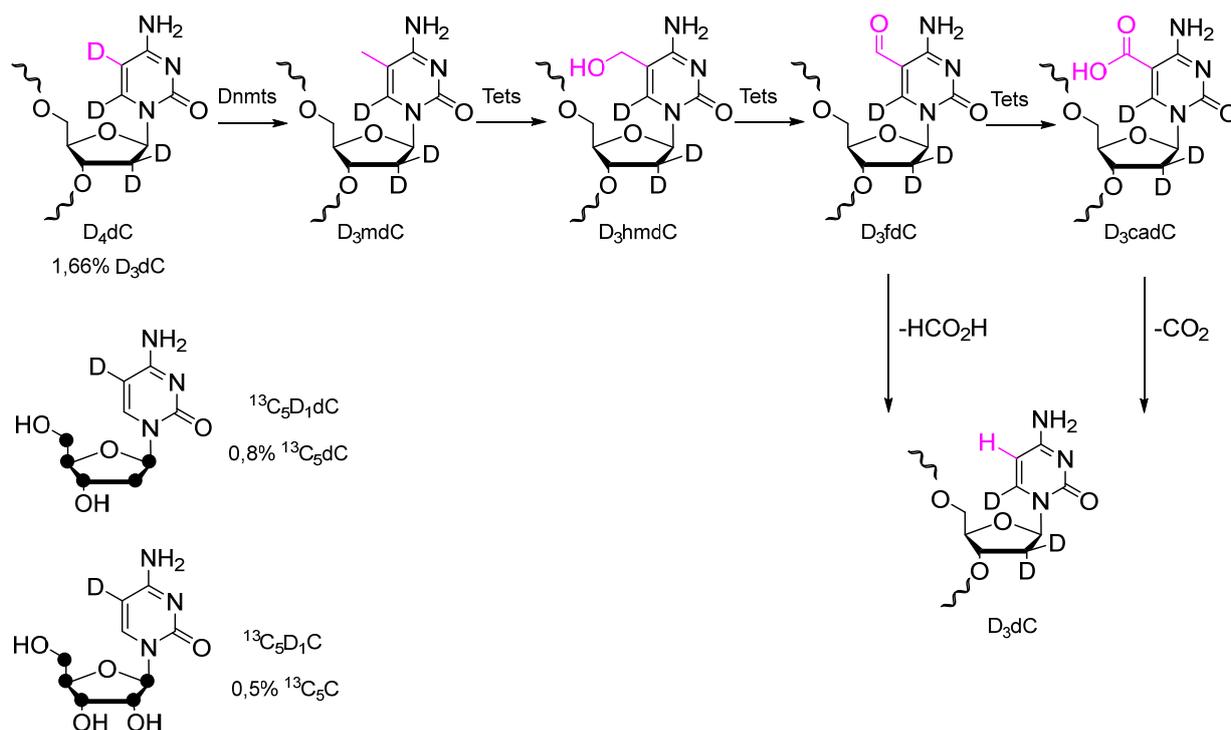
Figure 1: Schematic presentation of the process of DNA isolation, DNA digestion, the subsequent measurement and evaluation of the DNA samples using the isotope dilution method.

In the further course it was examined whether the 2'-F nucleoside analogues are also substrates for DNA polymerases and ten-eleven translocation (TET) enzymes. First, the 2'-(*R*)-fluorinated mdC, hmdC, fdC and cadC triphosphates were synthesized. A fragment of the octamer-binding transcription factor 4 (*Oct4*) promoter sequence was selected as a possible template. This sequence is known for its high density of epigenetically relevant nucleosides, as it contains many CpG dinucleotides, which often makes it an inaccessible template for the polymerase chain reaction (PCR). With all 2'-(*R*)-F-xdCTPs the *Oct4* promoter sequence could be produced via PCR. However, the activity of the TET enzymes is somewhat reduced by the F-substituent, but these fluorinated nucleoside analogs are still excellent for further studies of epigenetic metabolism.

Subsequently, these fluorinated nucleoside analogs were used to investigate BER-independent active demethylation. In this work it was shown that it is possible to incorporate isotope-labelled fdC, as well as 2'-(*R*)-F-fdC, into the genome of mammalian cells. The metabolism of these fdC derivatives in DNA can be traced by UHPLC-MS/MS measurements. These experiments showed quantitatively that, most likely after incorporation into the genome, a deformylation reaction of fdC to dC occurs. This C-C bond cleavage reaction leads to a direct conversion of fdC to an unmodified dC without the formation of an abasic site or a strand break.

In the search for the enzyme with the deformylase activity, different classes of enzymes were investigated. So far, no enzyme has been found whose inhibition led to a decrease of the deformylation activity to zero. In addition to a rather indirect effect of DNA methyltransferase 3b (DNMT3b) on deformylation, the enzyme class of the sirtuins in particular appeared to have an influence on the deformylation. By feeding F-fdC and simultaneous incubation with nicotinamide, a broadband inhibitor of sirtuins, it was shown that this reduces deformylation activity by approx. 24%. The addition of a specific sirtuin 2 inhibitor also led to reduced deformylation. It turned out that it is very difficult to obtain constant effects through the use of inhibitors. The subsequent investigation of a sirtuin 1 knockout cell line, compared to the corresponding wild type, showed a constant decrease in the deformylation reaction by approx. 60%. The enzyme class of sirtuins therefore seems to influence the deformylation activity. However, since the inhibition of sirtuins did not lead to a complete inhibition of the deformylation reaction, it can be assumed that there are further processes that can lead to deformylation.

Since all these studies were carried out with the 2'-fluorinated nucleosides and these do not occur naturally, various isotopically labelled dC and C derivatives were used in the following project to track methylation, oxidation and subsequent active demethylation in cells (Scheme 1).



Scheme 1: The possible pathway of an active demethylation process via a C-C bond cleavage reaction can take place either via deformylation ($-HCO_2H$) or decarboxylation ($-CO_2$). With the D-labelling of the nucleosides they can be distinguished from the natural nucleosides after total digestion.

The first test molecule, D_4dC , led to a more difficult mass spectrometric evaluation of the results due to contamination of 1.66% D_3dC . The further development of the test molecule to $^{13}C_5D_1dC$ led to an improved isotope purity with only 0.8% $^{13}C_5dC$ contamination. In order to improve incorporation and isotope purity, the test molecule was further developed. The latest test molecule, $^{13}C_5D_1C$ ribonucleoside, showed a much better incorporation via the ribonucleic acid (RNA) salvage pathway into the cells. The isotope purity has also been further improved and is now only 0.5% $^{13}C_5C$. The three described test molecules were fed to different knockout cell lines and their corresponding wild types and subsequently analysed by mass spectrometry. No further indications of an active demethylation reaction via a deformylation or decarboxylation reaction could be obtained from the data. Nevertheless, there are measurable differences between the respective knockout cell lines and their corresponding wild types. Especially when using Dnmt triple knockout cells compared to wild type, significantly more potential C-C bond cleavage product can be detected. However, this is not proof, as this product could also have been formed by a D/H exchange.

The second major topic of the doctoral thesis was the research of RNA metabolism and the necessary development of a quantification method for different RNA nucleosides.

RNA contains a large number of modified nucleosides, such as the methylated derivatives N^6 -methyladenosine (m^6A) and the dimethylated adenosine m^6_2A . For these two modifications

a UHPLC-MS/MS quantification method was developed. It should be investigated whether the modification m^6_2A can also be demethylated by the α -ketoglutarate-dependent oxidase ALKBH5. So far it has been assumed that only m^6A is mainly demethylated by ALKBH5. In this work it was shown that ALKBH5 can also demethylate m^6_2A .

In the second RNA project the hypermodification *N*⁶-isopentenyl-2-thiomethyladenosine ($ms^{2i^6}A$) and its occurrence in messenger RNA (mRNA) was investigated. For the modification $ms^{2i^6}A$ and its biological precursor modification *N*⁶-isopentenyl-adenosine (i^6A) a new and more sensitive UHPLC-MS/MS quantification method was developed to be able to use smaller amounts of PolyA-RNA. Applying this method, the amounts of $ms^{2i^6}A$ could also be quantified after two consecutive polyA-RNA enrichments and a subsequent rRNA depletion. Since an enrichment of $ms^{2i^6}A$ compared to total RNA occurred, it can be concluded that this modification occurs in polyA-RNA. In addition, it was assumed that $ms^{2i^6}A$ plays a role in splicing. Application of the splicing inhibitor isoginkgetin and subsequent quantification of the $ms^{2i^6}A$ and i^6A content by UHPLC-MS/MS showed that splicing inhibition led to a small but significant increase in the amount of $ms^{2i^6}A$ and i^6A . Although the measurement of relative expression showed an increase in pre-mRNA, it is difficult to establish a direct relationship between the modification $ms^{2i^6}A$ and splicing. Since the previous experiments did not allow final conclusions to be drawn, the phenotype of $ms^{2i^6}A$ knockout cells was examined more closely in order to obtain indications of the function of this hypermodification. However, after the evaluation of all results and the latest research in the field, it is very unlikely that the modification $ms^{2i^6}A$ occurs in mRNA. Nevertheless, a phenotyping of this $ms^{2i^6}A$ -deficient cell line could be achieved, which provides new research results and thus contributes to the understanding of this hypermodification.

1 Einleitung

1.1 Grundlagen

Im Jahr 2003 kostete die erste Sequenzierung eines gesamten menschlichen Genoms ca. 2,7 Milliarden Dollar. Durch die rasche Entwicklung neuer Sequenzierungsmethoden wird erwartet, dass man demnächst sein eigenes gesamtes Genom für weniger als 100 Dollar sequenzieren lassen kann.

Der Grundstein dafür liegt im letzten Jahrhundert, denn seit damals hat sich die Forschung rund um die DNA von den ersten, frühen Erkenntnissen ihrer chemischen Zusammensetzung und molekularen Struktur stark weiterentwickelt. Im Jahr 1919 konnte die erste korrekte Anordnung der einzelnen Bausteine als kettenartige Struktur in Hefe postuliert werden.^[1] Erst in den darauffolgenden Jahren wurde gezeigt, dass DNA aus paarweise angeordneten Nukleotiden besteht^[2], die schließlich eine doppelhelikale Struktur^[3, 4] ergeben und für die Weitergabe und Speicherung der genetischen Information verantwortlich sind^[5, 6]. Im Jahr 1953 wurde ein erster Replikationsmechanismus beschrieben^[7], welcher schließlich 1958 in *Escherichia coli* (*E. coli*) experimentell belegt werden konnte^[8]. Diese Erkenntnisse bilden noch heute im Wesentlichen die Grundlagen unseres Verständnisses der uni- und bidirektionalen Weitergabe sequenzieller Information.^[9] Das bereits 1970 postulierte "zentrale Dogma der Biochemie"^[10] ist auch noch gegenwärtig gültig. Unter dem zentralen Dogma der Biochemie versteht man die Transkription der DNA in mRNA und die anschließende Translation dieser in Proteine. Während der Translation kodiert jeweils ein Basentriplett (Kodon) für eine Aminosäure.^[9, 10]

Zu dieser Zeit standen sich zwei Weltanschauungen gegenüber. Zum einen existierte die Vorstellung der Präformationslehre, zum anderen bildete sich die Lehre der Epigenese (griech. epigenesis-nachträgliche, spätere Entwicklung) heraus. In der Präformationslehre ging man davon aus, dass die gesamte Komplexität bereits im frühen Embryo bzw. sogar in den Gameten ausgebildet ist und sich nur noch entfalten muss. Abgesehen von schädlichen Umwelteinflüssen, die schließlich zu Mutationen führen können, waren zu dieser Zeit keine weiteren chemischen Veränderungen des Genoms bekannt. Mit diesem quasi statischen Genom ist jedoch nicht erklärbar, wie man zu einer phänotypischen Diversität unterschiedlicher Zelltypen eines Organismus gelangen kann. Die Lehre der Epigenese hingegen besagt, dass der frühe Embryo noch undifferenziert vorliegt und erst im Laufe der Entwicklung bis hin zum adulten Organismus eine zunehmende Komplexität ausbildet. Das Schlüsselmerkmal ist die Entwicklung eines genetischen Plans, welcher im Grunde auf der Aktivierung oder Inaktivierung spezifischer Gene beruht. Diese Vorgänge sind auch im vollständig ausgebildeten, adulten Organismus gegenwärtig und stellen eine zusätzliche

Steuerungsebene der Genexpression dar, die es abgesehen von der reinen Sequenzinformation gibt. Diese vererbaren Programme genetischer Regulationsmuster werden der Epigenetik zugeschrieben. Diese Veränderungen können sowohl auf der Nukleinsäureebene als auch auf der Proteinebene vorkommen. Dort bewirken sie eine phänotypische Veränderung, die geschlechtlich wie auch ungeschlechtlich an die kommende Generation vererbt werden kann, aber nicht auf die zugrundeliegende DNA-Sequenz zurückzuführen ist.^[11-15]

DNA besteht hauptsächlich aus den kanonischen Nukleosiden 2'-Desoxyadenosin (dA), 2'-Desoxyguanosin (dG), 2'-Desoxycytidine (dC) und 2'-Desoxythymidin (dT). Diese sind jeweils über Phosphordiester miteinander verbunden und bilden zwei (dA:dT) bzw. drei (dC:dG) spezifische Wasserstoffbrückenbindungen zu ihrem komplementären Strang aus (siehe Abb. 2 B)^[1, 3] und ergeben die DNA-Doppelhelix-Struktur. Unter physiologischen Bedingungen liegt diese in der B-Form vor, d.h. sie weist eine kleine Furche auf, welche als Bindungsstelle für kleine Moleküle dient, und eine große Furche (Abb. 2 A), welche sequenzspezifisch mit Proteinen interagieren kann.^[16]

Lange Zeit dachte man, dass 5-Methyl-2'-desoxycytidin (mdC) in Säugetieren abgesehen von den kanonischen DNA Nukleosiden die einzige Modifikation ist.^[17] Bereits 1925 wurde dieses postuliert. 1950 wurde die Existenz von mdC schließlich bestätigt und seit 1969 wurde gemutmaßt, dass mdC eine Rolle in der Genregulation zukommt. Heute ist mdC die besterforschte epigenetische Modifikation.^[18-26] In den letzten Jahren wurden innerhalb kurzer Zeit noch vier weitere Modifikationen gefunden (Abb. 2 C): 5-Hydroxymethyl-2'-desoxycytidin^[27, 28], 5-Formyl-2'-desoxycytidin^[29], 5-Carboxyl-2'-desoxycytidin^[30, 31] und 5-Hydroxymethyl-2'-desoxyuridin (hmdU)^[32]. Auch diese Nukleoside stellen vermutlich epigenetische Marker dar und unterliegen einer dynamischen Regulation. Allerdings sind die Wirkung und Funktion der einzelnen Nukleoside weniger gut erforscht. Es ist nicht sicher, ob sie überhaupt definierte Funktionen besitzen oder ob sie nur die Intermediate eines oxidativen Demethylierungsprozesses sind.^[33, 34]

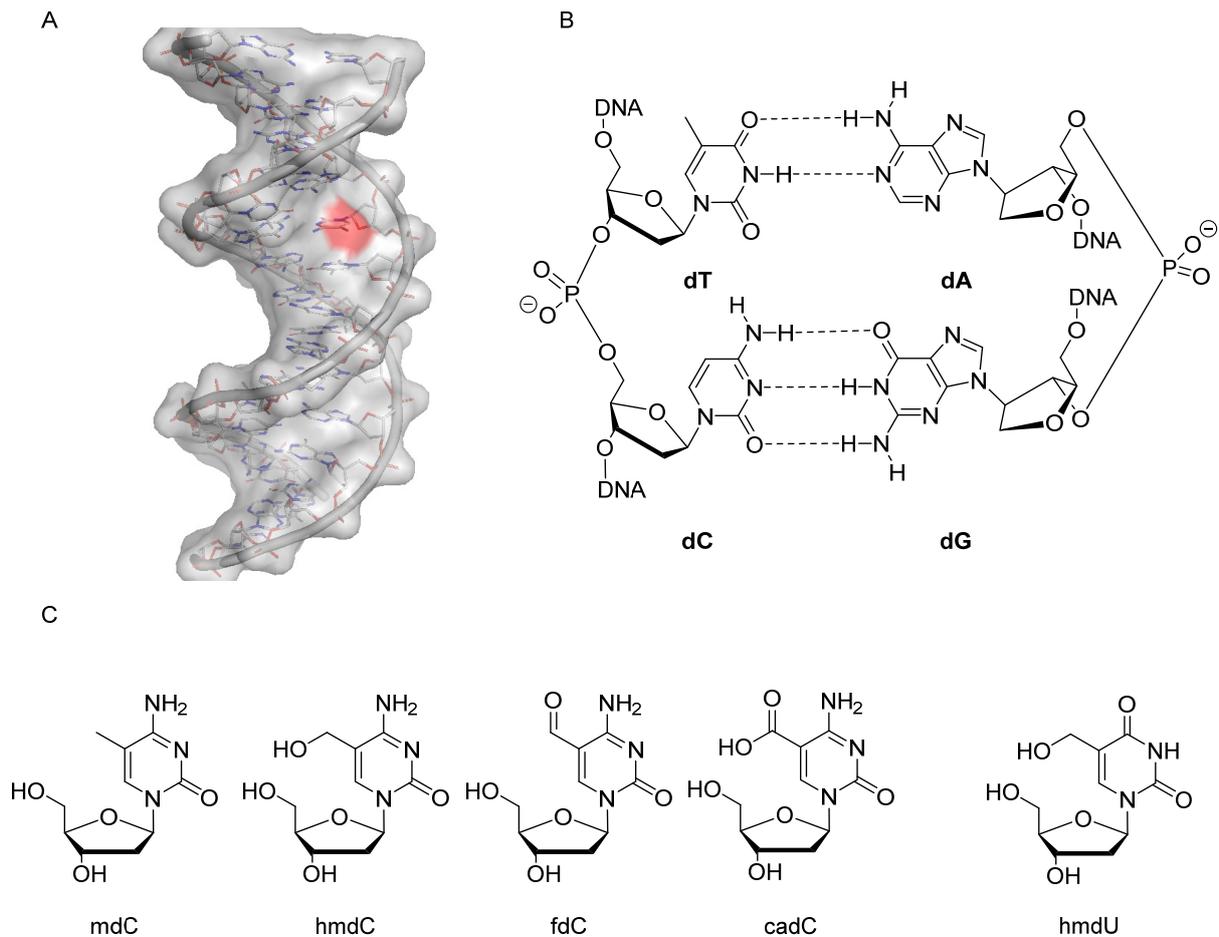


Abbildung 2: A): Die Modifikation mdC (rote Markierung) bzw. deren Methylgruppe deutet in die große Furche der DNA-Doppelhelix und kann an dieser Stelle Protein-DNA Wechselwirkungen beeinflussen (PDB ID 3a5t). B): Die Basenpaarungen der doppelsträngigen DNA: dT bildet mit dA zwei Wasserstoffbrückenbindungen und dC mit dG drei (angedeutet durch die gestrichelte Linie). C): Die chemischen Strukturen der epigenetischen Nucleoside der DNA.

Neben den epigenetischen Markern in der DNA existieren auch Histonmodifikationen (Abb. 3), denen eine vererbare Veränderung mit regulatorischen Funktionen zugeschrieben wird. Diese Veränderungen auf der Proteinebene können im direkten funktionellen Zusammenhang mit den epigenetischen DNA-Modifikationen stehen.^[35] Das Nukleosom ist die kleinste Packungseinheit der DNA im Zellkern. Eines besteht aus einem 145-147 Basenpaar (bp) langen DNA-Abschnitt und windet sich 1,65-mal um sogenannte Histonoktamere.^[36] Chemische Modifikationen dieser Histonproteine können nun direkten Einfluss auf die Packungsdichte des Chromatins nehmen und somit z.B. die Transkriptionsrate dieser gebundenen DNA beeinflussen.^[37-41] Für eine aktive Transkription muss die Zugänglichkeit der DNA für die an der Transkription beteiligten Enzyme gewährleistet sein.^[42] Histone haben N-terminal einen flexiblen Aminosäurerest, welcher posttranslational durch Enzyme modifiziert werden kann. Diese Modifikationen reichen von Methylierungen, Acylierungen (Acetyl-, Propionyl-, Butyryl-) und Phosphorylierungen bis hin zu komplexeren Modifikationen wie

Ubiquitinierung und Sumoylierung, bei welchen sogar kleinere Proteine angeknüpft werden.^[43-45] Dieses Muster an posttranslationalen Histonmodifikationen wird oft als Histon-Code bezeichnet.

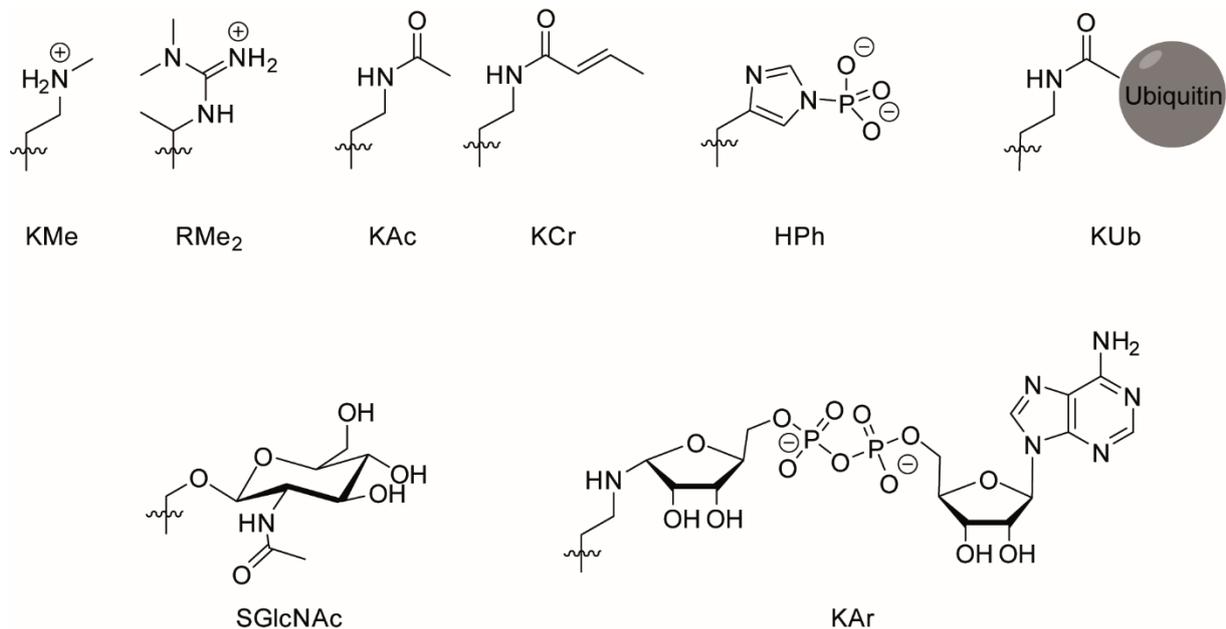


Abbildung 3: Die Aminosäuren K (Lysin); R (Arginin), H (Histidin) und S (Serin) können posttranslational in Histonproteinen (PTMs) modifiziert werden und sind beispielhaft gezeigt. Me = Methylierung, Ac = Acetylierung, Cr = Crotonylierung, Ph = Phosphorylierung, Ub = Ubiquitinylierung, GlcNAc = O-N-Acetylglucosamin; Ar = ADP-Ribosylierung.

Neben den epigenetischen Modifikationen der DNA ist auch eine RNA-Epigenetik^[46] denkbar. Die Forschung auf diesem Gebiet ist im Vergleich dazu noch jung. Bedenkt man aber, dass die RNA im Gegensatz zur DNA eine große Vielzahl an verschiedenen Modifikationen aufweist, so könnte es durch die vielen verschiedenen chemischen Strukturen zu einer Regulation kommen. Es wurde bereits gezeigt, dass diese posttranskriptionalen RNA Modifikationen auch in mRNA vorkommen und dynamischer Regulation unterliegen können.^[46-49]

1.2 Modifikationen der Ribonukleinsäuren

Ribonukleotide weisen eine deutlich größere strukturelle Vielfalt auf als DNA. Der hohe Modifikationsgehalt der RNA kann als eine Anpassung an ihre vielen, verschiedenen Funktionen im Organismus betrachtet werden.^[50] Neben dem Transfer der genetischen Information erfüllt sie ebenso regulatorische Aufgaben^[51, 52] und zeigt, in Form sogenannter Ribozyme, sogar katalytische Aktivität.^[53] Die enorme Vielfalt ihrer Aufgaben spiegelt sich auch in der Anzahl der unterschiedlichen RNA-Spezies wider. Diese reichen von den besser erforschten *transfer* RNAs (tRNAs), der *ribosomal* RNA (rRNA) und mRNA bis hin zu weniger bekannten Spezies wie *long non-coding* RNA (lncRNA), *small nuclear* RNA (snRNA), *micro* RNA (miRNA), *piwi-interacting* RNA (piRNA), *small-interfering* RNA (siRNA) und der *small nucleolar* RNA (snoRNA). Diese nicht-kodierenden RNAs erfüllen verschiedene Funktionen vor allem im Bereich regulatorischer Prozesse, wie dem Spleißen, sowie im Bereich der Differenzierung und Entwicklung, wie z.B. der X-Chromosom Inaktivierung.^[51, 54, 55]

Der Modifikationsgrad und die damit einhergehenden Funktionen sind bisher nur unvollständig untersucht. Bis heute sind 163 verschiedene posttranskriptionale Modifikationen der kanonischen RNA-Nukleoside bekannt,^[50, 56, 57] wobei die Komplexität dieser Modifikationen stark variiert (Abb. 4). Am häufigsten treten einfache Methylierungen auf, aber auch Alkylierungen, Acetylierungen und Thiomethylierungen sind bekannt. Wie die Modifikation Pseudouridin (Ψ) zeigt, ist auch eine Isomerisierung von einer C-N-glykosidischen Bindung zu einer C-C-glykosidischen Bindung möglich. Auch eine Erweiterung des Puringrundgerüsts, wie bei Wybutosin (yW), wurde bereits entdeckt. Generell können nahezu alle Positionen des Purin- und Pyrimidingrundgerüsts modifiziert vorliegen (Abb. 4 A und 4 B). RNA-Modifikationen können im Grunde in einfache Modifikationen (Abb. 4 C und 4 D) und Hypermodifikationen (Abb. 4 E) unterteilt werden. Als Hypermodifikationen gelten Nukleoside mit langen Seitenketten oder mit einem veränderten Heterozyklus.^[33] Chemische Modifikationen bringen eine große funktionelle Vielfalt mit sich. Zum einen können Modifikationen direkt die RNA-Struktur beeinflussen, z.B. durch Einfluss auf intramolekulare Interaktionen oder durch eine veränderte Flexibilität des RNA-Moleküls. Zum anderen können auch intermolekulare Interaktionen, z.B. zu Proteinen, beeinflusst werden. Auch kleine Veränderungen der chemischen Struktur der RNA können somit zu großen Veränderungen im Zellmetabolismus führen.^[56]

Vor allem die Nutzung von Flüssigchromatographie gekoppelt mit Massenspektrometrie (LC/MS) ermöglichte, dass stetig neue RNA-Modifikationen entdeckt und deren biologische bzw. physiologische Funktion aufgeklärt werden.^[56]

Gerade die tRNA ist mit mindestens 79 verschiedenen Modifikationen divers modifiziert und weist eine hohe Modifikationsdichte auf. Somit enthält eine tRNA durchschnittlich 8

Modifikationen.^[58] Dabei liegen die Positionen 34 (*Wobble-Position*) und 37 (*extended anticodon*) der Antikodonschleife fast quantitativ modifiziert vor.^[59-61]

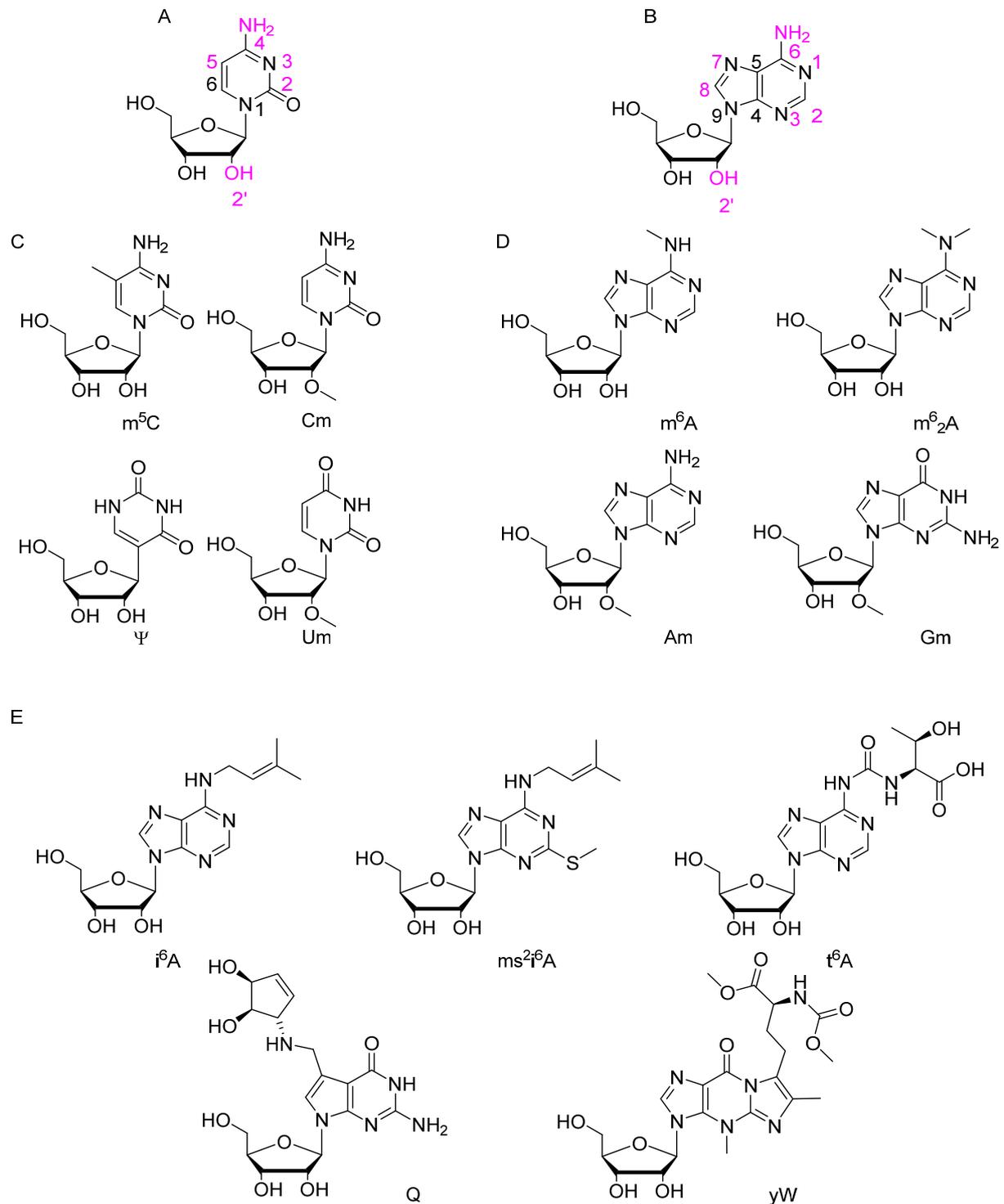


Abbildung 4: A): Cytidin als Beispiel für Pyrimidine. B): Adenosin als Beispiel für Purine. A) und B): Entsprechende Nomenklatur der Grundgerüste; die Pink markierten Positionen können chemischer Modifizierung unterliegen. C): Beispiele einfacher Modifikationen des Pyrimidingrundgerüsts. D): Beispiele einfacher Modifikationen des Puringrundgerüsts. E) Strukturen von bekannten Hypermodifikationen.

Modifikationen an diesen beiden Positionen führen zu starken Veränderungen der Struktur der Antikodonschleife und haben so Einfluss auf die Erkennung der korrekten Kodon-Antikodon Bindung.^[62] Viele verschiedene Modifikationen können an Position 37 vorkommen, zwei davon sind i^6A und ms^2i^6A . Diese kommen in tRNA exklusiv an dieser Position vor.^[63] Manche Modifikationen beschränken sich auf bestimmte RNA-Spezies, während andere omnipräsent sind, wie z.B. Pseudouridin (Ψ).^[33, 64] Etwa 1/3 aller bekannten Modifikationen wurde in rRNA und ca. 1/10 in mRNA nachgewiesen. Neben den starken Variationen des Modifikationsgehalts zwischen den einzelnen RNA-Spezies kommt es auch zu Unterschieden zwischen verschiedenen Organismen^[65, 66] sowie zwischen den unterschiedlichen Geweben innerhalb eines komplexen Organismus.^[67, 68] Zunächst sind die Sequenzen der jeweiligen RNA-Moleküle in den verschiedenen Zellen eines Organismus identisch; es können aber gewebespezifisch unterschiedliche Mengen an bestimmten Modifikationen eingebaut werden. Diese Abänderung könnte eine Konsequenz des zelltypischen Metabolismus sein, welche folglich eine Regulierung der Translation mit sich bringt. Ebenso wäre es möglich, dass das Modifikationsmuster von einer bestimmten Translationsaktivität verursacht wird.^[33] Weitere Faktoren, die im Rahmen von Modifikationsprozessen eine Rolle spielen können, sind äußere Einflüsse.^[69] So katalysieren in *Saccharomyces cerevisiae* (*S. cerevisiae*) Trm4 und Trm7 unter H_2O_2 Einfluss die Bildung von m^5C und Cm. Durch genaue LC/MS Quantifizierung von tRNA Nukleosiden konnte gezeigt werden, dass *S. cerevisiae* Trm4 bzw. Trm7 Mutanten eine erhöhte Empfindlichkeit gegenüber H_2O_2 aufweisen. Dadurch steigen die Mengen an m^5C und Cm. Die einfache Erklärung ist, dass die Zunahme einer spezifischen tRNA-Modifikation notwendig ist, um eine effiziente Stressreaktion zu fördern. Das gleiche gilt für m^2_2G , doch in diesem Fall hat der Verlust eines an seiner Biosynthese beteiligten Enzyms, Trm1, keine H_2O_2 -Empfindlichkeit zur Folge. Dieses Verhalten ist bereits in mRNA bekannt. Viele der stressinduzierten Transkripte sind nicht essentiell das das Überleben der Zelle unter Stresseinfluss. Dies zeigt, dass Modifikationen in der tRNA entscheidende Funktionen in der zellulären Stressreaktion wahrnehmen. Die zelluläre Antwort auf Reize wie toxische Stimuli ist vermutlich eine dynamischen Reprogrammierung der tRNA-Modifikationen.^[69]

1.2.1 Hypermodifikationen: i^6A und ms^2i^6A

Wie bereits erwähnt, gelten die beiden Modifikationen i^6A und ms^2i^6A aufgrund der langen Alkylkette als sogenannte Hypermodifikationen. Im Allgemeinen ist die bakterielle Biosynthese von ms^2i^6A ausgehend von Adenosin in tRNA ein gut erforschter Prozess (siehe Abb. 5). Die involvierten Enzyme konnten bereits identifiziert und teils auch charakterisiert werden: *Methylthio-isopentenyl-adenosine synthesis enzyme A* (MiaA) und *Methylthio-isopentenyl-adenosine synthesis enzyme B* (MiaB).^[70]

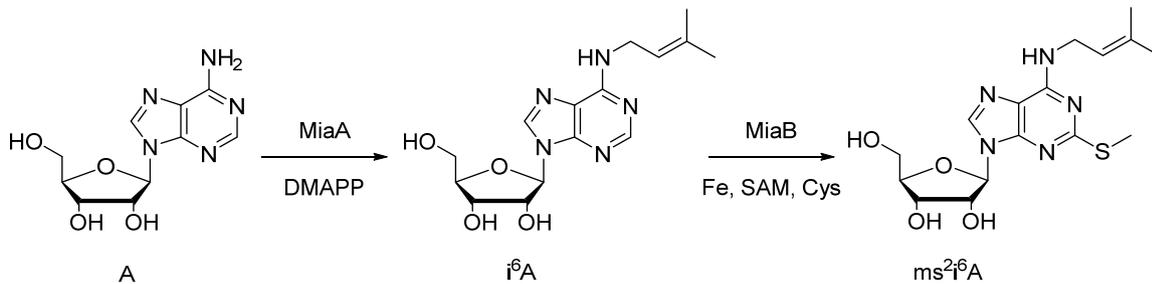


Abbildung 5: Schematische Darstellung der Biosynthese der Isopentenyl-Modifikationen in Bakterien. Die Prenyltransferase MiaA überträgt zunächst eine Dimethylallyl-Gruppe aus DMAPP (Dimethylallyl-diphosphat) auf A, welches anschließend durch die Methylthiotransferase MiaB in einer SAM-abhängigen Reaktion thiomethyliert wird.^[66, 70]

In einem ersten Reaktionsschritt katalysiert die Prenyltransferase MiaA die Übertragung einer Dimethylallyl-Gruppe auf das exozyklische N^6 des Adenosins mit Hilfe des Cosubstrates Dimethylallyl-diphosphat (DMAPP). Das so entstandene i^6A ,^[71, 72] wird durch die Methylthiotransferase MiaB in einem zweiten Reaktionsschritt zu ms^2i^6A umgesetzt.^[73] MiaB gehört zur Familie der Radikal-SAM-Enzyme. Bei der Bildung von ms^2i^6A überträgt MiaB mit Hilfe zweier [4Fe-4S]-Cluster eine Thiomethylgruppe auf i^6A .^[74-77] In einer Studie konnte beobachtet werden, dass in Abwesenheit von MiaA weder i^6A noch ms^2i^6A aufzufinden ist. Daraus könnte man ableiten, dass nur i^6A , nicht aber A als Substrat für MiaB dient.^[78] Das Problem dieser Studie ist, dass das potentielle Produkt der Methylthiotransferase MiaB mit dem Substrat A zu ms^2A ^[79, 80] führen würde, was allerdings nicht betrachtet wurde. In humanen Zellen wird die Biosynthese von i^6A durch das *MiaA* Homolog *tRNA isopentenyl transferase 1* (TRIT1) katalysiert.^[81] Das humane Ortholog zum bakteriellen MiaB stellt *CDK5 regulatory subunit associated protein 1* (CDK5RAP1) dar. Dieses Enzym gehört ebenfalls zur Familie der Radikal-SAM-Enzyme und katalysiert die Synthese von ms^2i^6A aus i^6A in nuklearer und mitochondrialer RNA.^[33, 82] CDK5RAP1 ist ein Repressor der gut untersuchten *cyclin-dependent protein kinase 5* (CDK5), die als Teil der *neuronal Cdc2-like kinase* (Nclk) bei der Regulation der Gehirnentwicklung wichtige Funktionen erfüllt. So beeinflusst sie die neuronale Zelldifferenzierung, die Apoptose, die Neuronen-Migration sowie die Bildung neuromuskulärer

Zellkontakte. Aus diesem Grund wird sie auch oft mit neurodegenerativen Krankheiten, wie Alzheimer oder Parkinson in Verbindung gebracht und spielt bei der Insulin-Ausschüttung eine wichtige Rolle.^[83, 84]

1.2.1.1 Einfluss auf die Translation

Der genetische Code wird mit Hilfe von tRNA Molekülen translatiert. Dabei vermitteln Ribosomen die Interaktion des Kodons der mRNA mit dem Antikodon der tRNA. Paarung von Kodon und Antikodon positionieren so die tRNA, die eine Kodon-spezifische Aminosäure trägt. Die herangeführten Aminosäuren werden so schrittweise durch Peptidbindungen zu einer Aminosäurekette verknüpft. Die tRNA wird während ihrer Reifung posttranskriptional modifiziert. Diese Modifikationen können an unterschiedlichen Stellen vorkommen. Von allen modifizierten tRNA-Nukleosiden in allen Domänen des Lebens kommen lediglich acht Modifikationen immer an der gleichen Stelle vor.^[85] Daraus lässt sich ableiten, dass tRNAs einen sehr konservierten Ursprung und feste Funktionen besitzen. Im Gegensatz dazu kommt die größte Modifikationsvielfalt an der *Wobble* Position 34 sowie der Position 37 vor.^[86, 87] Aufgrund der räumlichen Nähe zum Antikodon kann man davon ausgehen, dass Modifikationen an dieser Stelle Auswirkungen auf die Kodon-Antikodon-Paarung haben. Lange Zeit dachte man, dass ms²i⁶A ausschließlich an Position 37 der tRNA vorkommt^[88] und dort dafür verantwortlich ist, schwache Wechselwirkungen zwischen A und U zu verstärken.^[89] In einer neueren Studie wurden strukturelle und dynamische Änderungen, die durch die Modifikation i⁶A an Position 37 auftreten, untersucht. Der unmodifizierte tRNA^{Phe} *Hairpin* zeigt eine Helix in A-Form mit den Basenpaarungen U₃₂·A₃₈ und U₃₃·A₃₇ und bildet nicht das kanonische *U-turn* Motiv^[90] aus. Schmelzpunktmessungen des i⁶A₃₇tRNA^{Phe} *Hairpins* zeigten, dass es durch die Einführung der Modifikation zu einer Schmelzpunktniedrigung kommt. Die Prenylierung destabilisiert die *Hairpin*-Sekundärstruktur nahe des *Loops* und steigert so die Flexibilität des *Loops* und der benachbarten Nukleotide. Das deutet daraufhin, dass in Abwesenheit von z.B. i⁶A an Position 37 ein drei-Nukleotide-*Loop* entsteht. Indem i⁶A den Antikodon-*Loop* entspannt erleichtert es so die Kodon-Antikodon Erkennung.^[91] Das gleiche gilt für die Modifikation N⁶-Threonylcarbamoyladenosenin (t⁶A). Besetzt t⁶A die Position 37, so bildet es eine planare Struktur. Diese weist Ähnlichkeit zu einem dritten Heterozyklus auf und kann so zu vermehrtem *Stacking* mit den benachbarten Basen führen. Durch diese Rotation der sterisch anspruchsvollen Threonyl-Gruppe ist es nicht mehr möglich, die Position 37 in die helikale Struktur mit einzubinden und führt so zu einer Entspannung des *Loops* und einer erleichterten Kodon-Antikodon Bindung.^[92] Dieses Ergebnis zeigt sich auch bei ms²i⁶A₃₇.^[93] Jenner *et al.* sind der Frage nachgegangen, wie die Modifikation ms²i⁶A₃₇ durch schwache Kodon-Antikodon Wechselwirkungen die Verschiebung des translatorischen Leserahmens verhindern kann. Wie in Abbildung 6 gezeigt, stabilisiert ms²i⁶A an der Position

37 in tRNA^{Phe}_{GAA} die mRNA-tRNA Interaktion in allen drei tRNA-Bindungsstellen. Vor allem die Thiomethylgruppe des ms²ⁱ⁶ A₃₇ führt zu *Stacking* mit der ersten Base des abzulesenden Kodons und stabilisiert so in allen drei Bindungsstellen (A, P und E) des Ribosoms (Abb. 6) die korrekte Kodon-Antikodon Paarung.

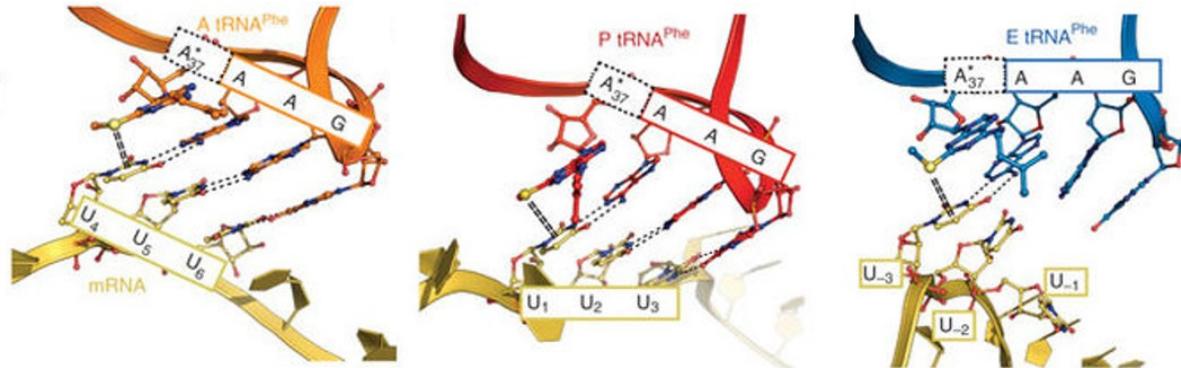


Abbildung 6: Die Modifikation ms²ⁱ⁶A an der Position 37 des Antikodons der tRNA stabilisiert die Basenpaarungen zum Kodon der mRNA während der Translation. Die Thiomethylgruppe des ms²ⁱ⁶A führt zu *Stacking* zwischen ms²ⁱ⁶A₃₇ und der ersten Base des Kodons. Dies ist bei allen drei Bindungsstellen des Ribosoms der Fall. (Abbildung nach Jenner *et. al.*)^[94]

Dies stellt sicher, dass der Leserahmen stabil gesetzt ist.^[94] Das ist ein Beispiel, wie ein Modifikationsmangel von ms²ⁱ⁶A₃₇ in tRNAs zu Verschiebungen des translatorischen Leserahmens führen könnte.^[94]

Zusammenfassend kann man also sagen, dass N⁶-modifizierte Nukleotide an der Position 37 durch die Erhöhung der Flexibilität im Antikodon-Loop zu einer besseren Bindung des Antikodons zum Kodon führen und der ms²-Substituent von ms²ⁱ⁶A eine stark stabilisierende Wechselwirkung mit der ersten Base des Codon-Tripletts eingehen kann.

1.2.1.2 Bedeutung in Krankheitsbildern

Für die meisten Modifikationen ist bisher unklar, bis zu welchem Grad sie das Krankheitsbild eines bestimmten Organismus prägen. In einigen Studien konnte gezeigt werden, dass Veränderungen der RNA-Modifikationen zur Entwicklung von Krankheiten beitragen können. Hierzu zählen insbesondere Erkrankungen des Stoffwechsels.

Die Existenz von i^6A und ms^2i^6A ist für das Überleben der entsprechenden Zellen zwar nicht existentiell, jedoch sind beim Fehlen des Enzyms MiaA bestimmte zelluläre Funktionen beeinträchtigt. Dazu zählt beispielsweise, dass die Translationsgeschwindigkeit verringert ist, was letztendlich auch das Zellwachstum reduziert. [70]

Um die Bedeutung der Modifikation ms^2i^6A in Krankheitsbildern zu erforschen ist es wichtig deren genaue Position zu kennen. Im Zuge dessen wurde in einer aktuellen Studie das Substrat von Cdk5rap1 in Mäusen identifiziert. Hierfür wurden mitochondriale RNAs (mt-RNA) aus Wildtyp-Mäusen mit denen der Cdk5rap1 *Knockout* (KO)-Mäuse verglichen. Die Modifikation ms^2i^6A wurde so in genau vier mt-RNAs an Position A₃₇ nachgewiesen: mt-RNA^{Phe}, mt-RNA^{Trp}, mt-RNA^{Tyr} und mt-RNA^{Ser(UCN)}. [95] Das Fehlen dieser Modifikation induziert eine sogenannte Verschiebung des Leserahmens an bestimmten Kodons. Der erhaltene Phänotyp dieser KO-Maus zeigt mitochondriale Dysfunktionen, aber keine morphologische Änderung des Gewebes. Des Weiteren gab es auch keine neurologischen Unterschiede zwischen der Wildtyp- und der KO-Maus. Genauere Untersuchungen in *mouse embryonic fibroblast* (MEF) Cdk5rap1 KO-Zellen zeigten, dass die mitochondriale Proteinsynthese stark verringert ist. Vor allem die Bildung der mitochondrialen Proteinkomplexe I, III und IV war stark reduziert. Es liegt deshalb nahe, dass ein Fehlen von ms^2i^6A die mitochondriale Proteinsynthese stark verringert. Als nächstes wurde ein Zusammenhang zur oxidativen Phosphorylierung (OXPHOS) gefunden. Die Instandhaltung der mitochondrialen OXPHOS Untereinheiten ist für die Elektronen-Transportkette von größter Bedeutung: Sie erhält das Membranpotential und ist somit für die Atmung verantwortlich. Unter den KO-Zellen wiesen Zellen mit sehr kleinem Membranpotential eine hohe Zelldichte auf, was zu einer erniedrigten Sauerstoff-Aufnahme führt. ms^2i^6A in mt-RNAs ist somit essentiell für die Translation in den Mitochondrien und für die Atmungskette. Die mitochondriale Aktivität der Skelettmuskulatur ist stark verringert. Das ist nicht verwunderlich, da von allen Geweben die Skelettmuskulatur und das Herzgewebe am stärksten von mitochondrialer Dysfunktion betroffen sind. Auch eine Verringerung des ATP-Levels war in den KO-Mäusen messbar. Die *reactive oxygen species* (ROS) waren im Vergleich zur Wildtyp-Maus hingegen erhöht. ROS ist ein Nebenprodukt des mitochondrialen Elektronentransports und wird vor allem in den mitochondrialen Proteinkomplexen I und III generiert. Da es an diesen in der KO-Maus aber mangelt, kommt es vermehrt zum Austritt von ROS. Dies wiederum kann zu mitochondrialen Erkrankungen führen. Diese Mäuse sind physisch weniger leistungsstark und anfälliger für

kardiale Dysfunktionen. Die Auswirkungen des KOs sind unter Stressbedingungen noch verstärkt.^[95]

In einer anderen Studie wurde überprüft, ob die $ms^{2i6}A$ -Mengen der $tRNA^{Trp}$, $tRNA^{Phe}$, $tRNA^{Tyr}$ und $tRNA^{Ser(UCN)}$ in der A3243G Mutante mit dem Expressionslevel von CDK5RAP1 korrelieren. Die A3243G Mutante ist in den peripheren Blutzellen von Patienten mit der Erkrankung *Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes* (MELAS) zu finden. MELAS und *Myoclonic epilepsy with ragged red fibers* (MERRF) werden durch ein Fehlen von Modifikationen (τm^5C bzw. τm^5s^2U) in der Wobble Position mitochondrialer tRNAs ausgelöst.^[96-98] Die Studie zeigte, dass die $ms^{2i6}A$ -Mengen der $tRNA^{Trp}$, $tRNA^{Phe}$, $tRNA^{Tyr}$ und $tRNA^{Ser(UCN)}$ in der Mutante verringert waren, aber nicht mit dem Expressionslevel von CDK5RAP1 korrelierten. Das schließt einen direkten Effekt aus. Die A3243G Mutanten weisen allerdings eine reduzierte mitochondriale Proteinsynthese und ein erhöhtes Stresslevel auf. Es könnte also sein, dass der erhöhte oxidative Stress zu einer Inhibierung des oxidationsempfindlichen CDK5RAP1 Enzyms führt. Dieses Ergebnis konnte durch die Behandlung mit H_2O_2 reproduziert werden.^[95] Somit führt also die Verringerung von $ms^{2i6}A$ durch oxidativen Stress indirekt zur einer Beeinflussung der Qualität der Mitochondrien und so zum Fortschreiten mitochondrialer Krankheiten.^[95]

Um zu überprüfen, ob die Modifikation $ms^{2i6}A$ in einem direkten Zusammenhang mit einer Krankheit steht, wurde eine Fall-Studie an Vitiligo (Weißfleckenkrankheit)-Patienten der koreanischen Bevölkerung durchgeführt. Es wurde untersucht ob, CDK5RAP1-Polymorphismen der Grund für die Ausbildung von Vitiligo sind; dafür wurden 296 Vitiligo-Patienten mit 426 gesunden Menschen/Kontrollen verglichen. Insgesamt wurden zwei *single nucleotide polymorphisms* (SNPs) von CDK5RAP1 festgestellt.^[99] Diese Ergebnisse deuten darauf hin, dass CDK5RAP1 in der koreanischen Bevölkerung zur Entstehung von Vitiligo beitragen könnte. Die Anzahl der Stichproben (269 Patienten) in dieser Studie ist allerdings klein und ein kausaler Zusammenhang konnte nicht zweifelsfrei belegt werden.

Diese Studien liefern Hinweise auf schädliche Veränderungen des Zellstoffwechsels, die indirekt oder direkt auf das Fehlen von $ms^{2i6}A$ zurückzuführen sind.

1.2.2 Modifikationen in mRNA

Im Gegensatz zur tRNA ist die mRNA weitaus weniger modifiziert, doch ist diese seit einiger Zeit Gegenstand intensiver Forschung. Die posttranskriptionalen Modifikationen dieser RNA Spezies sind von großem Interesse, da sie die Grundlage für dynamische Regulation an zellulären Prozessen sein könnte. Ein Grund für eine dynamische Regulation ist, dass Veränderungen in der mRNA nicht zwingend einen globalen Langzeit-Effekt mit sich bringen, wie es bei tRNA oder auch rRNA der Fall ist, sondern im Prinzip eine transkriptspezifische Regulation ermöglichen.

Einige Modifikationen, wie m⁷G, sind bereits sehr gut erforscht. Es ist unter anderem Bestandteil der 5'Cap-Struktur eukaryotischer mRNAs und definiert diese RNA dadurch als mRNA. Darüber hinaus konnte gezeigt werden, dass die Methylierung der 5'Cap-Struktur für die Translation der meisten mRNAs benötigt wird und diese gegen Angriffe von Exonukleasen schützt. Darüber hinaus trägt m⁷G zur Transkription, dem Spleißen, der Polyadenylierung und dem Export aus dem Nukleus bei.^[33, 100]

1.2.2.1 Inosin

Ein weiteres Beispiel einer bereits gut charakterisierten mRNA Modifikation ist Inosin, welches aus Adenosin durch *Adenosine Deaminases Acting on RNA* (ADARs) durch hydrolytische Desaminierung in doppelsträngiger RNA entstehen kann.^[101] Dieser Vorgang ist als Teil des RNA *editings* bekannt und führt zu einer veränderten Basenpaarung (C anstelle von U). Dadurch kann es zu spezifischen Aminosäure-Substitutionen, alternativem Spleißen, sowie Veränderungen in der Genexpression kommen, was die Vielfalt der mRNA und so auch die der Proteine in Eukaryoten erhöht.^[102] Wie oft es zu diesem A zu I *editings* kommt und welche spezifischen Funktionen es bei bestimmten Genen beeinflusst, ist bisher noch ungewiss. Man geht aber davon aus, dass zu hohes oder zu niedriges RNA *editing*, oder auch fälschliches *editing*, zu vielen humanen Krankheitsphänotypen bis hin zu Krebs führen kann.^[102-105] Ein Beispiel ist die Veränderung des RNA *editings* der prä-mRNA des Serotoninrezeptors 2C (HTR2C).^[106] Die Desaminierung beeinflusst die Zusammensetzung der Aminosäuren des 5-HT_{2C}R-Rezeptors. Dieses induziert Konformationsänderungen, die letztendlich die Signalweiterleitung des Serotonins verringern.^[107-110] Die Verteilung von RNA *editing* Enzymen, die eine Desaminierung von A zu I katalysieren, ist regional unterschiedlich im Gehirn. Ein deutlicher Anstieg des *editings* von 5-HT_{2C}R wurde vor allem im anterioren cingulären Cortex bei Selbstmördern beobachtet, was diesen kortikalen Bereich und die chemische Modifikation mit einem erhöhten Selbstmordrisiko in Zusammenhang bringt. Das deutet darauf hin, dass regionsspezifische Veränderungen des RNA *editings* von 5-HT_{2C}R

mRNA und mangelnde Rezeptorfunktion zur Entstehung einer schweren depressiven Störung beitragen können.^[106]

Es scheint naheliegend, dass es sich um eine aktiv gesteuerte Desaminierung in mRNA handelt, da die ADAR-Expression mit dem gemessenen Gehalt an Inosin korreliert und gewebsspezifische Unterschiede der Mengen auftreten.^[111] Neuere Erkenntnisse konnten durch Sequenzierungsdaten erhalten werden. Eine robuste Sequenzierung ist durch die chemische Derivatisierung von Inosin mit Acrylnitril möglich (Abb. 7).

Aufgrund des entstandenen *N*¹-Cyanoethylinosins kommt es zum Abbruch der reversen Transkription. Dieses ermöglicht so die Lokalisierung von Inosin-Loci.^[112, 113]

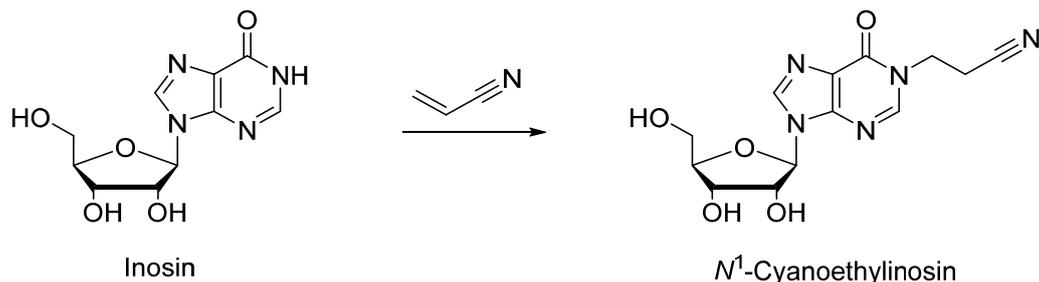


Abbildung 7: Chemische Derivatisierungsreaktion von Inosin zu *N*¹-Cyanoethylinosin.

Trotzdem ist das Ausmaß, in dem *A-to-I editing* vorkommt, noch nicht genau bestimmt. Vor allem bedingt durch die starken Abweichungen zahlreicher Sequenzierungsdaten voneinander kommt es zu unterschiedlichen Interpretationen der Ergebnisse.^[114, 115]

1.2.2.2 *N*⁶-Methyladenosin - ein potentieller epigenetischer Marker in mRNA

Bereits 1974 wurde die Modifikation *N*⁶-Methyladenosin (*m*⁶A) in humaner mRNA gefunden^[116, 117] und kurz darauf wurden erste Konsensmotive, wie G*m*⁶AC (70 %) und A*m*⁶AC (30 %), in der mRNA von HeLa Zellen entdeckt.^[118, 119] Heutzutage hat diese Modifikation einen besonderen Stellenwert unter den bekannten Modifikationen der mRNA, denn *m*⁶A ist sehr häufig und ubiquitär in humaner und muriner mRNA vorhanden. Man geht davon aus, dass im Durchschnitt drei *m*⁶A-modifizierte Nukleotide pro Transkript vorkommen.^[120-123] Die Erforschung der Verteilung von *m*⁶A über das gesamte Transkriptom konnte durch Fragmentierung und Immunopräzipitation von mRNA erfolgen, denn *m*⁶A wird, genauso wie A, zu T revers transkribiert. Auch chemische Derivatisierungsreaktionen der Methylgruppe sind nicht möglich.^[124, 125] Durch diese antikörperbasierte Sequenzierungsmethode (MeRIP-Seq) konnten ungefähr 12000 Positionen, die *m*⁶A enthalten, in ca. 7000 kodierenden und 300 nicht-kodierenden Genen humaner Zellen identifiziert werden. Dabei kann man sagen, dass *m*⁶A vor allem in kodierenden Sequenzen und der 3'UTR-Sequenz vorkommt. Des Weiteren wurde gezeigt, dass die globalen Methylierungsmuster in definierten Loci von Maus und Mensch

konserviert sind.^[124, 125] Das bedeutet, dass auch in nicht-verwandten Organismen die m⁶A-Positionen gleich sind.

Die damalige Auflösung von ungefähr 100-200 Nukleotiden konnte seither noch weiter verbessert werden, doch eine Auflösung bis hin zu einzelnen Nukleotiden ist auch heute noch eine große Herausforderung. Eine neue Methode verwendet hierfür Anti-m⁶A Antikörper, welche durch UV-induziertes *crosslinking* spezifische Veränderungen an m⁶A-Resten hervorrufen.^[126] Diese UV-induzierten Antikörper-RNA-Addukte können durch darauffolgende Sequenzabweichungen in der reversen Transkription detektiert werden. Diese neuartige Technologie heißt miCLIP (*m⁶A individual-nucleotide-resolution crosslinking and immunoprecipitation*). Die folgende Abbildung (Abb. 8) zeigt die Unterschiede der beiden Sequenzierungsmethoden.

Abgesehen von der verbesserten Auflösung hat die miCLIP Methode einen weiteren großen Vorteil, sie kann zusätzlich zu m⁶A auch N⁶,2'-O-Dimethyladenosine (m⁶Am) detektieren. Durch die Verwendung von zwei unterschiedlichen Anti-m⁶A-Antikörpern kommt es entweder zu Abbrüchen oder zu C→T Mutationen während der reversen Transkription. Der Antikörper, der zu Abbrüchen (*crosslinking-induced truncation sites* (CITS)) in der reversen Transkription führt, kann auch m⁶Am an den *transcription start sites* (TSSs) detektieren. Die Analyse von Abbrüchen in der 5'-UTR zeigt, dass diese Abbrüche entweder im bekannten m⁶A-Motiv DRACH (D = A, G oder U; H = A, C oder U) oder in einem neuartigen Motiv BCA* (B = C, U oder G; A* = methylierbares A) auftraten. Die 5'UTRs enthielten fast dreimal so viele methylierte BCA-Motive wie DRACH-Motive. Interessanterweise ähnelt das BCA-Motiv dem der bekannten pyrimidinreichen Sequenz an den Transkriptionsstartstellen^[127, 128], was darauf hindeutet, dass diese Stellen eher m⁶Am als m⁶A enthalten. Indirekt kann also auch zwischen m⁶A und m⁶Am unterschieden werden.

Normale Immunopräzipitationsexperimente sind bisher nicht dazu befähigt zwischen m⁶A und m⁶Am, denn die Base (6-Methyladenin) der beiden Modifikationen ist exakt gleich. Antikörper sollten deshalb in Zukunft nur als spezifische Antikörper gegen eine Base und nicht gegen ein Nukleotid/Nukleosid bezeichnet werden, sofern sie nicht auch Modifikationen am Zucker bzw. den Zucker unterscheiden.

Neue Sequenzierungsergebnisse zeigten nun auch das Transkriptom von m⁶Am. Warum die Unterscheidung zwischen den beiden Nukleotiden (m⁶A und m⁶Am) so wichtig ist, zeigten die jüngsten Ergebnisse. Lange Zeit wurde angenommen, dass die Methylgruppe des m⁶A durch das *Fat Mass and Obesity-associated Protein* (FTO) aktiv entfernt wird. Man dachte, dass FTO m⁶A in zwei Schritten zu hm⁶A und f⁶A oxidiert und diese Modifikationen wiederum nur bedingt stabil sind und somit zu Adenosin hydrolysiert werden.^[129] Nun konnte gezeigt werden, dass FTO bevorzugt die N⁶-Methylgruppe von m⁶Am gegenüber der von m⁶A entfernt.^[130]

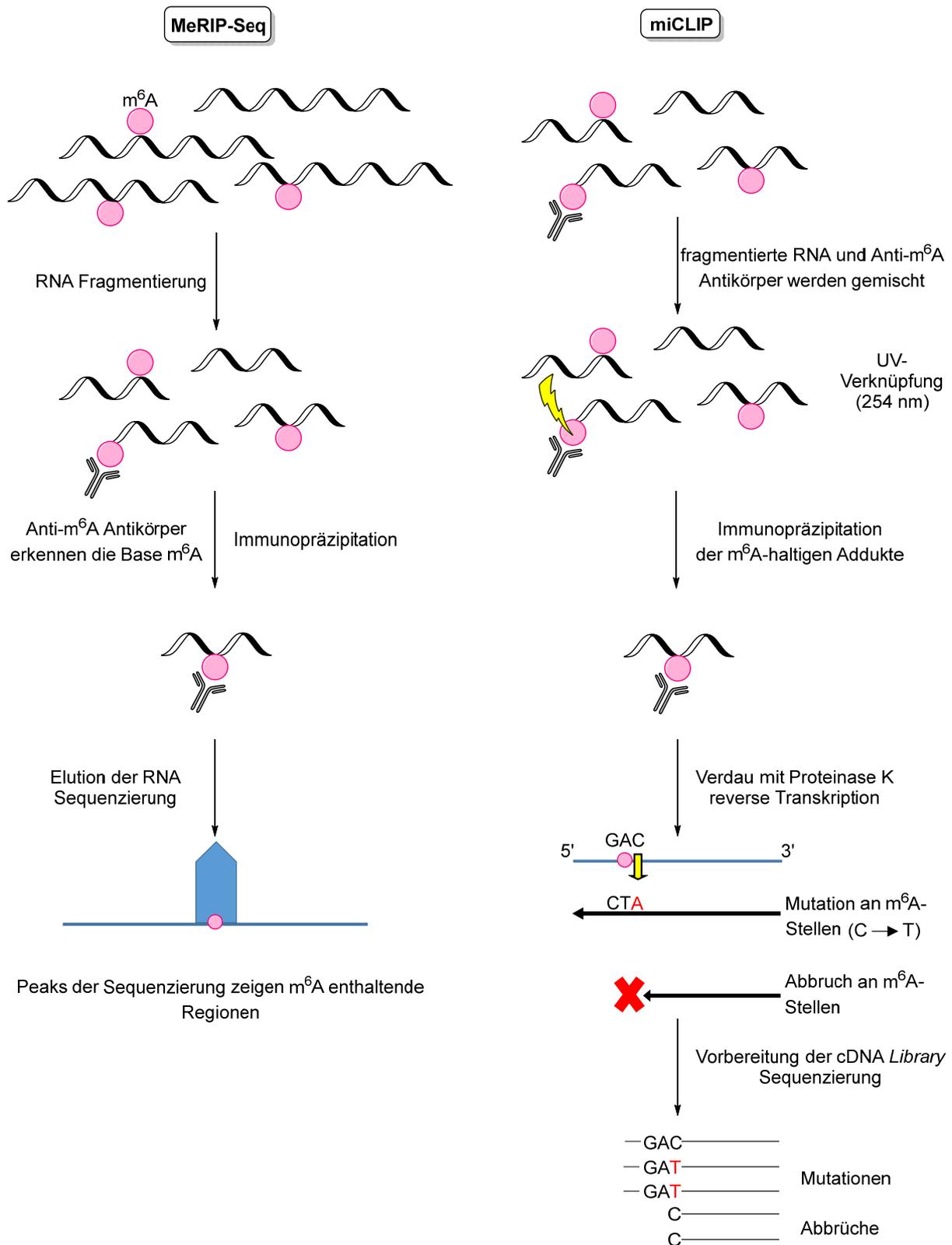


Abbildung 8: Zwei Methoden zur Bestimmung der m⁶A Positionen im Transkriptom: Beide Techniken benötigen fragmentierte zelluläre RNA und werden mit einem spezifischen 6-Methyladenin Antikörper inkubiert. In der MeRIP-Seq Technik benutzt man magnetische antikörperbindende *Beads* und zieht so die methylierten RNA-Moleküle aus der fragmentierten RNA. Die methylierten RNAs werden eluiert und sequenziert. In der miCLIP Methode werden die 6-Methyladenin Antikörper mit der methylierten RNA durch UV-Licht verknüpft und erst dann immunopräzipitiert.

Anschließend wird die methylierte RNA mit Proteinase K eluiert. Dadurch bleibt ein Addukt an der methylierten RNA zurück, welches bei der reversen Transkription entweder zu einem Abbruch oder einer Mutation in direkter Nachbarschaft zur methylierten Stelle führt. Im Anschluss wird die entstandene cDNA vervielfältigt und durch *next-generation sequencing* prozessiert. Stellen, an denen entweder m⁶A oder m⁶Am vorhanden sind, werden durch Detektion von Mutationen oder Abbrüchen bei der Sequenzierung ausgelesen. (Abbildung nach Meyer et. al.)^[131]

Im Folgenden wird nun der Metabolismus der Methylgruppe an der N⁶-Position beschrieben. Wenn davon ausgegangen wird, dass es sich bei RNA Modifikationen wie bei der DNA Methylierung und den Histon-Modifikationen um epigenetische Markierungen handelt, so könnte man auch hier die Bezeichnungen *Writer*, *Reader* und *Eraser* übernehmen. *Writer* katalysieren eine chemische Reaktion und installieren eine chemische Modifikation. Spezielle *Reader* Proteine können die modifizierten Stellen erkennen und in ein Signal verwandeln, während die *Eraser* für die Entfernung dieser Modifikationen zuständig sind.^[132]

Writer

Die Methylgruppe wird durch einen sogenannten m⁶A *Writer* Komplex eingeführt. Wie in der Abbildung 9 gezeigt ist, wird sie an der N⁶-Position durch einen Multikomponenten-Komplex eingeführt. Dieser besteht aus einem Heterodimer der *methyltransferase like 3* (METTL3) und *methyltransferase like 14* (METTL14) und einer dritten, essentiellen Komponente *WT1 associated protein* (WTAP). Eine Deletion von METTL3 führt zu einem fast kompletten Verlust an m⁶A in mRNA, weshalb es das Enzym ist, das m⁶A in mRNA hauptsächlich generiert.^[133, 134] Durch Proteomik Analysen wurden bekannt, dass METTL14 ein Interaktionspartner von METTL3 ist und dass es essentiell für die Methylierung ist.^[135-137] Einige Zeit galt METTL14 ebenfalls als Enzym mit Methyltransferaseaktivität.^[135] Kristallstrukturen zeigten aber, dass METTL3 das einzige Protein mit einer S-Adenosylmethionin (SAM) Bindungsstelle ist und diese bei METTL14 fehlt. Nichtsdestotrotz spielt es eine wichtige Rolle in der Methylierungsreaktion. METTL14 bindet die RNA, interagiert mit METTL3 und steigert dessen Aktivität, indem es die Methylgruppe für den folgenden Transfer auf ein Adenosin richtig positioniert.^[138-140] Die dritte Komponente WTAP ist ansonsten im Rahmen des Spleiß-Prozesses der mRNA involviert.^[141] Aber Knockdown-Experimente von WTAP zeigten eine signifikante Erniedrigung der m⁶A Mengen, obwohl WTAP selbst keine methylierende Aktivität zugeschrieben wird, da es keine konservierte katalytische Methylierungsdomäne aufweist.^[132, 136, 137, 142] Mittlerweile nimmt man an, dass WTAP die Lokalisierung des METTL3-METTL14 Komplexes in den sogenannten *Nuclear Speckles* koordiniert. ^[136, 143] Erst kürzlich konnten zwei weitere Proteine, *RNA-binding Motif* RBM15 und RBM15B, des Methyltransferase-Komplexes durch eine Proteomik Studie identifiziert werden.^[144, 145] Diese Proteine interagieren mit METTL3 in Abhängigkeit von WTAP. Ein Knockdown von RBM15 und RBM15B führte zu einem deutlich reduzierten Vorkommen von m⁶A und legt so die Vermutung

nahe, dass es sich dabei um funktionelle Komponenten des Methyltransferase-Komplexes handelt.^[145] Auch, wenn immer klarer wird, welche Proteine Teil des m⁶A *Writer* Komplex sind, so stellt sich noch immer die Frage, wie bestimmte Adenosine ausgewählt werden, um anschließend methyliert zu werden. Hierzu zeigten *Patil et. al.* interessante iCLIP Ergebnisse, die deutlich machen, dass RBM15 und RBM15B an Uridin-reiche Regionen in der Nachbarschaft von m⁶A Positionen binden. Anschließend rekrutieren sie dort den WTAP/METTL3/METTL14 Komplex. Da sich aber nicht alle m⁶A Bindemotive neben U-reichen Sequenzen befinden, wird durch RBM15 und RBM15B nur ein Teil der zellulären RNAs für die Methylierung ausgewählt.^[145] Über den weiteren Faktor *vir-like m⁶A methyltransferase associated* (KIAA1429) ist bisher nur bekannt, dass ein Mangel an KIAA1429 zu einer drastischen Verringerung an m⁶A führt.^[135-137, 146] Die genaue Funktion von KIAA1429 im Methylierungskomplex ist bis dato noch unbekannt.

Reader

Die Hauptfunktion von m⁶A in mRNA ist das Binden von spezifischen Proteinen. Die bisher bekannten *Reader* Proteine sind Proteine, die eine YTH (YTH521-B Homologie) Domäne aufweisen, und Translations-Initiationsfaktoren.

Das Genom eines Säugetiers weist fünf YTH-Proteine auf: YTHDC1 (DC1), YTHDC2 (DC2), YTHDF1 (DF1), YTHDF2 (DF2) und YTHDF3 (DF3).^[147] Aufgrund ihrer jeweiligen Sequenzen kann man diese Proteine in drei verschiedene Klassen einteilen. DC1, DC2 und die DF-Proteine bilden jeweils eine Klasse. Studien von *Patil et. al.* deuten darauf hin, dass die DF-Proteine und DC1 unter physiologischen Bedingungen m⁶A *Reader* sind. Hierbei bindet die Familie der DF-Proteine hauptsächlich m⁶A-Stellen in der mRNA, während DC1 nur manche m⁶A-Positionen in mRNA, aber auch in nuklearer nicht-kodierender RNA bindet. Es scheint so, als ob die DC1-Bindungsstellen in mRNA nur durch die Anwesenheit von m⁶A in der neu entstandenen mRNA entstehen und es zu einer Bindung kommt, bevor die mRNA diese ins Zytosol exportiert.^[131, 145] Ein weiterer Unterschied zwischen den DF-Proteinen und DC1 ist, dass DC1 vor allem ein Kernprotein ist und somit der hauptsächliche *Reader* von m⁶A im Nukleus ist. Die DF-Proteine hingegen sind vor allem zytoplasmatisch.^[131] Deren Funktion wurde in der letzten Zeit viel diskutiert. DF2 wird eine eigene Funktion zugeschrieben. Es soll die Halbwertszeit m⁶A-haltiger mRNA im Vergleich zu unmodifizierter mRNA herabsetzen.^{[134,}

148]

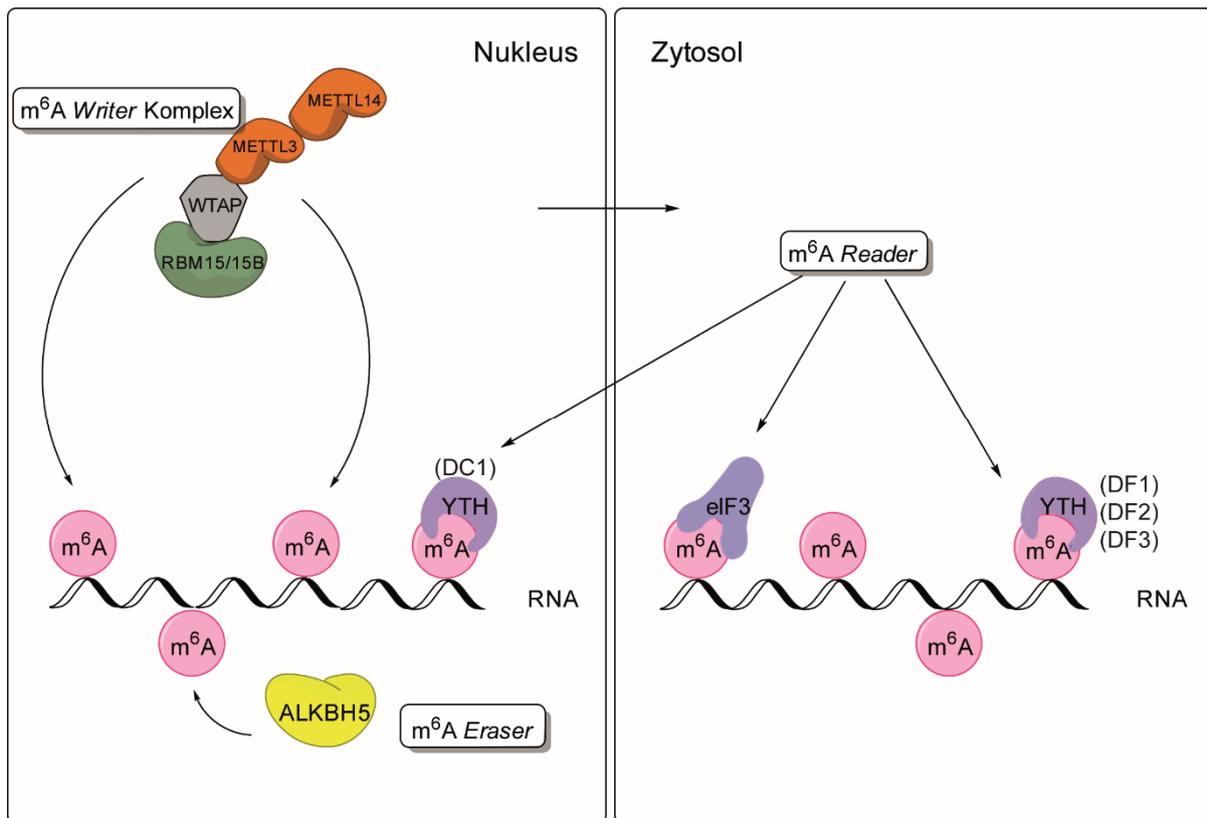


Abbildung 9: Adenine der mRNA können von einem *Writer*-Komplex bestehend aus METTL3, METTL14, WTAP und RBM15/15B im Nukleus zu m⁶A methyliert werden. *Reader* können diese Modifikation binden. Im Nukleus kann m⁶A von dem YTHDC1-Protein gebunden werden, im Zytoplasma kann es hingegen von den YTH-Proteinen DF1, DF2 und DF3 gebunden werden. Außerdem kann m⁶A in der 5'-UTR den eukaryotischen Initiationsfaktor eIF3 binden und so eine eIF4E-unabhängige Translation begünstigen. Auch der *Eraser* ALKBH5, welcher die Methylgruppe entfernen kann, ist im Nukleus positioniert. (Abbildung nach Meyer et. al.)^[131]

Man dachte, dass dieser Effekt durch die Interaktion von DF2 mit den sogenannten *Processing Bodies (P-Bodies)* zustande kommt. Allerdings zeigte sich im Folgenden, dass diese Interaktion nur sporadisch ausgebildet wird. Es könnte sein, dass DF2 immer nur vorübergehend mit den *P-Bodies* interagiert, um so bestimmte mRNAs abzubauen.^[149] *Edupuganti et al.* bestätigten, dass es eine negative Korrelation zwischen mRNA Halbwertszeiten und m⁶A aufgrund von DF2 induziertem mRNA Abbau gibt. Weiterhin zeigte die Gruppe, dass eine positive Korrelation zwischen mRNA Halbwertszeiten und der Bindung von G3BP1 auftritt. Bisher ist noch nicht genau geklärt, inwiefern Proteine, die speziell von m⁶A abgestoßen werden, die Funktionen der mRNA beeinflussen können. Dazu gehört zum Beispiel das Protein G3BP1, welches eine Präferenz für unmodifizierte mRNA gegenüber m⁶A-modifizierter mRNA aufweist und als Stress-Granulum-Protein bekannt ist. Des Weiteren nimmt es eine wichtige Rolle in der Embryonalentwicklung ein.^[150] Mit FMR1 wurde in dieser Studie ein weiterer sequenzabhängiger m⁶A *Reader* gefunden. Dieses Protein bindet an mRNAs, um deren Translation zu verhindern. Ein Fehlen des FMR1 führt zum *Fragiles-X-*

Syndrom,^[151-153] welches sich vor allem in einer kognitiven Behinderung des Betroffenen äußert. Weiterhin wird gezeigt, dass DF1 und FMR1 um die Bindung des m⁶A in mRNA konkurrieren. Dies zeigt, dass es sich um ein komplexes Zusammenspiel der verschiedenen *Reader* und somit der mRNA Homöostase handelt.^[150]

Im Gegensatz zu DF2 soll DF1 die Translation fördern, indem es mit dem eukaryotischen Initiationsfaktor (eIF3) und anderen Translationsinitiationsfaktoren interagiert. Aus diesen Funktionen von DF1 und DF2 wurde ein Modell entwickelt, welches vorschlägt, dass bestimmte m⁶A-haltige mRNAs zuerst durch das Binden von DF1 translatiert werden und dann durch DF2 gebunden und abgebaut werden.^[154] Seit kurzem weiß man, dass DF1, DF2 und DF3 eigentlich identische Bindungen zu jedem m⁶A in mRNA zeigen, was durch die hohe Sequenz-Homologie ihrer YTH-Domäne erklärbar ist. Es scheint, als ob alle DF-Proteine identische Bindungsaffinitäten zu m⁶A-Stellen aufzeigen.^[145] So ist wahrscheinlicher, dass die DF-Proteine gleiche statt unterschiedliche Funktionen haben.^[131] Bevor man wusste, dass DC1 mit m⁶A interagiert, war DC1 als Splicing Regulator bekannt, der das alternative Spleißen in endogenen Transkripten reguliert.^[155] Erst später wurde deutlich, dass DC1 zumindest teilweise für m⁶A-abhängiges Spleißen verantwortlich ist.^[156] Eine andere Studie zeigte hingegen, dass es durch m⁶A zu einer Entfaltung der RNA kommt und so den Zugang für den bekannten Spleiß-Regulator HNRNPC zur RNA erleichtert.^[142] Die gleiche Schwierigkeit tritt bei dem potentiellen *Reader* HNRNPA2B1 auf. Entweder bindet HNRNPA2B1 direkt an m⁶A^[157] oder m⁶A sorgt für die Entfaltung der RNA und ermöglicht so wiederum eine bessere Zugänglichkeit dieser RNA Regionen.^[158, 159]

Bemerkenswert ist die Interaktion von DC1 mit m⁶A-Stellen in *X-inactive specific transcript (XIST)*. *XIST* ist eine nicht-kodierende RNA, die ausschlaggebend für das Stilllegen von Genen auf einem der beiden X-Chromosomen in weiblichen Zellen ist.^[160] *XIST* enthält mindestens 76 m⁶A-modifizierte Stellen. Die Modifikation m⁶A wird für die genstilllegenden Effekte in *XIST* benötigt, denn die Modifikation bindet DC1 an *XIST* und leitet so die transkriptionale Repression ein. Es ist allerdings noch unklar, ob die Funktion direkt auf DC1 zurückzuführen ist, oder ob es sich um einen indirekten Effekt handelt, da bekannt ist, dass das Interaktom von DC1 auch viele epigenetische Regulatoren enthält.^[145] Im Gegensatz zu DC1 sind die Funktionen von DC2 nicht ausreichend verstanden.^[131]

Ein weiterer m⁶A *Reader* ist der eukaryotische Initiationsfaktor (eIF3). Dieser *Reader* wurde gefunden, als versucht wurde, den translationssteigernden Effekt von m⁶A zu verstehen. Nun ist klar, dass ein einziges m⁶A in der 5'-UTR dafür verantwortlich ist, dass es zu Veränderungen der Zusammensetzung der Initiationsfaktoren kommt. mRNAs, die an dieser Stelle m⁶A enthalten, benötigen kein eIF4E (und andere Komponenten, die den eIF4E Komplex ausmachen), um eIF3 für die mRNA zu rekrutieren. Diese mRNAs, die m⁶A an der 5'UTR enthalten, zeigen eine erhöhte Translation.^[161] Eine andere Möglichkeit ist, dass m⁶A-Stellen

in der Nähe des Stop-Kodons DC1 binden und über einen noch unbekanntem Mechanismus eIF3 zur 5'-UTR bringen.^[149] Der eukaryotische Initiationsfaktor eIF3 ist also über zwei verschiedene Wirkmechanismen mit der m⁶A-abhängigen Translationssteigerung in Verbindung zu bringen.^[131]

Generell erscheint es so, als würden die dynamischen m⁶A Mengen einer Zelle als eine Art Schalter wirken.^[150]

Eraser

Bis vor kurzem dachte man, dass es zwei verschiedene *Eraser* für m⁶A gibt: FTO^[162] und ALKBH5^[163]. Seit dem letzten Jahr ist bekannt, dass nicht m⁶A, sondern m⁶Am das Substrat von FTO ist. Somit ist FTO der *Eraser* von m⁶Am und hat dadurch funktionelle Auswirkungen auf die Stabilität der mRNA.^[130] Im Zuge dessen wurde auch untersucht, ob ALKBH5 m⁶Am demethyliert. Es zeigte aber keine Demethylierungsaktivität.^[130] Nichtsdestotrotz ist bisher unklar, wie spezifisch ALKBH5 m⁶A-Stellen demethyliert und wie groß die Bedeutung der aktiven Entfernung der Methylgruppe durch ALKBH5 ist, denn die Abwesenheit von ALKBH5 in Tieren hat zwar negative Auswirkungen auf die Spermatogenese, darüber hinaus sind die Tiere normal.^[131] Dadurch, dass ALKBH5 im Nukleus lokalisiert ist, ist es eher unwahrscheinlich, dass es reife mRNA im Zytoplasma demethylieren kann.^[163] Daraus folgt, dass eine ALKBH5-abhängige Demethylierung von m⁶A nur während der Biogenese der mRNA im Zellkern stattfinden kann. Wie ALKBH5 Knockout-Mäuse gezeigt haben, kann die ALKBH5-abhängige Demethylierung von m⁶A nur geringe Auswirkungen auf den Phänotyp haben. Zukünftige Studien könnten aber noch feine Unterschiede aufzeigen.

2 Aufgabenstellung

In dieser Arbeit werden modifizierte Basen sowohl in der DNA als auch in der RNA untersucht.

Seit der Entdeckung von 5'-Methyl-2'-desoxycytidin (mdC) und dessen höher oxidierten Derivaten 5-Hydroxymethyl-dC (hmdC), 5'-Formyl-dC (fdC) und 5'-Carboxyl-dC (cadC) stellt sich die Frage, wie diese Modifikationen der DNA wieder entfernt werden können. In dieser Arbeit soll nicht auf die Entfernung der höher oxidierten Spezies durch die Basenexzisionsreparatur eingegangen werden. Vielmehr soll die Möglichkeit eines direkten C-C Bindungsbruchs in Form einer Deformylierungs- bzw. Decarboxylierungsreaktion untersucht werden.

Das erste Ziel dieser Arbeit war die Erforschung eines möglichen aktiven Demethylierungsweges mittels C-C Bindungsbruch mit Hilfe von massenspektrometrischen Messmethoden (UHPLC-MS/MS) und synthetischen Testmolekülen. Dieses Problem sollte auf zwei verschiedene Arten erforscht werden: Zum einen durch die direkte Inkorporation von isotopenmarkiertem 5-Formyl-2'-desoxycytidin (fdC) bzw. von dessen 2'-fluoriertem Derivat (F-fdC), zum anderen durch die Verwendung verschiedener isotopenmarkierter dC Spezies um tiefere Einblicke zu erhalten. Im Wesentlichen sollen alle Testmoleküle an verschiedene Zelllinien verfüttert und deren Metabolismus in der genomischen DNA verfolgt werden. Dafür mussten geeignete massenspektrometrische Analysemethoden entwickelt werden. Vor allem die Entwicklung der Messmethode zur Quantifizierung aller fluorierten Nukleoside sollte im Vordergrund stehen. Das Testmolekül F-fdC sollte an verschiedene Zelllinien, z.B. Knockout- oder Inhibitor-behandelte Zelllinien, verfüttert werden. Durch die neu entwickelte Messmethode sollte anschließend der Metabolismus dieses Moleküls verfolgt und die fluorierten Derivate quantifiziert werden. Entdeckt man durch diese Experimente tatsächlich einen C-C Bindungsbruch, so wäre ein weiteres Ziel dieser Arbeit, den Mechanismus einer solchen C-C Bindungsbruchreaktion zu identifizieren.

Im zweiten Teil dieser Arbeit sollte die Funktion und das Vorkommen der Hypermodifikation m^2i^6A untersucht werden. Um dieser Fragestellung nachzugehen, musste auch hierfür eine massenspektrometrische Analysemethode (UHPLC-MS/MS) entwickelt werden. Im Zuge dessen sollte eine generelle Messmethode für verschiedene interessante RNA-Modifikationen entwickelt werden, um in Zukunft auch diese erforschen zu können.

3 Veröffentlichte Arbeiten

3.1 Isotope-dilution mass spectrometry for exact quantification of non-canonical DNA nucleosides

Franziska R. Traube[#], Sarah Schiffers[#], Katharina Iwan[#], Stefanie Kellner, Fabio Spada, Markus Müller and Thomas Carell*

([#] geteilte Erstautorenschaft)

Prolog

Es ist bekannt, dass die DNA nicht nur kanonische Nucleoside enthält, sondern auch eine Vielzahl an modifizierten Nucleobasen. Die Bestimmung der genauen Menge dieser nicht-kanonischen Basen kann zur Charakterisierung des Zustands eines biologischen Systems, wie zur Bestimmung von Krankheiten oder Entwicklungsprozessen, beitragen und ist daher von großer Bedeutung. Da die meisten dieser Nucleoside nur in sehr geringen Mengen vorkommen, bedarf es einer Ausarbeitung präziser Probenvorbereitungsschritte sowie einer sensitiven Messmethode.

In diesem Protokoll wird die Extraktion von DNA durch eine optimierte *Spin Column* Methode und der enzymatische Verdau der DNA beschrieben. Die Verwendung von isotopenmarkierten Standards erlaubt eine exakte Quantifizierung mittels der Isotopenverdünnungstechnik, welche sehr präzise und robuste Werte liefert. Da der limitierende Faktor solcher Analysen meist die Messdauer ist, wurde eine kurze, ca. 14-minütige, UHPLC-MS/MS (Triple Quadrupol) Massenspektrometrie-Methode entwickelt. Durch die zur Verfügung stehenden Excel-Tabellen ist die Berechnung der Mengen der jeweiligen Modifikation im Genom möglich. Das in diesem Protokoll beschriebene Verfahren zur Quantifizierung ermöglicht im Vergleich zu alternativen Methoden eine parallel, schnelle, ultrasensitive und hochreproduzierbare Quantifizierung verschiedener Nucleoside. Inklusiv Probenvorbereitung und Auswertung ist die Quantifizierung von DNA-Modifikationen in weniger als einer Woche möglich.

Autorenbeitrag

Eine detaillierte Auflistung der Autorenbeiträge ist im nachfolgenden Manuskript zu finden.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags

Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides

Franziska R. Traube^{1,2}, Sarah Schiffrers^{1,2}, Katharina Iwan^{1,2}, Stefanie Kellner¹, Fabio Spada¹, Markus Müller¹ and Thomas Carell^{1*}

DNA contains not only canonical nucleotides but also a variety of modifications of the bases. In particular, cytosine and adenine are frequently modified. Determination of the exact quantity of these noncanonical bases can contribute to the characterization of the state of a biological system, e.g., determination of disease or developmental processes, and is therefore extremely important. Here, we present a workflow that includes detailed description of critical sample preparation steps and important aspects of mass spectrometry analysis and validation. In this protocol, extraction and digestion of DNA by an optimized spin-column and enzyme-based method are described. Isotopically labeled standards are added in the course of DNA digestion, which allows exact quantification by isotope dilution mass spectrometry. To overcome the major bottleneck of such analyses, we developed a short (~14-min-per-sample) ultra-HPLC (UHPLC) and triple quadrupole mass spectrometric (QQQ-MS) method. Easy calculation of the modification abundance in the genome is possible with the provided evaluation sheets. Compared to alternative methods, the quantification procedure presented here allows rapid, ultrasensitive (low femtomole range) and highly reproducible quantification of different nucleosides in parallel. Including sample preparation and evaluation, quantification of DNA modifications can be achieved in less than a week.

Introduction

In addition to the canonical nucleotides, both DNA and RNA contain a variety of modifications of the bases. In the DNA of vertebrates, for example, modified cytidines such as 5-methyl-2'-deoxycytidine (m⁵dC), 5-hydroxymethyl-dC (hm⁵dC)^{1,2}, 5-formyl-dC (f⁵dC)^{3,4} and 5-carboxy-dC (ca⁵dC)^{4,5} have been discovered. It is well established that m⁵dC and hm⁵dC are particularly epigenetically relevant^{6,7}. The levels of hm⁵dC are often altered by several orders of magnitude in tumor tissues, and this has been shown to correlate with the aggressiveness of tumors^{8–10}. In stem cells, f⁵dC and ca⁵dC were detected at substantial levels^{3,4} and their abundance changes during differentiation^{11,12}. Both f⁵dC¹³ and ca⁵dC are thought to be involved in a process of active demethylation. Whether f⁵dC has additional distinct epigenetic functions is unclear^{14–18} and, for ca⁵dC, no such function has yet been found. Neither f⁵dC nor ca⁵dC has yet been explored in regard to potential level changes in response to disease states. In bacteria, two major modifications are N⁴-methyl-dC (m⁴dC)¹⁹ and N⁶-methyl-2'-deoxyadenosine (m⁶dA)²⁰. The latter is a well-studied modification with extreme importance in host defense mechanisms²¹. m⁴dC has not been found in vertebrate DNA, and conflicting results have been found for research into the presence of m⁶dA in vertebrate DNA^{22–26}.

In addition to these noncanonical bases that are actively generated for partially unknown purposes, genomic DNA (gDNA) contains modified bases that are generated as DNA lesions. In particular, oxidative DNA lesions such as 8-oxo-7,8-dihydro-deoxyguanosine (8oxodG) can easily form²⁷, and, again, the levels of such base lesions can correlate with diseases²⁸. Measurement of 8oxodG has proven to be particularly difficult, because its amounts can easily be overestimated because of additional oxidation during sample preparation or when bringing it into the gas phase in the mass spectrometer (gas chromatography MS)²⁹.

Determination of the exact amount of all noncanonical DNA nucleosides (epigenetic nucleosides or DNA lesions) is consequently an important task and requires a fast, ultrasensitive and highly reproducible approach.

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

²These authors contributed equally: Franziska R. Traube, Sarah Schiffrers, Katharina Iwan. *e-mail: thomas.carell@lmu.de

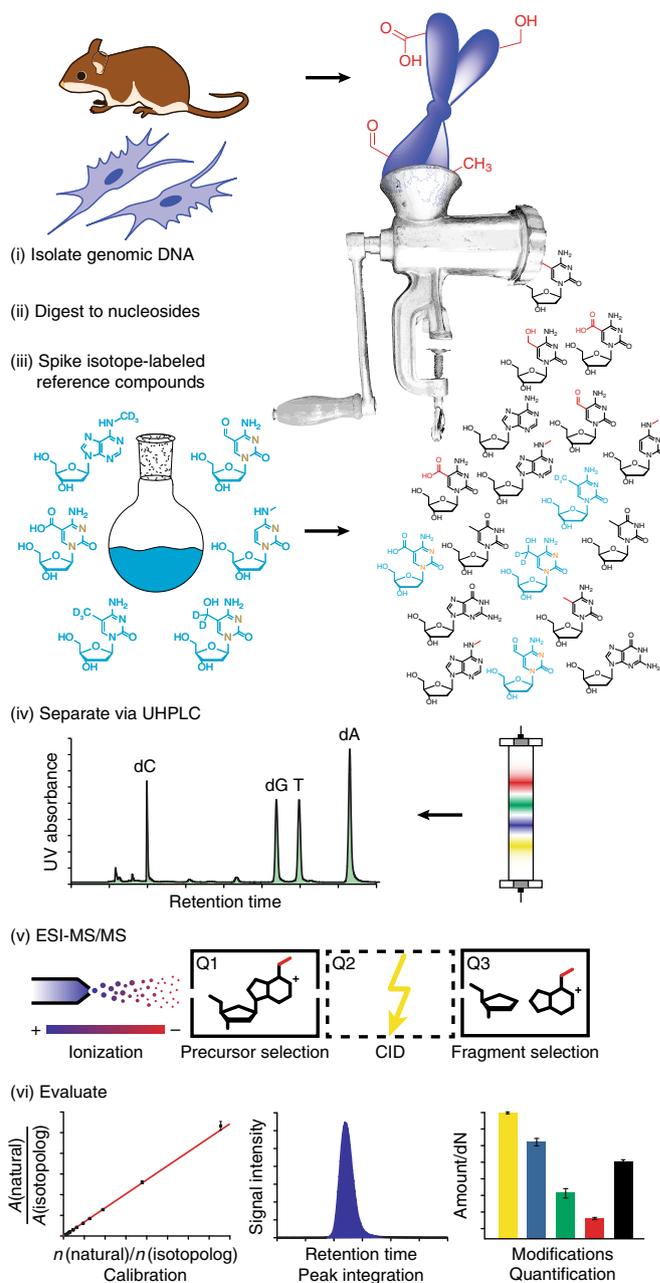


Fig. 1 | Procedure overview. General workflow for rapid DNA extraction, digestion and ultrasensitive quantification of the obtained nucleosides via UHPLC-*QQQ*-MS. T, deoxythymidine.

Here, we present a detailed workflow for the quantification of noncanonical nucleosides. It includes DNA extraction from cell culture or tissues, total enzymatic digestion of the DNA, and ultrasensitive quantification of the obtained nucleosides via UHPLC and *QQQ*-MS (UHPLC-*QQQ*-MS) (Fig. 1). The major advantage of UHPLC-*QQQ*-MS is its capability of chemically determining the true identity of the modified nucleoside and its exact abundance in a relatively short time frame. Furthermore, the application of the multiple reaction monitoring (MRM) mode enables simultaneous fragmentation of different precursor ions into one or multiple fragment ions and their subsequent detection³⁰. The fundament of the method is the isotope dilution technique, in which stable isotopologs of the nucleosides of interest are added to the specimen (spiking) as internal standards. This analytic approach results in highly reproducible parallel quantification of modified DNA nucleosides in only ~14 min per sample and makes this method the ideal choice when sequence resolution is not required and sufficient sample material is available.

Development of the protocol

Quantification of very abundant DNA modifications, such as m^5dC in vertebrates, has been possible for decades without the need for ultrasensitive quantification methods³¹. Owing to their abundance, small quantification errors do not affect interpretation of the data. However, accurate quantification of DNA nucleosides can be challenging when the modification is very rare. Contaminations at the cell culture level or during tissue dissection are a particular problem if they go unnoticed. These contaminations include bacteria and fungi, which often carry DNA modifications that might be rare or not present in mammals and can have a substantial effect on the resulting values. Therefore, the biological material that is investigated must be carefully tested for the presence of such contaminations. With respect to the investigation of DNA lesions, it must be noted that lesions are also formed during DNA isolation, for example, by releasing reactive oxygen species in the course of cell lysis. Most problematic are unspecific oxidations and deaminations. To overcome these problems, addition of butylated hydroxytoluene (BHT) and deferoxamine (DFOA) during isolation is essential. BHT is a radical scavenger that prevents DNA from being oxidized, and DFOA inhibits deaminases during the isolation process^{11,32}.

The biggest challenge of exact quantification is the generation of correct calibration curves. Mass spectrometry is a highly sensitive, but inherently not quantitative, method. The signal intensity reflects not only the amount of a molecule of interest (MOI) in the specimen, but also its ionizability, which is, in complex mixtures, strongly affected by other co-eluting molecules. To correlate the signal intensity with the exact amount of the MOI, calibration curves are essential. In principle, there are two different procedures for quantifying the amount of an MOI. One can use either external or internal calibration. Both approaches depend on the availability of the MOI in weighable quantities from, e.g., chemical synthesis. For external calibration, a serial dilution of the synthetic MOI is measured before analysis of the samples. Here, it is important to use the same buffer and the same LC column, and to perform the calibration on the same day, ideally once before the sample measurement, once after the sample set and, in the case of many samples, once during the sample worklist. This is important in order to counteract differences in instrument performance. The advantage of the external method is that it requires only the pure MOI as the standard. However, it also has many disadvantages. For best reproducibility, at least five different dilutions are measured as technical triplicates, which increases the measurement time substantially. Although this extensive calibration effort reduces errors from instrument performance fluctuations, it cannot overcome quantification errors stemming from matrix effects.

Matrix effects are mainly caused by salts, solvents and other undefined components within the sample³³. It must be noted that a biological sample, e.g., gDNA, is much more complex than a solution of pure, synthetically generated nucleosides and that this complexity impacts the measurement. The matrix of the samples or the pressure on the column may furthermore lead to shifts in retention time (RT) of the MOI, so that a reliable statement on identification of the MOI may be infeasible as well. The sample matrix may also affect the signal intensity of the MOI: certain nucleosides suffer from ion suppression, which occurs when high amounts of interfering ions are present that may co-elute with the MOI and therefore compete for charge and space in the mass spectrometry device^{33,34}. Taken together, it is not predictable whether a certain MOI can be reliably quantified using external calibration. Therefore, the gold standard in MS quantification is internal calibration, whereby a specific amount (n^*) of a reference molecule is added to the sample and used as an internal standard at all times.

Our group utilizes the isotope dilution technique, in which the reference compound is an isotopolog of the MOI (designated MOI*)^{13,26,35–44}. This molecule has very similar chromatographic and fragmentation properties as compared to the MOI, and as such it elutes ideally with an identical RT. The isotopes of choice should be ¹³carbon (¹³C) and ¹⁵nitrogen (¹⁵N), because no RT shifting is observed with ¹²C/¹³C and ¹⁴N/¹⁵N substitutions. By contrast, replacing hydrogen (H) atoms with deuterium (D) atoms in the MOI* affects the physicochemical properties of the molecules^{45,46}. This leads to an observable shift in RTs, and, more importantly, the acid–base properties and thus the ionizing efficiency are affected in the MOI* as compared to the MOI.

The chosen MOI* is then used as an internal standard to reference and identify the correct peak of the MOI in the chromatogram. As such, it is essential to achieve accurate quantification. Calibration curves based on internal standards have the advantage that they can be used for multiple measurements on different days, because they are independent of the instrument performance on a specific occasion.

Applications of the method

The described protocol for DNA isolation is optimized for cell culture and vertebrate tissues. Isolation of DNA from other organisms, especially from those consisting of cells with cell walls, require harsher conditions. Once the DNA is isolated, our protocol can be applied to any kind of sample and the robust method reported here provides reliable quantification data. In addition, the method is highly sensitive, which enables quantification of nucleosides in the low femtomole and even attomole range, as described in the following section. Thereby, depending on the amount of DNA available for the measurements and the genome size of the organism of interest, even the determination of nucleosides that are very rare (a couple hundred nucleosides per genome) is possible. With this specificity, one can reliably demonstrate the presence of a modification in a genome²⁶ and determine even small changes of its abundance in response to disease progression⁴³. Furthermore, the effects of stress factors, as well as those of cellular differentiation and mutation/knock-down of involved proteins, can be measured. Such biological conditions influence the biological pathways that lead to the formation or removal of noncanonical nucleosides in the genome^{13,44}.

Comparison with other methods

The workflow described here involves isolation and total digestion of gDNA with subsequent (parallel) analysis and quantification of nucleosides of interest, even in complex mixtures. This method provides robust and highly reproducible data in a fast manner due to the utilization of stable isotopologs as internal standards. The analysis time for parallel quantification of nucleosides routinely takes <15 min.

DNA digestion can be performed in a variety of ways, depending on the research focus. The method described here is optimized toward native DNA modifications, and other protocols that require harsh conditions or reactive substances will not be discussed in detail in this section. A comparative summary is nevertheless provided in Supplementary Table 1. For the optimized digestion of DNA adducts, see Liu et al.⁴⁷.

Antibody-based methods, such as dot blots, often represent the method of choice for nucleoside analysis when no detector-based system is available. These methods always depend on the specificity and binding affinity of the antibody. Antibodies have the major disadvantage that they often show cross reactivity, particularly between RNA and DNA bases²⁴, and it was recently shown that IgG antibodies have an intrinsic affinity for short DNA repeats⁴⁸. This might obscure the obtained data. Our method is a combination of two different analytical techniques, (i) separation of analytes with an LC device and (ii) detection of these analytes with a detector, and it can distinguish between DNA and RNA nucleosides. In general, LC can be accomplished by applying HPLC or UHPLC. The following detection of the eluting MOI is achieved by using a UV detector, a fluorescence detector (FLD), an electrochemical detector or an MS device.

Chromatographic separation

Separation of the nucleosides of interest can be done by either reversed-phase HPLC or UHPLC. Both methods can apply either isocratic or binary gradient elution. Binary gradients usually shorten the analysis time, and the resolution is similar to that for isocratic elution. Gradient elution requires more regular maintenance, and the accuracy of the peak area and peak height is often questionable, which could impede reliable quantification^{49,50}. This does not apply to our isotope dilution method, as a variation in size or shape of the peak will also affect the corresponding isotopologs and is therefore taken into account. On the chromatographic side, the generally high polarity of nucleosides requires the use of reversed-phase C8 or C18 columns to achieve good separation. Recently, hydrophilic interaction chromatography (HILIC) columns were alternatively used for separation of nucleosides⁵¹. The most recent developments regarding the high-throughput analysis of nucleosides is the application of UHPLC. The UHPLC columns have smaller particle size (typically $\leq 2 \mu\text{m}$, in comparison with $>4 \mu\text{m}$ for HPLC), and they are shorter and have a smaller inner diameter (50–150 mm length, $<2.5\text{-mm}$ diameter for UHPLC as compared to 150–250 mm length, 4.6-mm diameter for HPLC). As such, they provide superior separation in a short time (~ 15 min for UHPLC as compared to ~ 45 min for HPLC). Most importantly, these columns give extremely narrow and sharp peaks that are essential to achieving the highest possible sensitivity. In comparison to ordinary HPLC, this UHPLC technique has lower solvent consumption due to low flow rates <0.5 mL/min. With less solvent, analytes are also more concentrated after elution and can be more easily vaporized and ionized, which is a major advantage for the subsequent MS measurement. In principle, the injection volumes should be in the

range between 1 and 10 μL ; otherwise, this will impair efficient separation. As a rule of thumb, 10% of the flow rate should be used as the maximum injection volume. In our system, the flow rate is 0.35 mL/min, which allows for a maximum injection volume of 35 μL . However, the presented system was found to accept injection volumes up to 40 μL without loss of chromatographic resolution.

Detectors

Previously established methods for detection use, for example, the combination of reversed-phase HPLC and a UV detector. A UV detector is often the detector of choice, because the resulting peak area, and therefore the amount of a nucleoside, depends only on its extinction coefficient. UV-based quantification entails no matrix effects. Determination of the amount of each nucleoside can therefore be reliably accomplished by applying external calibration curves. One disadvantage is that baseline separation of the nucleosides of interest is critical. Moreover, UV detection suffers from lower sensitivity as compared to the MS- and fluorescence-based methods, which is reflected in the limits of quantification (here, lower limit of quantification (LLOQ)) of ~ 0.3 – 1.4 pmol. Correspondingly, the limits of detection (LOD) are typically not <0.08 – 0.42 pmol per nucleoside⁵². The absolute numbers depend on the type of nucleoside.

Higher sensitivity can be achieved when an LC device is coupled to an FLD. This combination results, for example, in an LOD for deoxyguanosine (dG) of 0.24 pmol⁵³. Because most nucleosides do not show strong enough autofluorescence, however, they must be derivatized for fluorescence detection. For example, phenylglyoxal can be used to modify guanine-containing compounds⁵³. $m^5\text{dC}$ can be converted into the corresponding fluorescent 3, N^4 -etheno-5-methyl-2'-deoxycytidine ($\epsilon m^5\text{C}$). These compounds can then be quantified by reversed-phase HPLC-FLD with a typical LOD of 0.02 pmol⁵⁴. One major disadvantage of this method is that derivatization chemistry is never quantitative (never 100% efficient). This leads to an underestimation of the abundance of the MOI and may distort the resulting data.

For detection and quantification of a rare MOI, an MS detector with higher sensitivity compared to UV and FLD is required. To analyze the MOI, it must be brought into the gas phase and be ionized so that it can enter the vacuum of the mass spectrometer. This critical step is achieved in the ion source, and it is the first critical parameter that must be optimized for high sensitivity. Next, efficient fragmentation and detection of the specific fragments must be achieved. Our MS device is equipped with an ESI source, which is perfectly suited to the evaporation and ionization of nucleosides and nucleotides. ESI is considered to be a mild ionization source that applies relatively low energy. This results in less in-source fragmentation of the MOI in comparison to those of electron ionization and chemical ionization, and therefore less loss of signal intensity due to a broad distribution of fragment ions⁵⁵. Furthermore, this technique provides the highest possible flow rate to the following interface, e.g., the mass analyzer⁵⁶. This is essential for maximal sensitivity.

We use QQQ mass spectrometers for our analyses because they have the highest sensitivity among the available mass analyzers. Time-of-flight (ToF) analyzers⁵⁷, ion-trap analyzers^{58,59} or hybrid analyzers⁶⁰, which are a combination of different mass analyzers, can also be used in this protocol. Compared to ToF and ion-trap mass detectors, the QQQ detector has high sensitivity but low resolution. This means a QQQ mass spectrometer cannot distinguish between ions that differ in <1 AMU. However, high resolution is not necessary for quantification of known compounds for which synthetic standards are available.

The greatest advantage of the QQQ mass detector is the possibility of monitoring multiple mass/nucleoside signals in the MRM mode, e.g., examination of fragmentation of the precursor ion and observation of the resulting product ion simultaneously, within a fraction of a second. The first quadrupole can be programmed to rapidly switch between monitoring molecules of various specific m/z values. The selected MOI ions are then fragmented in the second quadrupole, which represents the collision cell. Subsequently, the third quadrupole can be stepped to different m/z values to identify a specific fragment ion generated from the MOI in the collision cell. The detector ascertains the analytes, and the signal is enhanced by an electron multiplier. This technique improves the detection limits of analytes because only the m/z values from the molecules of interest are recorded, instead of scanning across the whole mass spectrum⁵⁵.

The method presented here is a combination of a UHPLC device coupled to a UV detector and a QQQ analyzer. The less sensitive UV detector is used to quantify the highly abundant canonical nucleosides, whereas the quantification of the less abundant noncanonical nucleosides is achieved by the QQQ detector. Applying this method leads to LLOQs in the femtomole range or lower for nucleosides of interest and enables the determination of their quantities in the genome.

Experimental design

Our protocol covers all steps to establish the whole procedure from the beginning, when neither expertise nor the material for quantification of DNA nucleosides is available. For development of a quantification method, several machine parameters must be determined, apart from the *m/z* values for the precursor and product ion of the MOIs and MOI*s.

The Procedure provides detailed steps for two methods:

- (i) Quantification of the so-called epigenetic DNA modifications (m⁵dC, hm⁵dC, f⁵dC and ca⁵dC), as well as a common DNA lesion (8oxodG), with highest sensitivity toward the less abundant modifications. This method uses a C8 column and a water/MeCN solvent system (Supplementary Tables 2 and 3).
- (ii) Quantification of further DNA modifications (m⁴dC and m⁶dA) known from bacteria, in which they are highly abundant. For investigation of their existence in higher organisms, in which their abundance is expected to be low, the use of a C18 column enables the detection and distinction of m⁴dC and m⁵dC. Furthermore, applying a water/MeOH solvent system provides the highest sensitivity for the detection of m⁶dA. This method is used in two variations with adjusted collision energies (CEs) (sensitive and insensitive) for m⁶dA to enable quantification of the modification in organisms with low or high abundance, respectively, without reaching limitations for other important components of quantification, e.g., the LLOQ of canonical nucleosides (Supplementary Tables 4 and 5).

A compilation of linear equations is given in Supplementary Tables 2 and 4; these can be applied only to the respective solvent systems and columns (in Supplementary Table 2, water/MeCN on a C8 column; in Supplementary Table 4, water/MeOH on a C18 column). For m⁴dC, two linear equations are given because they span two ranges (big = values in the pmol range; small = values in the low fmol range), and for m⁶dA, two linear equations are shown for the different CEs (sensitive and insensitive), thereby resulting in two respective ranges.

If you want to measure all the nucleosides given in Supplementary Tables 2 and 4 at once, new calibration curves are necessary for the modifications with the different solvent system. In this case, we recommend sticking to the water/MeOH system with a C18 column.

If identical equipment and methods are used, it is possible to use the calibration curves provided in Supplementary Tables 2 and 4.

Internal calibration

We recommend internal calibration with a stable isotopolog MOI* as a reference for the MOI. This MOI* is spiked into the DNA sample. If several MOIs must be quantified in parallel, a mixture of the required MOIs* is spiked into the sample in the form of a spiking mixture. This mixture contains not only the needed MOIs* but also the digestion enzymes, and it is added to all samples of one experiment. At best, the MOI* should be at least two atom units heavier than the MOI to allow a clear MS-based differentiation between the MOI* and the naturally occurring MOI. Natural ¹³C has an abundance of 1%. Thus, any molecule with 10 C atoms, such as m⁵dC, will have a 10% chance to carry at least one ¹³C atom and a 1% chance to carry two ¹³C atoms. The presence of these natural [+1] and [+2] isotopologs limits the sensitivity, especially for highly abundant MOIs. Therefore, a mass difference between the MOI and the MOI* of three or more units is recommended for high fidelity.

The mass spectrometer is set to monitor the nucleoside signals of the MOI and the MOI*, which will be displayed in two separate chromatograms: one for the MOI signal and one for the MOI* signal. Integration of the peak areas for MOI and MOI* gives areas *A* and *A**. To calculate the amount of the MOI from *A* and *A**, calibration curves are required. These calibration curves are generated before the measurement by mixing the MOI and its MOI* at different ratios, in which the amount of the MOI (*n*) is stepwise diluted by a factor of two, while the amount of MOI* (*n**) is kept constant.

The ratios of *A/A** and *n/n** correlate in a linear manner. This allows the use of the following linear equation:

$$\frac{A}{A^*} = m \times \frac{n[\text{pmol}]}{n^*[\text{pmol}]} + t.$$

For the determination of the slope *m* and the *y*-intercept *t* of the calibration curve, the aforementioned ratios for *n/n** are measured in technical triplicates, and the resulting ratio of *A/A** is plotted against *n/n**. The standard deviation for each data point must not exceed 20%, and the accuracy must be between 80 and 120%.

The lowest amount that meets the requirements for the standard deviation and accuracy determines the LLOQ. The same criteria regarding standard deviation and accuracy apply for the highest amount, which is therefore defined as the upper limit of quantification.

Once these prerequisites are met, the specific amount of the MOI (n) can be calculated using the following equation:

$$n[\text{pmol}] = \frac{A - t}{m} \times n^*[\text{pmol}].$$

If the calibration curve meets those requirements, it is considered reliable, independent of the machine performance, as long as the components are the same. The specific parameters for each compound, e.g., collision energy and cell accelerator voltage, must be set identically for the MOI and the MOI*. This saves measurement time and secures high fidelity. In addition, RT shifts can be accepted if the internal standard shows the same shift.

However, if the MOI and the MOI* do not co-elute, it must be assumed that one of the signals is not caused by either the MOI or MOI*. In particular, in complex mixtures, it may be generated by an unknown analyte that shows the same mass signal. In general, RT differences that are, for example, caused by the isotope difference between the MOI and the MOI*, are tolerated when they fall within a 2.5% limit. In particular, when deuterium is used as a source of labeling, slightly shifted RTs are common because of diverging binding strengths. Still, the MOI and MOI* must meet the above criterion. If the deviation is >2.5%, the apparent peak is not the peak of the MOI⁶¹. Whether the 2.5% limit is exceeded can be calculated with the following equation:

$$\frac{|RT^{\text{MOI}}[\text{min}] - RT^{\text{MOI}^*}[\text{min}]|}{RT^{\text{MOI}^*}[\text{min}]} \times 100\% \stackrel{!}{\leq} 2.5\%.$$

The limit of detection

The LLOQ does not equal the lower LOD, which is important for evaluation of the data. Even if the LLOQ criterion is fulfilled, the sample data may not meet the LOD criterion when the signal-to-noise ratio (SNR) is too low. The LOD_{theoretical} is set at an SNR of 3 for the pure MOI in water. But analyzing MOIs in biological samples demands harsher constraints because of matrix effects. Therefore, the LOD_{practical} is defined as the SNR of 3 of the MOI in a biological sample. To determine the LOD_{practical} in a biological sample, we always evaluate so-called digestion blanks (in our protocol, samples 1-3) for each measurement. The digestion blanks contain the spiking mixture, but no DNA. For the digestion blanks, we analyze A (in this case, ideally A = 0) at the respective RTs of each MOI and A* of each MOI* of the spiking mixture. Therefore, the LOD_{practical} relates to the background signal of the spiking mixture in the digestion blanks. It can be calculated in two different ways:

- (i) One can calculate the LOD_{practical} as the mean of the ratio A/A* for the digestion blanks multiplied by the factor of 3.
- (ii) One can calculate the LOD_{practical} as the mean of A for the digestion blanks multiplied by the factor of 3.

Both definitions can be used for further data evaluation. Method (ii) is even stricter than method (i). However, it is important to choose one method consistently. Applying this LOD_{practical} value will prevent the emergence of false-positive data and is therefore highly recommended. Only the ratios of A/A* (method (i)) or the areas A (method (ii)) of the subsequent samples that exceed this LOD_{practical} represent true peaks. Ratios of A/A* that exceed the LLOQ determined by the calibration curve, but not the LOD_{practical}, might be false-positive hits and therefore have to be excluded from further analysis. To overcome this problem in future experiments, we recommend the use of more DNA.

Calibration curves

A calibration curve spans a specific concentration range of at least five data points, in which the curve is linear. The limitations of this so-called dynamic range become problematic if an experiment contains samples in which the abundance of the MOI varies by several orders of magnitude. In this case, a single calibration curve might be insufficient if it does not cover the resulting intensities of signals. Then a second calibration curve with a different range is needed. The two resulting calibration curves will ideally cover the lowest and highest concentrations of the MOI in the specimen. To achieve consistent quantification results across multiple biological samples, it is important to use a single calibration curve for the whole dataset. It may, for example, occur that a knock-out cell line

cannot be quantified with the same calibration curve as its corresponding wild type, because the amount of the MOI in the latter exceeds the values in the knockout by a factor of ten or more. For this biological question, a combination of different calibration curves might seem a valid solution, but it is not recommended. It is better to adjust the amount of DNA by dilution, so that the total amount of the MOI in both samples falls within the same range and hence a single calibration curve can be applied. It is also important that the concentrations of the spiked-in MOI*s are in the range of the expected concentrations of the MOI. If dilution of the DNA is not applicable, one can prepare two different spiking mixtures to meet the requirements of a single calibration curve; one with a higher amount of MOI*, which is added to the biological samples when a higher amount of MOI is expected and one with a lower amount of MOI* for the corresponding sample with a low abundance of MOI. In the subsequent evaluation of the data, the respective amount of MOI* (n^*) must be adjusted.

For handling samples with unknown and potentially strongly deviating MOI content, we recommend generating a calibration curve, in which you start with an amount n that is more than four times greater than n^* and dilute this amount n by a factor of three instead of two. We used this procedure successfully in some cases, but for certain modifications, the smaller dilution factor resulted in a better calibration curve.

Evaluation of UHPLC-MS/MS data

The provided Excel sheet (Supplementary Methods 1 and 2) includes all functions that are needed for evaluating the data. It allows calculation of the ratio of A/A^* and thereby the unknown amount n of the MOI with the following equation, where m represents the slope and t is the y -intercept. This equation is unique for each MOI and only valid within its concentration ranges.

$$\frac{A}{A^*} = m \times \frac{n[\text{pmol}]}{n^*[\text{pmol}]} + t,$$

$$n[\text{pmol}] = \frac{\frac{A}{A^*} - t}{m} \times n^*[\text{pmol}].$$

Note that in this protocol an injection volume of 39 μL is applied. This was chosen because it is the largest amount that can be injected without a loss of chromatographic resolution. The amount of the MOI (pmol/sample), e.g., in 50 μL of digestion mixture, is calculated by multiplication of the upper equation by (50/39 μL). The complete calculation is described in the following equation:

$$n[\text{pmol}] = \frac{\frac{A}{A^*} - t}{m} \times n^*[\text{pmol}] \times \frac{50 \mu\text{L}}{39 \mu\text{L}}.$$

If <39 μL is injected, this value must be adjusted in the provided Excel sheet (Supplementary Methods 1 and 2). This additional calculation will provide comparable data for your technical and later biological replicates. The injection volume is automatically recorded in the QQQ quantitative analysis program.

In the case that the GC content of the DNA is known, division of the amount of MOI (in pmol) by the amount of dG (in pmol) will provide the term ‘MOI per dG’. For example, for mouse gDNA, further multiplication by 0.21 (adjusting for the 42% GC value of mouse gDNA⁶²) will yield the term ‘MOI per dN’, which is mainly used for presenting quantification data.

$$\frac{n(\text{MOI})[\text{pmol}]}{n(\text{dG})[\text{pmol}]} = \text{MOI/dG},$$

$$\text{MOI/dG} \times 0.21 = \text{MOI/dN}.$$

In the case that the GC content is unknown, the amounts of dA, dC, dG and thymidine (T) must be determined using UV detection, and the amount of MOI divided by the amount of $\Sigma(\text{dA}, \text{dC}, \text{dG}, \text{T})$ will directly provide the term ‘MOI per dN’.

If one is interested in the derivatives of dC, namely $m^5\text{dC}$, $hm^5\text{dC}$, $f^5\text{dC}$ and $ca^5\text{dC}$, it is important to perform additional calculations because these modifications are measured in the MS mode, whereas canonical dC is measured in the UV mode. The combination of UV- and MS-derived data often results in summed values for the total amounts of all dC derivatives (dC, $m^5\text{dC}$, $hm^5\text{dC}$, $f^5\text{dC}$, and $ca^5\text{dC}$ and/or dC*) that deviate from the dG content, which is quantified by UV detection. Therefore, the expected exact 1:1 ratio of $\Sigma x\text{dC}/\text{dG}$ is almost never reached. Nevertheless, the amounts of dC and $x\text{dC}$ derivatives can be expressed relative to the amount of dG, but a corrective

factor is needed. To this end, the amount of each dC derivative is divided by the amount of dG. Multiplying by 100 results in the percentage of every dC derivative. Summarizing these percentages will lead to a total percentage as a corrective factor (%dC^{total}). Dividing the measured %xdC by this factor will provide the term 'MOI/dG'. For mouse gDNA, further multiplication by 0.21 will lead to the desired term 'MOI/dN'.

$$\begin{aligned}
 & \text{dC[pmol]}/\text{dG[pmol]} \times 100 = \% \text{dC}, \\
 & \text{m}^5\text{dC[pmol]}/\text{dG[pmol]} \times 100 = \% \text{m}^5\text{dC}, \\
 & \text{hm}^5\text{dC[pmol]}/\text{dG[pmol]} \times 100 = \% \text{hm}^5\text{dC}, \\
 & \dots \\
 & \sum (\% \text{dC}; \% \text{m}^5\text{dC}; \% \text{hm}^5\text{dC}; \dots) = \% \text{dC}^{\text{total}}. \\
 & \% \text{dC}/\% \text{dC}^{\text{total}} = \text{dC}/\text{dG}^{\text{correct}}, \\
 & \% \text{m}^5\text{dC}/\% \text{dC}^{\text{total}} = \text{m}^5\text{dC}/\text{dG}^{\text{correct}}, \\
 & \% \text{hm}^5\text{dC}/\% \text{dC}^{\text{total}} = \text{hm}^5\text{dC}/\text{dG}^{\text{correct}}, \\
 & \dots \\
 & \text{m}^5\text{dC}/\text{dG}^{\text{correct}} \times 0.21 = \text{m}^5\text{dC}/\text{dN}, \\
 & \text{hm}^5\text{dC}/\text{dG}^{\text{correct}} \times 0.21 = \text{hm}^5\text{dC}/\text{dN}, \\
 & \dots
 \end{aligned}$$

The values of each MOI per sample can be presented as the mean and standard deviation of a technical triplicate. Combination of at least three biological replicates, e.g., their means, leads to reliable data.

Level of expertise needed to implement the protocol

A trained technician, graduate student or postdoctoral researcher can perform all the steps from DNA isolation to DNA digestion and sample preparation. For working with a UHPLC-QQQ-MS system, at least basic knowledge of how to use the machine is required. Core facilities for MS measurements typically operate the LC-MS/MS instrument and perform the standard LC-MS/MS analysis, but it is recommended to use a facility focused on small molecules, with a dedicated instrument and experienced personnel to avoid contamination with distinct analytes that would impair sensitivity. If this facility is unavailable on-site, the DNA samples can also be shipped to a respective facility on dry ice. But any researcher with an interest in nucleoside research can generate and evaluate data when he or she uses the provided material. Nevertheless, the problems addressed in the Troubleshooting section require a more profound knowledge of the mass spectrometer, so adapting the provided method to different nucleosides of interest or mastering upcoming challenges calls for a well-trained researcher.

Limitations

The high sensitivity of a triple quadrupole mass spectrometer originates from the selection for the specific mass signals of the molecules of interest by disregarding other potential contents of the analysis mixture. Before analysis of a sample of interest, it is therefore critical to define all nucleosides that are to be quantified. Molecules that were not considered in the method are not monitored, and their data can therefore not be extracted retrospectively.

The development of measurement protocols for new molecules requires optimization of the LC-based separation part, adjustment of MS parameters such as the optimal collision energy and validation of the nucleoside signal of the MOI. Furthermore, quantification with internal standards requires the availability of suitable isotopologs with a $\Delta m/z$ of at least 2, preferably more. If such an isotopolog is not available commercially, the corresponding molecule must be synthesized chemically or metabolically^{63,64} or one needs a collaborator who can provide it.

Because the origin of the naturally occurring nucleosides is DNA, the extraction and digestion efficiency are critical. Extracting DNA from cell culture usually leads to high yields and can typically also be scaled up. DNA isolation from tissues, however, is more demanding and may not provide a sufficient amount of DNA to detect a modification with low abundance. In this case, it might be necessary to combine several biological samples. The isolation of DNA from only a few cells and subsequent quantification of modified nucleosides has been reported⁶⁵, but it is not routinely possible. At the other extreme, excessive amounts of DNA and corresponding nucleosides can result in

so-called matrix effects, which can suppress the signal of a nucleoside. It is therefore required to determine the optimal amount of DNA to be analyzed to obtain the best signal. The reported method provides global quantitative data; sequence information is not available.

The whole method depends on the quality of the input material, as the origin of the resulting DNA, and the respective nucleosides, cannot be determined at a later stage. Contaminations, e.g., from a bacterial or fungal source during cell culture work or from the microbiome of a higher organism, might lead to false-positive results for certain modified nucleosides. In addition, the abundance of modifications may vary substantially during cell differentiation or in response to stress. Therefore, the timing of the cell culture work, e.g., harvesting time points, is very important to obtaining reliable values for the biological replicates.

There are also some challenges when performing the measurements, as not every rare nucleoside of interest can be quantified with every setup. Sometimes it may not be possible to measure all nucleosides with the same setup and conditions. If no further optimization is possible, one needs to process the sample with two different methods, but this of course requires increased instrument time and it also requires more material. For nucleosides that are quantified using the UV trace (dC and dG), it is critical that the peaks be baseline separated. For MS analysis, chromatographic separation of nucleosides is mostly not necessary because of the separation in the mass spectrometer according to their unique m/z values and fragmentation patterns. Only when analyzing samples that might contain isomers, e.g., m^5dC and m^4dC , which have identical precursor and product ions, chromatographic separation becomes necessary in order to clearly determine the identity of the detected signal.

To further increase the sensitivity and the number of MS data points, it is beneficial to subdivide the table of analytes into several segments according to their RTs. In each time segment, only certain nucleosides are monitored, which increases the dwell time for each analyte and thus the strength of the signal. The price of this increased MS sensitivity is that, depending on chromatographic performance, the RTs of the nucleosides might change, and then these nucleosides may escape their time segment. Then the corresponding data of the analyte in the time segment are irretrievably lost.

Materials

Biological materials

- Cell line: cell lines of various sorts, as well as animal tissues, have worked well in our experience. Specifically, we have performed this protocol using iNGN cells⁶⁶ (hPSCreg no. [HVRDi004-B-1](#)); HEK293T cells (ATCC, cat. no. CRL-3216); mES wild-type cell line J1 (ref. ⁶⁷); and mouse cerebellum from a C57-BL6/J wild-type genetic background, provided by S. Michalak (Department of Pharmacy, Ludwig-Maximilians-Universität München). **! CAUTION** The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma. **! CAUTION** Ensure sterile work in order to avoid cross-contamination of the extracted tissue with bacteria or other organisms from the environment. **! CAUTION** All animal experiments must be performed according to the relevant guidelines and regulations and must be approved by your institutional animal care and use committee.

Reagents

- 2-Mercaptoethanol (β ME, CAS no. 60-24-2; Sigma-Aldrich, cat. no. M3148-25mL)
! CAUTION 2-Mercaptoethanol is toxic, so avoid exposure.
- Acetonitrile (MeCN, 99.95% (vol/vol), LC-MS grade; Roth, cat. no. AE70.2) **▲ CRITICAL** Sensitivity of the mass spectrometer might vary when a different supplier is used. Check new suppliers carefully.
- Antarctic phosphatase (New England Biolabs, cat. no. M0289L)
- BHT (CAS no. 128-37-0; Sigma-Aldrich, cat. no. B1378-100G) **▲ CRITICAL** To keep background oxidation at a minimum, it is recommended to store BHT powder under vacuum.
- Benzonase nuclease, 10 KU (VWR, cat. no. 70746-3)
- Blood & Cell Culture DNA Midi Kit (Qiagen, cat. no. 13343)
- Degradase Plus (Zymo Research, cat. no. E2021)
- DFOA (Sigma-Aldrich, CAS 138-14-7, cat. no. D9533-1G)
- Dimethyl sulfoxide (DMSO, CAS no. 67-68-5; Acros, cat. no. 327182500)
- DNA pre-wash buffer (Zymo Research, cat. no. D3004-5)
- Dulbecco's PBS without $MgCl_2$, $CaCl_2$ (DPBS; sterile filtered, suitable for cell culture; Sigma-Aldrich, cat. no. D8537-500 mL)
- EDTA disodium salt ($Na_2[EDTA]$; CAS no. 6381-92-6; VWR, cat. no. 33600.267)

- Formic acid (CAS no. 64-18-6; Fluka Honeywell Chemicals, cat. no. 94318-50 mL-F) **!CAUTION** Formic acid is highly corrosive and can lead to severe burns when inhaled. Use only in a fume hood or in a highly ventilated area and protect your skin and eyes carefully.
- gDNA wash buffer (Zymo Research, cat. no. D3004-2)
- Genomic lysis buffer (Zymo Research, cat. no. D3004-1) **▲ CRITICAL** Do not add β ME in advance.
- Glycerol (CAS no. 56-81-5; Roth, cat. no. 3783.2)
- Magnesium chloride hexahydrate ($\text{MgCl}_2 \cdot \text{H}_2\text{O}$, CAS no. 7786-30-3; Merck, cat. no. M8266-100g)
- Methanol (MeOH; 99.9% (vol/vol), LC-MS grade VWR, cat. no. HONC34966-1L) **!CAUTION** Methanol is toxic, so avoid exposure. **▲ CRITICAL** Mass sensitivity might vary when a different supplier is used. Check new suppliers carefully.
- Nuclease S1 from *Aspergillus oryzae* (Merck, cat. no. N5661-50 KU) **▲ CRITICAL** Stock solution should be stored at -20°C and kept on ice while making aliquots for usage.
- Parafilm M (4×125 inches, clear; Bemis, cat. no. 52858-000)
- Phosphodiesterase I from *Crotalus adamanteus* venom (Abnova, cat. no. P5263)
- PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, cat. no. P7589)
- Buffer RLT (Qiagen, cat. no. 79216)
- RNase A (100 mg/mL, 7,000 U/mL; Qiagen, cat. no. 19101)
- Sodium chloride (NaCl; CAS no. 77-86-1; Bernd Kraft, cat. no. 10724344)
- Tetrahydrouridine (THU; CAS no. 18771-50-1; Merck Millipore, cat. no. 584222)
- Tris(hydroxymethyl)aminomethane (Tris base; CAS no. 77-86-1; Fisher Scientific, cat. no. BP152-5)
- Water, LC-MS grade (Honeywell, cat. no. 39253-1L) **▲ CRITICAL** Mass sensitivity might vary when a different supplier is used. Check new suppliers carefully.
- Zinc sulfate (ZnSO_4 , CAS no. 7446-20-0; Grüssing, cat. no. 14039)

Nucleosides

- 2'-Deoxyadenosine (dA) (Carbosynth, cat. no. ND04011, CAS 16373-93-6)
- 2'-Deoxycytidine (dC) (Carbosynth, cat. no. ND06286, CAS 951-77-9)
- 2'-Deoxyguanosine (dG) (Carbosynth, cat. no. ND06306, CAS 961-07-9)
- Thymidine (T) (Carbosynth, cat. no. NT02592, CAS 50-89-5)
- 5'-Methyl-2'-deoxycytidine (m5dC) (Carbosynth, cat. no. ND06242, CAS 838-07-3)
- 5'-Hydroxymethyl-2'-deoxycytidine (hm5dC) (Carbosynth, cat. no. NH15898, CAS 7226-77-9)
- 5'-Formyl-2'-deoxycytidine (f5dC) (Carbosynth, cat. no. ND63556, CAS 137017-45-9)
- 5'-Carboxy-2'-deoxycytidine (ca5dC) (Carbosynth, cat. no. ND158446, CAS 46003-72-9)
- 8-Oxo-7,8-dihydro-deoxyguanosine (8oxodG) (Carbosynth, cat. no. ND06344, CAS 88847-89-6)
- 15N5-8oxodG (Cambridge Isotope Laboratories, cat. no. NLM-67 15-0, CAS NA)
- D3-m5dC (synthesis described in ref. 36)
- 15N2-hm5dC (synthesis described in ref. 41)
- 15N2-f5dC (synthesis described in ref. 41)
- 15N2-ca5dC (synthesis described in ref. 41)
- N4-methyl-2'-deoxycytidine (m4dC) (synthesis described in ref. 26)
- 15N2-m4dC (synthesis described in ref. 26)
- N6-methyl-2'-deoxyadenosine (m6dA) (synthesis described in ref. 26)
- D3-m6dA (synthesis described in ref. 26)

Equipment

- Stainless-steel beads (5 mm; Qiagen, cat. no. 69989)
- Syringe filter (0.2- μm cellulose acetate; VWR, cat. no. 514-0061)
- Snap ring cap 11 mm tr. (natural rubber/TEF, 60° , 1.0 mm, HPLC vial cap; VWR cat. no. 548-0014)
- Snap ring micro-vial (0.3 mL, polypropylene, 32×11.6 mm, transparent, HPLC vial; VWR, cat. no. 548-0120) **▲ CRITICAL** If using different HPLC vials, they must have a volume-reducing insert.
- Poroshell 120 SB C8 column (2.7 μm , 2.1×150 mm; Agilent Technologies, cat. no. 683775-906)
- Poroshell 120 SB-C18 column (2.7 μm , 2.1×150 mm; Agilent Technologies, cat. no. 683775-902)
- PCR plate (skirted, 96-well, 0.2 mL; VWR, cat. no. 732-3225)
- Falcon tubes (15 mL; VWR, cat. no. 188271)
- Falcon tubes (50 mL; VWR, cat. no. 227161)
- AcroPrep Advance 96-well, 350- μL , 0.2- μm Supor short-tip natural polypropylene plates (Pall, cat. no. 518-0022)

- Centrifuge tubes (0.5 mL; Eppendorf, cat. no. 211-2140)
- Centrifuge tubes (1.5 mL; Eppendorf, cat. no. 211-2130)
- Centrifuge tubes (2.0 mL; Eppendorf, cat. no. 211-2120)
- $-20\text{ }^{\circ}\text{C}$ Freezer (e.g., Bosch, model no. GSN58AW45)
- $-80\text{ }^{\circ}\text{C}$ Freezer (e.g., Eppendorf, Innova U725, model no. U9440-0002)
- 0.5- to 10- μL , 2- to 20- μL , 10- to 100- μL , 20- to 200- μL , and 100- to 1,000- μL pipettes (Eppendorf)
- $4\text{ }^{\circ}\text{C}$ Refrigerator (e.g., Liebherr, model no. LCv 4010)
- $37\text{ }^{\circ}\text{C}$ Heat block (Eppendorf, ThermoMixer Comfort model, device, cat. no. 5382000015, plus top, cat. no. 5360000011)
- MM400 bead mill (Retsch, cat. no. 20.745.0001)
- Refrigerated benchtop microcentrifuge (e.g., Centrifuge 5424R; Eppendorf, cat. no. 5404000014)
- Refrigerated swinging-bucket rotor centrifuge (Centrifuge 5810R; Eppendorf, cat. no. 5811000428 with A-4-81 rotor and holders for MTP plates)
- Microscale (e.g., Sartorius, cat. no. RC 210 P)
- Ultrapure water system (e.g., arium pro DI; Sartorius Stedim Biotech, cat. no. H2OPRO-DI-B)
- UV/Vis spectrophotometer (NanoDrop; NanoDrop Technologies, cat. no. ND-1000)
- Triple quadrupole LC/MS system with iFunnel technology (Agilent Technologies, model no. 6490)
- UHPLC system (Agilent Technologies, model no. 1290 Infinity II LC)
- Vortex mixer (Scientific Industries, model no. Vortex-Genie 2)
- Zymo-Spin IIC-XL column (Zymo Research, cat. no. C1102)

Software

- Microsoft Office Excel 2016 (Microsoft, <https://products.office.com/en-us/compare-all-microsoft-office-products?activetab=tab%3aprimaryr1>)
- OriginPro 2016G b.9.3.226 (<https://www.originlab.com/2016>)

Reagent setup

▲ **CRITICAL** Deionized water is used for all solutions, unless otherwise indicated. ▲ **CRITICAL** For the gDNA isolation, work at room temperature ($23\text{ }^{\circ}\text{C}$) all the time. Some buffers and solutions will freeze at $4\text{ }^{\circ}\text{C}$. ▲ **CRITICAL** For each procedure using stock solutions, let all nucleoside and salt dilutions thaw and equilibrate at room temperature and vortex them vigorously (for at least 1 min). ▲ **CRITICAL** Perform all pipetting steps with well-calibrated pipettes.

Internal standard mastermix

To make the internal standard (ISTD) mastermix, add 58.0 μL of $\text{D}_3\text{-m}^5\text{dC}$ (m^5dC^* , concentration (c) = 264.1 μM), 90.0 μL of $\text{D}_2,^{15}\text{N}_2\text{-hm}^5\text{dC}$ (hm^5dC^* , c = 25.5 μM), 90.0 μL of $^{15}\text{N}_2\text{-f}^5\text{dC}$ (f^5dC^* , c = 0.152 μM), 120.0 μL of $^{15}\text{N}_2\text{-ca}^5\text{dC}$ (ca^5dC^* , c = 0.108 μM), 180.0 μL of $^{15}\text{N}_5\text{-8oxodG}$ (8oxodG^* , c = 0.181 μM). Sum the real volume and bring the volume to 900.1 μL by adding water. Vortex rigorously for 1 min. The prepared ISTD mastermix, which is sufficient for 300 measurements, can be stored at $-20\text{ }^{\circ}\text{C}$ for up to 1 year, and multiple thawing and freezing cycles are acceptable. ▲ **CRITICAL** To ensure accurate preparation of the ISTD mastermix, the added volume after each pipetting step is controlled by using microscopes. For each MOI^* , a deviation in weight of $\pm 5\%$ is tolerated. If the deviation is $> -5\%$, one could add the required volume of the MOI^* , but if the deviation is $> 5\%$, the sample must be discarded. ▲ **CRITICAL** If you are interested in MOI^* s other than those listed in the ISTD mastermix, e.g., m^6dA or m^4dC , the MOI^* s in the ISTD mastermix can be adjusted accordingly.

1,000 \times BHT

Prepare a 200 mM stock solution of BHT (1,000 \times BHT) in DMSO. Make 50- μL aliquots and store them at $-80\text{ }^{\circ}\text{C}$ for up to 3 years. 1,000 \times BHT can be thawed and refrozen up to three times but must be refrozen as soon as possible, and multiple freeze-thaw cycles should be avoided. Before use, the 1,000 \times BHT must be diluted 1:10 to make it soluble in water (100 \times BHT). Combine 1 equiv. of 1,000 \times BHT with 3 equiv. of DMSO and vortex briefly. Add 6 equiv. of H_2O dropwise and vortex thoroughly between drops to make sure that the BHT does not precipitate. 100 \times BHT should be a clear solution at the end. If the solution is turbid, vortex longer. If too much BHT has already precipitated, discard the dilution and prepare a new 100 \times BHT dilution from the 1,000 \times BHT stock. 100 \times BHT must be prepared immediately before use. ▲ **CRITICAL** Do not refreeze or store the 100 \times BHT; instead prepare it fresh each time before the isolation.

1,000× DFOA

Prepare a 200 mM stock solution of DFOA (1,000× DFOA) in degassed water. Make 20- μ L aliquots and store them at -80 °C for up to 3 years. **▲ CRITICAL** 1,000× DFOA should not be refrozen. **▲ CRITICAL** When DFOA is dissolved in water, the final volume increases. Therefore, dissolve DFOA in only 80% of the calculated amount of degassed water, measure the volume afterward and add the missing volume at the end.

Lysis buffer

If the option to isolate RNA and total protein (denatured) is desirable, use Buffer RLT as the base lysis buffer. If only isolation of gDNA is required, use genomic lysis buffer (already contains RNase) instead. The yield of gDNA will be slightly higher if you use genomic lysis buffer. To 1 equiv. of base lysis buffer, add 0.01 equiv. of β ME (14.3 mM final concentration), 0.002 equiv. of 1,000× DFOA (400 μ M final concentration) and 0.02 equiv. of 100× BHT (400 μ M final concentration). For example, to 1 mL of base lysis buffer, add 10 μ L of β ME, 2 μ L of 1,000× DFOA and 20 μ L of 100× BHT. The ready-to-use lysis buffer is referred to as Buffer RLT⁺ or genomic lysis buffer⁺, respectively. These buffers without β ME, BHT and DFOA can be stored at room temperature until the expiration date given by the manufacturer. The buffers supplemented with β ME, BHT and DFOA must be prepared immediately before use and cannot be stored. **▲ CRITICAL** If you are interested in any kind of deamination, you should consider adding a broadband deamination inhibitor such as tetrahydrouridine (THU).

RNase wash buffer

To 1 equiv. of genomic lysis buffer, add 0.002–0.01 equiv. of RNase A (0.2–1 mg/mL final concentration), 0.002 equiv. of 1,000× DFOA (400 μ M final concentration) and 0.02 equiv. of 100× BHT (400 μ M final concentration). The amount of RNase A to be added can be adjusted according to the RNA content of the cells. If cells are known to contain a high amount of total RNA, we recommend using 0.01 equiv. of RNase A, but 0.002 equiv. is sufficient in most cases. For example, to 1 mL of genomic lysis buffer, add 2–10 μ L of RNase A, 2 μ L of 1,000× DFOA and 20 μ L of 100× BHT. RNase wash buffer must be prepared immediately before use and cannot be stored. **▲ CRITICAL** If you are interested in any kind of deamination, you should consider adding a broadband deamination inhibitor such as THU.

Washing of stainless-steel beads

To wash the beads, shake them in a soap and water mix for 10 min and rinse them with plenty of water to remove all the remaining soap. Wash them once with acetone and twice in pure ethanol. If necessary, sonicate between washes. Let the beads dry in an oven and let them equilibrate to room temperature before use. The washed beads should be protected from dust and can be stored at room temperature for an infinite amount of time without further washing.

Phosphodiesterase I buffer

Phosphodiesterase I buffer is 5.5 mM Tris (pH 8.9), 5.5 mM NaCl, 0.7 mM MgCl₂·H₂O, 50% (vol/vol) glycerol and 50% (vol/vol) water. The buffer can be stored at -20 °C for long-term storage (up to 5 years) and should be filtered with a syringe filter (0.2- μ m cellulose acetate) before use. **▲ CRITICAL** Sodium (Na⁺) adducts form easily, and these can potentially distort the data (the mass of the MOI-Na⁺ adduct is not selected). In addition, an abundance of ions in the mass device could enhance matrix effects. However, for this application, Na⁺ cannot be avoided and, in our experience, does not cause any problems.

Snake venom phosphodiesterase I

Use a syringe filter (0.2- μ m cellulose acetate) to filter 10 mL of phosphodiesterase I buffer. Add 1 mL of this sterile buffer to the phosphodiesterase I pellet and dissolve it by slowly inverting the vessel (enzyme concentration = 100 U/mL). If the enzyme does not dissolve, vortex briefly. Make 100- μ L aliquots and store them at -20 °C until usage for up to 1 year.

ZnSO₄ stock solution

Prepare a 4 mM solution of ZnSO₄ in water. The stock solution can be stored at -20 °C for up to 1 year.

EDTA stock solution

Prepare a 1 mM solution of Na₂[EDTA] in water (pH 8.0). The stock solution can be stored at –20 °C for up to 1 year. **▲ CRITICAL** Na⁺ is critical for sensitivity in mass spectrometry; however, for this application, Na⁺ cannot be avoided and, in our experience, does not cause any problems.

Nuclease S1 solution

Dilute the stock nuclease S1 (100,000 U/mL) to a concentration of 18,400 U/mL for usage. Mix this dilution by pipetting up and down. For example, add 20.4 μL of water to 4.6 μL of nuclease S1. Nuclease S1 solution must be prepared immediately before use and cannot be stored. **▲ CRITICAL** Pipette the enzyme very slowly because of its high viscosity, which is due to its glycerol-containing storage buffer. The nuclease S1 stock solution should be stored for only a few minutes on ice until dilution. Prepare freshly diluted nuclease S1 for each DNA digestion and prepare at least 1 μL more than needed for the mastermix 1.

Mastermix 1

For each sample to be digested, you need 7.5 μL of mastermix 1. Per 7.5 μL of mastermix 1, add the calculated amount of water, ZnSO₄ stock solution to a final amount of 3.6 nmol, nuclease S1 solution to a resulting amount of 18.4 U, antarctic phosphatase to a resulting amount of 5 U and specific amounts of labeled internal standards (pH 6.0) (add the reagents in the specified order). See the provided Excel Sheet (Supplementary Methods 3) for details. **▲ CRITICAL** The amount of enzyme can be adjusted according to the amount of DNA you want to digest. It is possible to digest at least 10 μg of DNA per sample using this mastermix. Mastermix 1 must be freshly prepared for each digestion and should be stored on ice only for a few minutes before usage. **▲ CRITICAL** If you are interested in any kind of deamination, you should consider adding a broadband deamination inhibitor such as THU. This will ensure that any detected deaminated nucleosides are native and not produced due to the digestion conditions, as deaminases are often contaminations of commercially available nucleases.

Mastermix 2

For each sample to digest, you need 7.5 μL of mastermix 2. Per 7.5 μL of mastermix 2, add the calculated amount of water, EDTA stock solution to a final amount of 3.9 nmol and snake venom phosphodiesterase to a resulting amount of 0.15 U (add in the specified order). See the provided Excel Sheet (Supplementary Methods 3) for details. **▲ CRITICAL** The enzyme concentrations can be adjusted according to the amount of DNA you want to digest. It is possible to digest at least 10 μg DNA per sample using this mastermix. Mastermix 2 must be freshly prepared for each digestion and should be stored on ice for only a few minutes before usage.

Procedure

General procedure for dissolution and dilution of nucleosides ● Timing 30 min

- 1 Dissolve a small amount (typically 1–2 mg in 1,000 μL) of each desired nucleoside (natural and isotopically labeled) in a 1.5-mL centrifuge tube in water.
- 2 Measure the absorption of the solution at the respective wavelength for the maximum absorption of the nucleoside on a photometer.

▲ CRITICAL STEP The extinction coefficient must be known for the respective nucleoside (see Supplementary Table 6 for the extinction coefficients of m⁵dC, m⁴dC, hm⁵dC, f⁵dC, ca⁵dC and m⁶dA). If it is unknown, use microscales (Box 1) instead. When using a photometer, make sure that your absorption is within the linear range.
- 3 Determine the concentration *c* of the nucleoside using the Beer–Lambert law:

$$c = \frac{E_{\lambda}}{\epsilon_{\lambda} \times d}$$

with wavelength-dependent extinction E_{λ} , wavelength-dependent extinction coefficient ϵ_{λ} and path length d .

- 4 Dilute the nucleoside with water to the desired concentration.

▲ CRITICAL STEP Very accurate pipettes are needed. Always use the pipette with the smallest margin of error. If you cannot ensure that your pipettes are accurate, use calibrated microscales (Box 1).

■ PAUSE POINT The dissolved nucleoside can be stored at –20 °C for up to 1 year. If you want to store it longer, we recommend storing it at –80 °C.

Box 1 | Use of microscales for dilution of nucleosides ● Timing 15–30 min

- 1 Use microscales to weigh small amounts of the natural nucleoside and the isotopically labeled nucleoside into separate tared 1.5-mL centrifuge tubes.
- 2 Add an appropriate amount of water to each tube and weigh the tube again.
- 3 Calculate the respective concentrations.
- 4 If your stock solution is too concentrated and you want to dilute it further, calculate the necessary volume of water and use microscales to control the addition until the desired concentration is reached.

Preparation of ISTD mastermix ● Timing 1–1.5 h

▲ **CRITICAL** Very accurate pipettes are needed. Always use the pipette with the smallest margin of error. If you cannot ensure that your pipettes are accurate, use calibrated microscales.

- 5 Thaw all labeled nucleosides needed for the ISTD mastermix for at least 30 min at room temperature, vortex rigorously and spin down (5,000g, room temperature, 3 s).
 - ▲ **CRITICAL STEP** If the concentrations of the labeled nucleosides do not match the concentration needed for the ISTD mastermix, dilute them further (or if the concentration is too low, repeat Steps 1–4 for the desired nucleoside).
- 6 Prepare the ISTD mastermix.
 - ▲ **CRITICAL STEP** If your ISTD mastermix deviates from the one given in the ‘Reagent setup’ section, consider adjusting the solvent, column and method for the UHPLC–MS/MS in Steps 7 and 8.
 - **PAUSE POINT** If the ISTD mastermix is not evaluated at once, store it at -20°C for up to 1 year.

Preparation for UHPLC–MS/MS and performance of sensitivity checks ● Timing 1–1.5 h

▲ **CRITICAL** From now on, the procedure is described for the epigenetic modifications and the provided acquisition method (Supplementary Methods 4) must be applied. For the modifications m^4dC and m^6dA , use MeOH as solvent B, the Poroshell 120 SB C18 as a column and the acquisition methods for m^4dC - and m^6dA -sensitive mode (Supplementary Methods 5) or m^4dC - and m^6dA -insensitive mode (Supplementary Methods 6).

- 7 Set up the UHPLC–MS/MS system by installing the Poroshell 120 SB C8 column and preparing new buffers for LC. Therefore, add 75 μL of formic acid to a full 1.0-L bottle of MS-grade water (solvent A) and 187.5 μL of formic acid to a full 2.5-L bottle of MS-grade acetonitrile (solvent B). (Alternatively, if solvent B is MeOH, use a full 1.0-L bottle of MS-grade MeOH and add 75 μL of formic acid.)
- 8 Attach the bottles to the UHPLC system (solvent A to port A, solvent B to port B) and purge at 50% A/50% (vol/vol) B for 5 min with a flow of 5 mL/min. To provide reproducible separation efficiency, the columns must be equilibrated. Therefore, you need to perform 20 chromatographic runs using the method for epigenetic modifications (Table 1) without injection for new columns and three runs before measurement of the first sample per set. The flowrate is 0.35 mL/min, the pressure is 600 bar, the temperature of the column oven of the UHPLC is 35°C and the gas temperature is 80°C . The UV detector monitors an absorption signal at the 260-nm wavelength. (Alternatively, if the method for m^6dA and m^4dC is used, see Supplementary Table 7. The temperature of the column oven of the UHPLC is 30°C ; flowrate, pressure and gas temperature are not changed.)
- 9 If this is the first ever produced ISTD mastermix, add 12 μL to an HPLC vial and analyze it three times, each with an injection volume of 3 μL . Evaluate the ISTD mastermix according to Steps 49–53 and Steps 55 and 56. If a previous ISTD mastermix is available, add 12 μL of the previous one and 12 μL of the new one each to separate HPLC vials. Measure both three times, alternating with an injection volume of 3 μL , and evaluate the respective data. Determine the resulting areas for each labeled nucleoside, calculate the mean of this technical triplicate and evaluate the deviation of the mean as a percentage between the previous ISTD mastermix and the new one.
 - ▲ **CRITICAL STEP** The deviation of the resulting areas must be smaller than 5% for each labeled nucleoside in order for this ISTD mastermix to qualify for exact quantification.
- 10 For the sensitivity check, add a few microliters ($3 + 3 \times n$ μL , n = number of performance checks) of the ISTD mastermix to an HPLC vial.
- 11 Measure 3 μL of the ISTD mastermix with the provided measuring method, integrate the areas of each labeled nucleoside and compare the areas of the labeled nucleosides and the corresponding SNRs from this measurement with those of the measurements from Step 9.

? TROUBLESHOOTING

Table 1 | UHPLC gradient elution table for the method for epigenetic modifications

	Time (min)	Solvent A (%)	Solvent B (%)
1	0.00	100.0	0.0
2	4.00	96.5	3.5
3	6.90	95.0	5.0
4	7.20	20.0	80.0
5	10.50	20.0	80.0
6	11.30	100.0	0.0
7	14.00	100.0	0.0

Table 2 | Example of a dilution series from ~10 pmol to ~10 fmol with the corresponding calibration levels L11-1

	L11	L10	L9	L8	L7	L6	L5	L4	L3	L2	L1
$n(\text{MOI})$ [pmol]	10	5	2.5	1.25	0.625	0.313	0.156	0.078	0.039	0.020	0.010
$n(\text{MOI}^*)$ [pmol]	2.48	2.48	2.48	2.48	2.48	2.48	2.48	2.48	2.48	2.48	2.48
n/n^*	4.037	2.018	1.009	0.505	0.252	0.126	0.063	0.032	0.016	0.008	0.004

Table 3 | Example for setting up the dilution mix and the calibration mix (Calmix)

	Calc. for the highest data point (μL)	Dilution mix		Calmix	
		Factor	Vol (μL)	Factor	Vol (μL)
MOI	2	—	—	8	16
MOI*	4	37	148	8	32
Water			925		184
Total volume			1,073		232

Calculation of calibration curves ● Timing 1 d

- 12 Decide which molar range the calibration curve should span; the calibration curve will consist of 11 different data points (levels, L) with serial 1:2 dilutions of the natural nucleoside while the amount of the labeled nucleoside is kept constant. Note that the highest data point should contain about four times more of the natural nucleoside (MOI) as compared to the isotopically labeled nucleoside (MOI*) and consider this in the next step. Example: if an amount of 500 fmol of the MOI is expected, span the calibration curve from ~10 fmol to ~10 pmol (Table 2).
- 13 Calculate which volumes (in μL) of the stock solutions (natural and labeled) equal the amount of substance for the highest point of the calibration curve. Example: The calibration curve should span 10 fmol to 10 pmol; thus the highest data point for MOI is 10 pmol. If the concentration of the stock solution for MOI is 5 μM (5 pmol/ μL), a volume of 2 μL of stock solution is needed for the highest data point. If the concentration of the stock solution for MOI* is 0.62 μM , a volume of 4 μL of stock solution is needed in this case to obtain 2.48 pmol (~1/4 of 10 pmol) of MOI* (Table 3).
- 14 Vortex all nucleoside solutions vigorously (for at least 1 min).
- 15 Prepare the dilution mix and the Calmix with amounts as calculated in Table 3.
- 16 Prepare eleven 1.5-mL centrifuge tubes and label them from L1 to L11. Add 100 μL of the dilution mix to each of the tubes L1–L10.
- 17 Add 100 μL each of the Calmix to L10 and L11. Set L11 aside and vortex L10 vigorously (for at least 1 min).
- 18 Make a serial dilution as shown in Fig. 2.

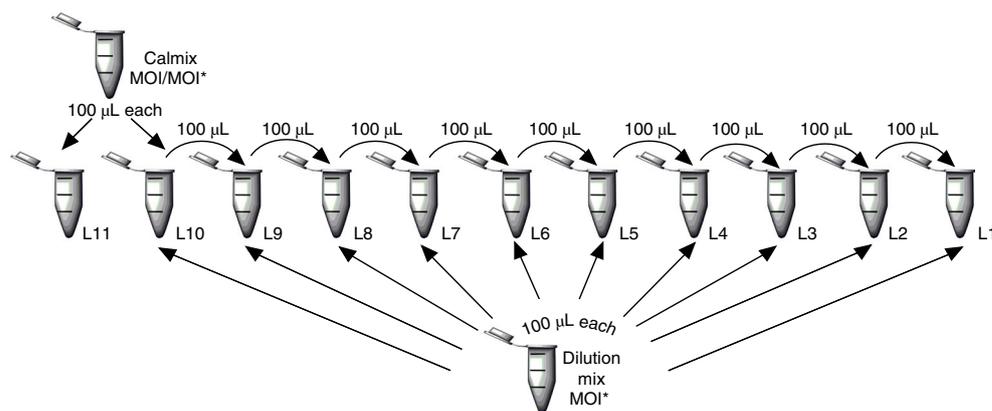


Fig. 2 | Calibration curve preparation. Pipetting scheme for generating a calibration curve of MOI and MOI* with a fixed specific amount of MOI*. L, level.

Table 4 | Example of evaluation of a calibration curve for m^4dC

Calibration level	n/n^*	$A(m^4dC)$	$A^*(^{15}N_2-m^4dC)$	A/A^*	Replicate no.
L1	0.004	4,797	1,745,218	0.003	1
L2	0.008	7,214	1,737,978	0.004	
—	—	—	—	—	
L1	0.004	5,094	1,611,819	0.003	2
L2	0.008	6,145	1,658,872	0.004	
—	—	—	—	—	
L1	0.004	4,728	1,689,815	0.003	3
L2	0.008	6,595	1,680,195	0.004	
—	—	—	—	—	

- 19 Transfer the contents of each tube to an HPLC vial and close the vial with an HPLC vial cap. The injection volume for one measurement is 29 μ L. L1 contains the lowest concentration and is therefore the first vial to be measured. Proceed then with measuring L2-11 in their respective order. After measuring each level once, insert a blank measurement into the sample queue without injection. Then repeat the procedure twice, starting from L1 to produce a technical replicate.
- 20 For the evaluation, apply Steps 49–53 and 55, integrate the peaks for the MOI (A) and the MOI* (A^*) and transfer the values for A and A^* to an Excel sheet, then calculate A/A^* for each data point. In three additional columns, record the amounts for MOI and MOI* and the respective n/n^* values (Table 4).
- 21 Calculate the mean, the standard deviation and the standard deviation as percentages (%s.d.) for the A/A^* values of all data points from the technical replicates. If the %s.d. exceeds 20%, the data point is not valid and must be excluded, if it is the highest or lowest data point. Invalid data points in between render the whole calibration curve invalid.
- 22 Plot the values for the mean of A/A^* value against the respective n/n^* and perform a linear regression of the data points using Origin or similar calculation software. Record the values for m and t .
- 23 Perform an accuracy check (backfit) by inserting the value of n/n^* into the calculated linear equation and calculating the A/A^* value of every single calibration level. Then determine the percentage of the calculated A/A^* value in comparison to the intended A/A^* value. The resulting percentage provides the accuracy of the linear equation and should not be <80% or >120%.
- 24 If the accuracy check fails, remove the highest or lowest data point and repeat Steps 22 and 23 with only the remaining data points. Repeat this step until only data points with sufficient accuracy make up the linear equation. A valid linear equation must consist of at least five consecutive levels.

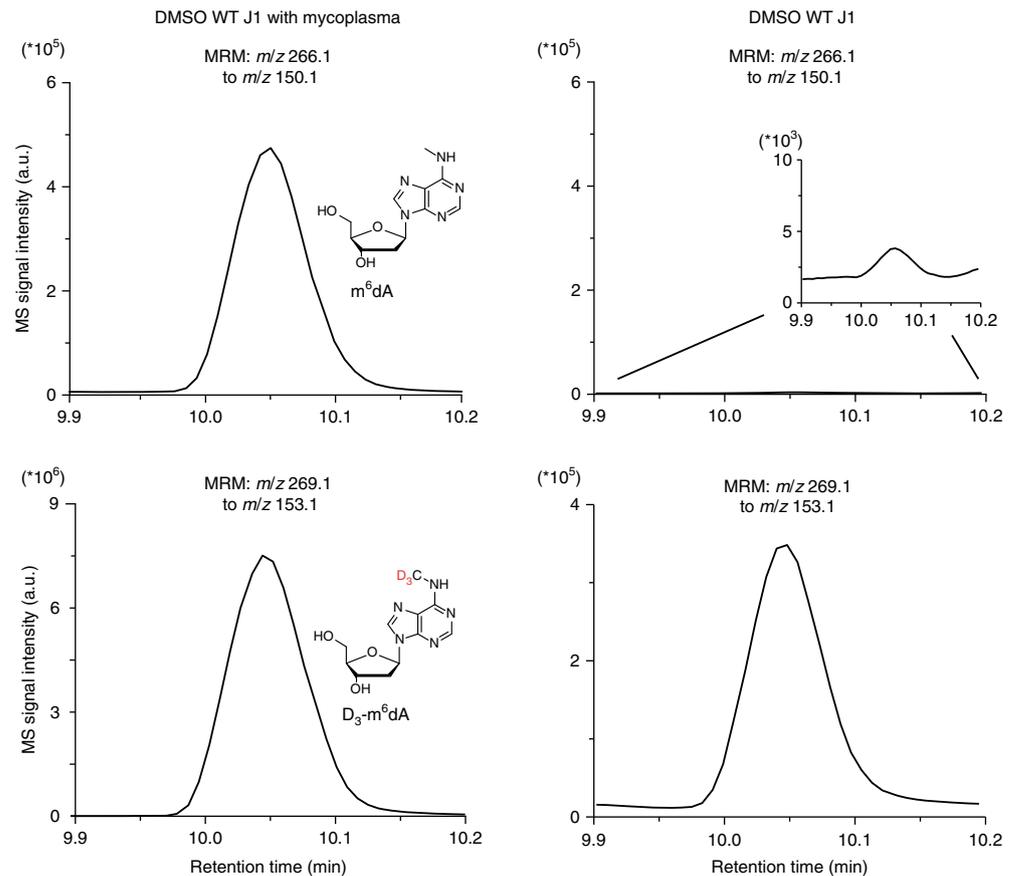


Fig. 3 | Exemplary impairment of results by mycoplasma contamination. Chromatograms with nucleoside signals for the internal standard D_3 - m^6dA (lower row) and natural m^6dA (upper row) in 100 ng of DNA from WT J1 mouse embryonic stem cells (mESCs) with a mycoplasma contamination (left column) or 10 μ g from mycoplasma-free DNA of WT J1 cultures (right column). a.u., arbitrary units; WT, wild-type.

Isolation of gDNA ● Timing 1.5-3 h

- 25 Before lysing the cells, wash them once with DPBS and remove all DPBS before proceeding.
 - ▲ **CRITICAL STEP** If the cells are loosely adherent and will be washed away with DPBS, they can be directly lysed after removing the medium. However, this will lower the yield of isolated gDNA.
 - ▲ **CRITICAL STEP** Make sure that the cultures are not contaminated with mycoplasma or other microorganisms, as this may substantially alter the abundance of certain DNA modifications (Fig. 3).
- 26 Add ~200-300 μ L of lysis buffer per 10^6 cells. For a typical six-well plate with 75% confluent mouse embryonic stem cells (mESCs; $5-8 \times 10^6$ cells, depending on specific culture conditions), use 1.6 mL of lysis buffer per well. Pipette up and down several times to reduce lysate viscosity and ensure homogeneity. For lysing tissue samples, add ~1.5 mL of lysis buffer to 50 mg of tissue. Transfer the lysate to 2-mL centrifuge tubes.
 - ▲ **CRITICAL STEP** We recommend direct lysis in the tissue culture vessel because it is convenient and, in direct comparison tests, we have experienced increased quantities of oxidative lesions after harvesting by trypsinization. However, if oxidative lesions are not a major concern, cells can be harvested and further processed, for example, for counting or flow-cytometric analysis, before being lysed as a pellet. Whenever possible, lyse samples before freezing, as the highly chaotropic lysis buffer will inactivate all enzymatic processes. If not avoidable, add lysis buffer directly to frozen cell pellets or frozen soft tissue in very small pieces without letting them thaw. To ensure quick and complete lysis, vortex immediately and thoroughly until no pellet/tissue fragments are visible. Before lysing larger or harder frozen tissue samples, pulverize them using pestle, mortar and liquid nitrogen to prevent them from thawing.
 - ▲ **CRITICAL STEP** The amount of lysis buffer affects the efficiency of gDNA isolation. Bead milling or sonication is used to further homogenize the lysate and shear the gDNA. If the concentration of

DNA in the lysate is too high (high viscosity), DNA shearing and the final yield of DNA will be reduced because of inefficient elution from the spin column.

■ PAUSE POINT Lysate can be stored at -80°C for up to 5 years. Frozen lysate can be thawed at room temperature before further processing.

- 27 Add one stainless-steel bead per tube. Further homogenize/shear DNA in the bead mill for 1 min at 30 Hz. For tissue samples, first apply 30 Hz for 1 min and then apply an additional 20 Hz for 4 min.

▲ CRITICAL STEP Sonication can also be used for homogenization/DNA shearing. However, we detected increased quantities of abasic sites using this method. If no bead mill or sonicator is available, it is also possible to skip Step 27 and continue directly with Step 29. In this case, use larger volumes of lysis buffer instead and vortex more extensively to provide sufficient DNA shearing, which is important to allow efficient elution from the spin column at the end.

? TROUBLESHOOTING

- 28 Centrifuge for 5 min at 21,000g at 23°C to remove the foam (denatured proteins). If there is no foam, the lysis did not work or the number of cells in the tube was very low. If you expect low cell numbers, we highly recommend proceeding directly with Step 29.

- 29 Load the lysate onto a Zymo-Spin IIC-XL spin column and centrifuge for 1 min at 10,000g at 23°C . Up to 800 μL of lysate can be loaded at once. If your volume is larger, load several times. If the cells were lysed with Buffer RLT⁺, the flow-through fraction contains total RNA and denatured proteins, which can be isolated subsequently. In the case of lysis with genomic lysis buffer⁺, the RNA is already degraded, but the proteins can still be precipitated for subsequent analysis. If desirable, the flow-through can be stored at -80°C for up to 5 years.

- 30 Add 400 μL of RNase wash buffer to each spin column and incubate for 10–15 min to make sure that the residual RNA is degraded. After incubation, centrifuge for 2 min at 10,000g at 23°C , using a new collection tube.

▲ CRITICAL STEP The amount of RNA varies greatly from cell type to cell type. RNA impurities can be separated from DNA modifications in the UHPLC–MS/MS analysis. However, it is difficult to estimate the amount of isolated gDNA if the sample is contaminated with a substantial amount of RNA.

? TROUBLESHOOTING

- 31 From now on, either discard the flow-through by pipetting it out of the collection tube or use a new collection tube for each step.

▲ CRITICAL STEP Take care when removing the collecting tubes of the centrifuge. The flow-through should neither touch the end of the spin column nor become trapped between the spin column and the collection tube, as this might lead to alcohol or salt contaminations in the eluate that will massively impair the subsequent UHPLC–MS/MS analysis.

- 32 Add 400 μL of DNA pre-wash buffer per spin column and centrifuge for 1 min at 10,000g at 23°C . Discard the flow-through.

? TROUBLESHOOTING

- 33 Add 600 μL of gDNA wash buffer per spin column and centrifuge for 1 min at 10,000g at 23°C . Discard the flow-through and repeat this step at least once.

? TROUBLESHOOTING

- 34 Put the column into a new collection tube and centrifuge for 1 min at 10,000g at 23°C to remove all the wash buffer.

- 35 To elute, set the spin column into a new 1.5-mL centrifuge tube, add 50–150 μL of water supplemented with 0.001 equiv. of 100 \times BHT directly to the matrix of the column and incubate for 10 min. Centrifuge for 2 min at 10,000g at 23°C .

- 36 Measure the gDNA concentration with a photometer. The optical density (OD)₂₆₀/OD₂₈₀ ratio should be between 1.85 and 1.90, and the OD₂₆₀/OD₂₃₀ ratio should be >2 , ideally between 2.3 and 2.5.

▲ CRITICAL STEP If a sample is contaminated with RNA, the OD₂₆₀/OD₂₈₀ ratio is often >1.9 . If the OD₂₆₀/OD₂₃₀ ratio is too low, the sample probably contains salt impurities. High concentrations of BHT can also result in a low OD₂₆₀/OD₂₃₀ ratio, but they do not impair the UHPLC–MS/MS analysis. Note: If the amount of RNA impedes determination of DNA content, PicoGreen can be used to clearly determine the content of double-stranded DNA.

? TROUBLESHOOTING

■ PAUSE POINT Isolated gDNA can be kept at -80°C for an infinite amount of time. For short-time storage, -20°C (1–7 d) or even 4°C (up to 1 d) is suitable. If you freeze your isolated gDNA, thaw your samples on ice before starting the DNA digestion steps. After thawing, vortex the samples thoroughly before use and do a quick spin (5,000g, 23°C , 5 s) in order to collect the solution at the bottom.

Table 5 | Example of calculating the amount of gDNA and water for the DNA digestion

Sample ID	c (ng/μL)	m (μg)	V (H ₂ O) (μL)	V (DNA) (μL)	Sample no.
1_Blank 1	—	—	35.0	0	1
2_Blank 2					2
3_Blank 3					3
4_1.DNA_1	444.61	10.0	12.5	22.5	4
5_1.DNA_2					5
6_1.DNA_3					6
7_2.DNA_1	405.34	10.0	10.3	24.7	7
8_2.DNA_2					8
9_2.DNA_3					9

V, volume.

Box 2 | Testing digestion conditions for an unknown sample ● Timing 8 h

- 1 Produce a single technical replicate, instead of a technical triplicate as described in Step 37, by digesting 1–5 μg of DNA. Then continue with Step 38.
- 2 Check if all requirements of the protocol are fulfilled:
 - (i) Complete digestion of DNA,
 - (ii) All MOIs and MOI*s are detectable,
 - (iii) All MOIs can be evaluated with the existing calibration curves.

If there are problems with (i), your sample contains impurities (e.g., salt) or shows an unfavorable pH. See the Troubleshooting section (incomplete digestion).

If there are problems with (ii):

Neither the MOI nor the corresponding MOI* is detectable: see the Troubleshooting section (peak shifting).

Only the MOI*, but not the corresponding MOI, is detectable: increase the amount of DNA and check the pH.

If there are problems with (iii): first, try to adjust the amount of DNA and/or the amount of MOI*. If this does not solve the problem, generate a new calibration curve.

? TROUBLESHOOTING

DNA digestion ● Timing 7 h or overnight

37 Create a digestion sheet (Supplementary Methods 3) to easily calculate the required volume of DNA and water per sample. Each DNA sample should be processed from now on as a technical triplicate. Your sheet starts with samples 1–3 as digestion blanks, which contain no DNA, but to which mastermixes 1 and 2 will be added. For all the other samples, DNA and water should have a total volume of 35 μL (Table 5). Number the samples consecutively, as these will help you to identify your samples afterwards.

▲ CRITICAL STEP The volume of DNA must not exceed 35 μL per sample.

38 Start with the calculated volume of water, and add the DNA according to the digestion sheet.

▲ CRITICAL STEP Use fresh water for each digestion. Autoclaved water or water that has been stored for a long time might contain dust or additional contaminations that could impair the digestion or the measurements.

▲ CRITICAL STEP If this is the first digestion of your DNA sample and/or you are interested in a new DNA modification, you should consider testing your digestion protocol (Box 2). This may help to determine the appropriate addition of labeled nucleosides, e.g., spiking and the proper amount of DNA in general for fulfilling the requirements of the calibration curves.

39 Create a pipetting scheme for your mastermixes 1 and 2, using our provided Excel template (Supplementary Methods 3). Calculate the number of DNA samples (Step 37) you want to digest and add three samples more as spares.

▲ CRITICAL STEP Instead of following the two-step digestion protocol, which was designed to give an optimal SNR in the MS, you can use Degradase Plus from Zymo Research, which requires only one digestion step. See Box 3 for details.

Box 3 | DNA digestion with Degradase Plus (replaces Steps 39–42) ● Timing 5 h

- 1 Create a pipetting scheme for the Degradase mastermix using our provided Excel template for Degradase digestion (see the tabs for 'Degradase digestion' in Supplementary Methods 3). Calculate the number of DNA samples and add at least one more sample as a spare.
- 2 Prepare the Degradase mastermix, add 15.0 μL to each sample and incubate for 4 h at 37 °C. When using Degradase Plus, adjust the enzyme concentration if >5 μg of DNA/sample must be digested (Supplementary Methods 3). The additional use of benzonase is required for residual RNA content within your DNA sample. The enzyme mixture in Degradase Plus cannot digest RNA. If benzonase is not used, this will lead to an overpressure in the LC device due to residual undigested RNA, and subsequent clogging of the column. However, there is no need to add the enzyme benzonase in the case of digestion of pure DNA, e.g., synthetic DNA oligonucleotides.

- 40 Prepare mastermix 1, then add 7.5 μL of it to each sample and incubate for 3 h at 37 °C.
- 41 Before the incubation time for Step 40 runs out, start preparing mastermix 2.
- 42 Add 7.5 μL of mastermix 2 to each sample and incubate for 3 h at 37 °C.
 - ▲ **CRITICAL STEP** If >5 μg of DNA/sample is digested, incubate overnight at 37 °C.
 - **PAUSE POINT** Samples can be stored at 4 °C (overnight) or at –20 °C for long-term (up to 5 years) storage.

Sample preparation for UHPLC-MS/MS analysis ● Timing 1 h

- 43 Tape an AcroPrep Advance 96-well Supor plate (0.2 μm) to a skirted 96-well PCR plate.
- 44 After the digestion, pipette the full 50 μL of digestion mixture for each sample into a separate well of the assembled filtration plate.
- 45 Cover the filtration plate with Parafilm M and secure it with tape.
 - ▲ **CRITICAL STEP** Water should not enter the plate from an external source, but the Parafilm M should not seal the plate airtight.
- 46 Centrifuge the filtration plate, using a plate rotor for 35 min at 3,220g at 4 °C.
 - ? **TROUBLESHOOTING**
- 47 Transfer 43 μL of the filtered digestion mixture to an HPLC vial and close the vial with an HPLC vial cap. If the volume of the filtered digestion mixture is <43 μL , note which volume is transferred to the vial.
 - ▲ **CRITICAL STEP** The unused wells of the filtration plate can be used for later filtrations. For storage, seal the filtration plate, which is taped to the 96-well PCR tube plate, with Parafilm M and store it at room temperature for an infinite amount of time.
- 48 For each sample from Step 47, inject 39 μL of sample into the UHPLC-MS/MS system when the machine is ready. Record the signal of the MOIs with the relevant method for your MOIs (Supplementary Methods 4 for epigenetic modifications or Supplementary Methods 5 for m^6dA - and m^4dC -sensitive or Supplementary Methods 6 for m^6dA - and m^4dC -insensitive). For details about the measurement methods, see Table 1 and Supplementary Tables 3, 5 and 7.
 - ▲ **CRITICAL STEP** Inject 3 μL less than the transferred volume, if the transferred volume is <43 μL .
 - ? **TROUBLESHOOTING**

Evaluation ● Timing 20 min–1.5 h

- 49 Create a new folder called 'QuantResults' in the folder with the measured data.
- 50 Open the program QQQ Quantitative Analysis and click on 'File' > 'New Batch' and label the batch according to your experiment name/number.
- 51 Import all samples of interest.
- 52 Set up a method for analysis. Click on 'Method' > 'New' > 'New Method' under 'Acquired MRM Data' > 'Select a sample which was measured with this method' > 'Method Setup Tasks' > 'MRM Compound Setup'. Add UV, UV-dG and UV-dC as new compounds in the table (right-click on the table). Set the time segment 'TS' to '3' and the 'Type' to 'Target'. Enter the appropriate RT in the 'RT setup'. For UV, set it to 6 min with a 'Left RT Delta' of 6 min and a 'Right RT Delta' of 6 min. Go next to 'ISTD Setup'. In the 'ISTD conc.' column, add '1' for the ISTD concentration, make a tick in the 'ISTD Flag' column for each ISTD and select the appropriate ISTD for each target in the 'ISTD Compound Name' column. Finally, select 'Advanced Tasks' > '2D Compound Setup' and then select 'VWD' for the compounds UV, UV-dC and UV-dG. Click on 'Validate' to ensure that

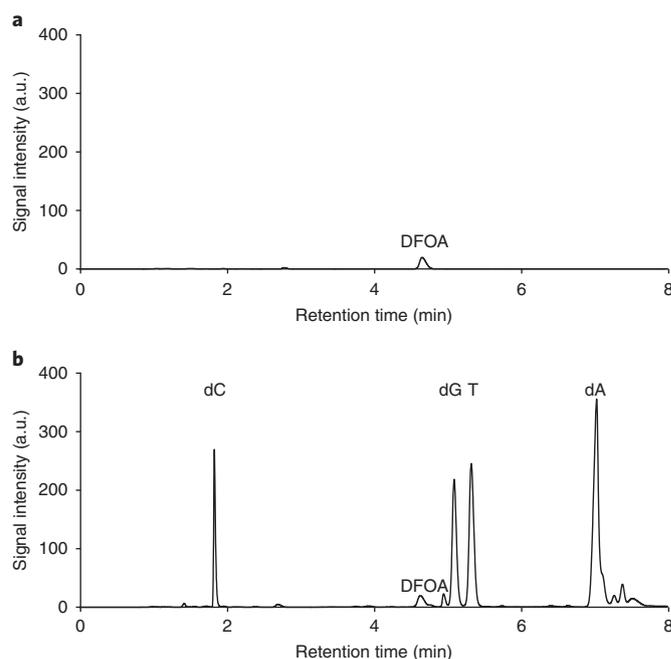


Fig. 4 | Chromatogram of different UV traces as represented in the QQQ quantitative analysis program. a,b, The chromatogram of a digestion blank sample (a) and the chromatogram of a gDNA sample from mESC line J1 (b).

your method does not have any errors. If errors or warnings occur, the program will tell you where the problem is.

▲ CRITICAL STEP The protocol described here uses the provided Excel sheets (Supplementary Methods 1 and 2) for quantification. Therefore, the concentration that you define in the ‘ISTD conc.’ column is not relevant.

53 Click on ‘Exit’ and let the program analyze your batch.

54 Select at first for the compound ‘UV’: click on ‘manual integration’ and start with the integration of dG (RT ~5.3 min for method with buffer B: MeCN or ~8.3 min for method with buffer B: MeOH) for all samples. Click in the table, then select ‘File’ > ‘Export’ > ‘Export Table’ (this will open an Excel sheet, which should be saved as ‘dG’). Continue with the integration of the dC Peak (RT ~1.9 min for method with buffer B: MeCN or ~1.8 min for method with buffer B: MeOH).

▲ CRITICAL STEP In the case that you also need dA (RT ~7.3 min for method with buffer B: MeCN or ~9.1 min for method with buffer B: MeOH) and T (RT ~5.5 min for method with buffer B: MeCN or ~8.8 min for method with buffer B: MeOH), additionally integrate those peaks.

▲ CRITICAL STEP The digestion blanks should not show any peaks for dA, dC, dG or T in the UV trace. If you see any peaks at the anticipated RT, your blanks are probably contaminated with DNA, and the whole measurement is invalid. Peaks at different RTs, not overlapping with dA, dC, dG and T, in the UV trace of the digestion blank are impurities or small molecules from the digestion mix (e.g., DFOA) that can be ignored (Fig. 4). RNA impurities in the DNA samples should not overlap with dA, dC, dG and T peaks either.

55 In the Batch table, select the first compound measured in the Signal type MS. The upper row shows the peak for the natural nucleoside (MOI), e.g., the one defined as Target and the lower row shows the peak for the labeled standard (MOI*), e.g., the one defined as ISTD (Fig. 5). Integrate both peaks the same way. This is very important for the reliability of the method. Export the Batch table. This will create an Excel sheet automatically, which you should label according to the MOI.

▲ CRITICAL STEP Carefully integrate the peaks of your digestion blanks, e.g., samples 1-3, even if they do not contain any DNA. These samples will serve as your negative control and will reveal the limits of quantification and detection; see Experimental setup.

56 Repeat this procedure for all nucleosides (MOI) of interest (Fig. 6).

? TROUBLESHOOTING

57 Use the provided Excel sheet (Supplementary Methods 1 or 2) for the evaluation of your data. For this, you need to copy the values of the peak areas (for A (MOI) and A* (MOI*)) resulting from the

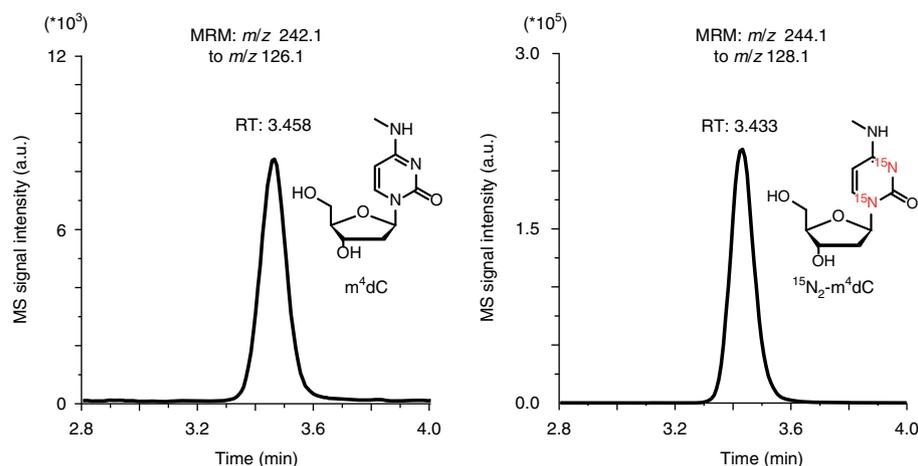


Fig. 5 | Chromatograms for m⁴dC and ¹⁵N₂-m⁴dC. The left panel shows the nucleoside signal for the natural m⁴dC and the right panel shows the nucleoside signal for the internal standard ¹⁵N₂-m⁴dC.

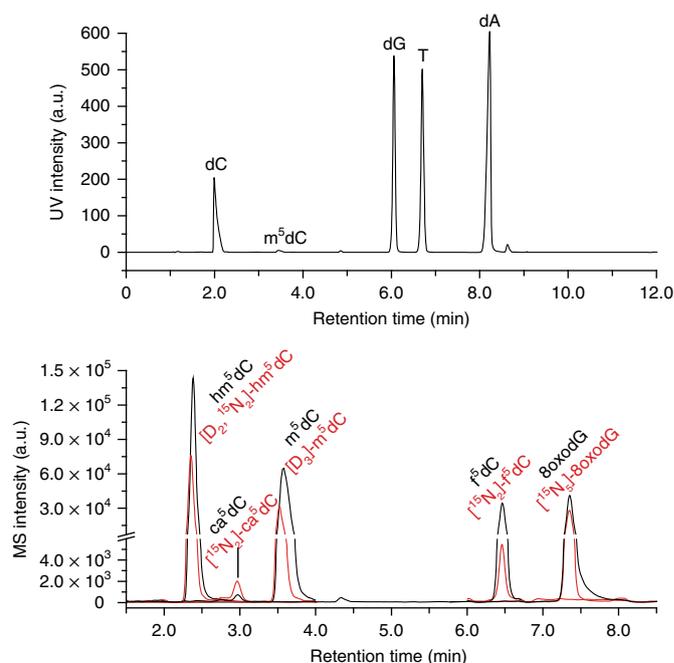


Fig. 6 | Chromatogram showing the analysis of hm⁵dC, ca⁵dC, m⁵dC, f⁵dC and 8oxodG with their corresponding isotopically labeled standards. The upper chromatogram shows the UV trace and the lower chromatogram shows the MS trace¹¹.

integration of Steps 54–56 from the created Excel sheet into the evaluation Excel sheet (Supplementary Methods 1 or 2).

▲ CRITICAL STEP Before processing the data, you must calculate the LOD of every single MOI by analyzing the SNR. Furthermore, the ratios of A/A^* and thereby the amount n of the MOI need to meet the requirements of the corresponding calibration curve. Valid data that are consistent with these conditions can be processed further and will provide the amount of the MOI in pmol within one sample, more specifically within the volume of injection (39 μ L). If you did not inject 39 μ L for the measurement of a sample, you need to adjust this value in the provided Excel sheet for this sample. This additional calculation will provide comparable data for your technical and later biological replicates. The injection volume can be looked up with the QQQ Quantitative Analysis program.

Troubleshooting

Troubleshooting advice can be found in Table 6.

Table 6 | Troubleshooting table

Step	Problem/Observation	Possible reason	Solution
11	The sensitivity of the mass spectrometer is too low: The sensitivity check for the ISTD mastermix shows substantially smaller peaks than in previous measurements	(i) Dirty ionization source (ii) Old buffers (iii) Old column (iv) Molecule insoluble (pH variation)	(i) Clean ionization source (ii) Exchange buffers (iii) Exchange column (iv) Check and adjust pH (v) Check sensitivity between (i)-(iv)
11, 32, 33 and 56; Box 2	Peak shifting: The peak of interest occurs at a different retention time	(i) Impurities (e.g., salt) (ii) pH variation (iii) Column batch shows different properties	(i) Wash samples more rigorously during DNA isolation (ii) Check pH regularly (iii) Adjust time segments, if modifications escape
11, 46 and 56	Peak splitting: The chromatogram for a modification shows an unexpected number of signals; usually one is at the expected retention time and a second one elutes at an earlier retention time. The two signals might be connected	(i) The filtration was not successful (ii) The inlet filter is clogged (iii) The nucleoside was only partially charged during the chromatography and is therefore eluting at two different time points	(i) Optimize the filtration step (ii) Use guard columns or inlet filters (iii) Check the pH of the sample
11 and 56	Peak broadening	(i) Amounts of modification are too high (ii) Not enough data points in peak(s) (iii) Poor performance of the column	(i) Try setting 'MS1 resolution' or 'MS2 resolution' to 'enhanced' in the method (ii) Decrease dwell time to receive more cycles per second (iii) Test a new column
	Two molecules cannot be baseline-separated: The signal for the MOI shows at least a shoulder or a clear second maximum	(i) Similar retention time due to physical properties of the molecule (often in the UV trace, but also possible with MS analysis; e.g., in isomers) (ii) Altered performance of the column, which is possible after change of the column batch	(i) Switch the machine setup by testing a different column or different buffer system (ii) Adjust column temperature and check whether molecules with the same mass signal still elute in the same order
27 and 36	Low yield of DNA: Yield of DNA after isolation is lower than expected	(i) Insufficient DNA shearing (ii) Spin column did not bind DNA (iii) Spin column was overloaded (iv) Insufficient DNA elution from column	(i) Use more lysis buffer (ii) Test a different batch of spin columns (iii) Distribute the lysate on more spin columns (iv) Increase the volume for elution or elute twice (i,iv) Use the Blood & Cell Culture DNA Midi Kit from Qiagen (anion-exchange columns allow elution of larger gDNA fragments than spin columns)
30, 36 and 56	Too much RNA in DNA preparation: (i) The ratios of the absorption at 260 nm and 280 nm are >2 (ii) The UV chromatogram shows additional peaks	(i) Cell line contains higher amounts of RNA (ii) The amount or performance of the RNase was insufficient	Adjust the amount of RNase during the DNA isolation and increase the incubation time
36	Low DNA concentration: DNA volume for digestion exceeds 35 µL	Elution volume too high	Precipitate DNA (Supplementary Methods 7) and re-dissolve in a smaller volume
48, Box 2	The pressure of the column increases during one sample set: Between samples of the same dataset, an increase in the column pressure is observed	(i) Incomplete digestion (ii) Salt impurities	(i) Exchange the inlet filter (ii) Add runs without injection to the worklist
56, Box 2	Incomplete digestion: (i) No or substantially smaller signals for the canonical nucleosides are observed in the UV trace (ii) Sometimes additional UV signals appear at early (-1 min) or late (-10 min) retention times (iii) The pressure of the column increases due to clogging	(i) Enzyme performance is reduced (ii) Enzyme performance is reduced and therefore residual single nucleotides or oligonucleotides exist (iii) pH variation of the sample and/or salt impurities	(i) Test enzymes regularly with appropriate amounts of DNA (ii) See (i) (iii) Test pH compatibility of sample and enzyme and/or wash DNA more often with DNA pre-wash and gDNA wash buffer during the isolation

Timing

Steps 1–4, general procedure for dissolution and dilution of nucleosides: 30 min
 Steps 5–11, preparation of ISTD mastermix and UHPLC–MS/MS system, and performance of sensitivity checks: 2–3 h
 Steps 12–24, calculation of calibration curves: 1 d
 Steps 25–36, isolation of gDNA: 1.5–3 h
 Steps 37–42, DNA digestion: 7 h or overnight
 Steps 43–48, sample preparation for UHPLC–MS/MS analysis: 1 h
 Steps 49–57, evaluation: 20 min–1.5 h
 Box 1, use of microscales for dilution of nucleosides: 15–30 min
 Box 2, testing digestion conditions for an unknown sample: 8 h
 Box 3, DNA digestion with Degradase Plus: 5 h

Anticipated results

The yield of gDNA depends on the cell/tissue type and isolation method. We adapted the procedure delineated in the Quick-gDNA MidiPrep Kit in combination with Zymo-Spin IIC-XL spin columns from Zymo Research. Using this method, ~6 or 10 μg of gDNA can be isolated from 10^6 human embryonic kidney 293T (HEK293T) cells or 10^6 mESCs, respectively. The yield of gDNA from 1 mg of mouse cerebellum will be ~1 μg per mg of tissue. The maximal binding capacity of the Zymo-Spin IIC-XL spin columns is 20–25 μg of gDNA. If more gDNA is required, we recommend using a proportional number of columns and combining the eluates at the end. Earlier, we had successfully used spin columns from Zymo Research with higher binding capacity. These columns yielded DNA of good quality, but recent production batches seemed to contain impurities, which perturbed our LC–MS/MS setup, thus leading to a loss of sensitivity. Note, however, that the sensitivity to impurities may vary with the specific UHPLC–MS/MS equipment. As an alternative to spin column–based isolation, we have successfully used gravity flow columns with anion exchange properties (Qiagen Blood & Cell Culture DNA Midi Kit). This method is recommended for abasic site analysis⁴⁴, as it generates lower background quantities of abasic sites, probably as a result of the gentler elution procedure, but it takes much more time than the spin column–based method. An additional advantage of the spin column–based method is that total RNA and (denatured) protein can be subsequently isolated from the very same lysate.

The protocol described above details the steps needed to quantify DNA modifications and lesions such as m^5dC , hm^5dC or 8oxodG in mammalian cells are probably between 0.07 and 0.9×10^{-6} 8oxodG/dN, based on the average of the results from several laboratories and various methods⁶⁸. The background level of 8-oxodG has been estimated at 0.5 lesions per 10^{-6} dN^{69,70}. Depending on the modification, there are big differences in the abundance of certain modifications between different cell lines or culture conditions. Whereas the abundance of m^5dC for most

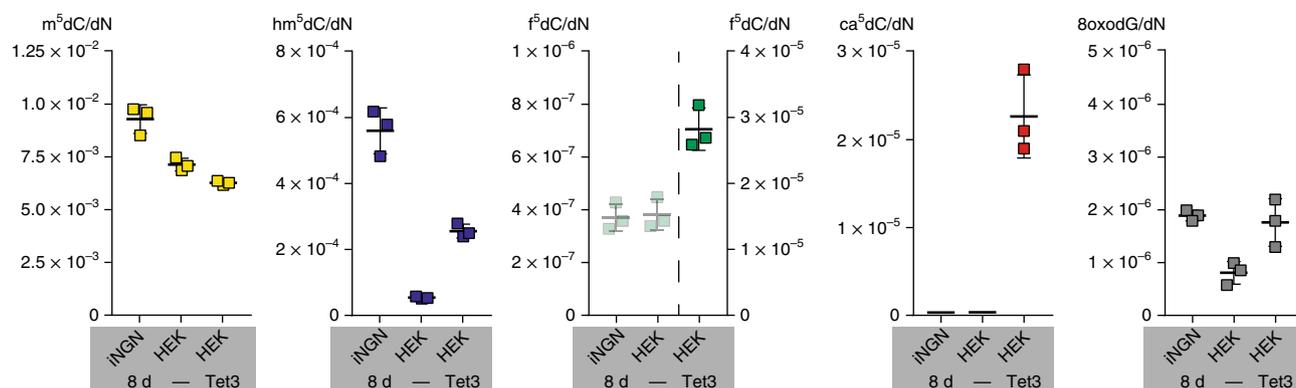


Fig. 7 | Quantities of m^5dC , hm^5dC , f^5dC , ca^5dC and 8oxodG in small molecule–inducible neurogenin human-induced pluripotent stem cells lines (iNGNs) (8 d after induction of differentiation), untransfected HEK293T cells and HEK293T cells that were transfected for 24 h with a plasmid coding for Tet3. The transparent data points at f^5dC for iNGNs and untransfected HEK293T cells indicate that f^5dC was above the LLOQ but did not meet the strict $\text{LOD}_{\text{practical}}$ criterion. Each data point represents the mean of one biological replicate that was measured in three technical replicates; the bar represents the mean of the biological replicates and error bars represent s.d.

Table 7 | Measurements for an exemplary calibration curve for m⁴dC in technical triplicates

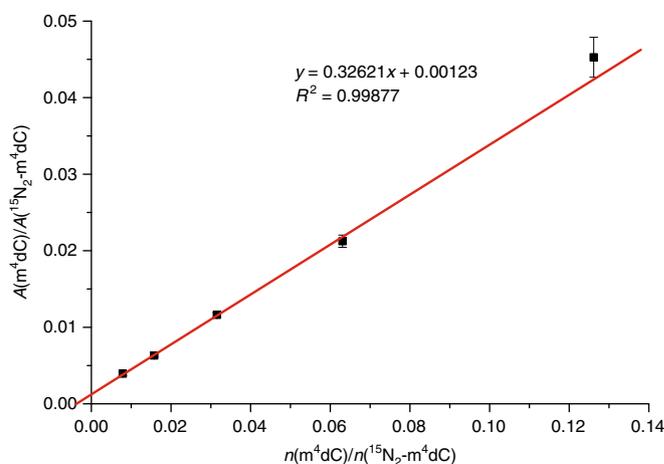
Calibration level	A(m ⁴ dC)	A*(¹⁵ N ₂ -m ⁴ dC)	A/A*	Replicate
L1	4,797	1,745,218	0.0027	1
L2	7,213	1,737,978	0.0042	
L3	10,664	1,692,721	0.0063	
L4	19,426	1,670,953	0.0116	
L5	34,395	1,677,652	0.0205	
L6	74,308	1,646,471	0.0451	
L7	164,901	1,566,133	0.1053	
L8	379,446	1,467,230	0.2586	
L9	869,825	1,340,740	0.6488	
L10	1,960,129	1,181,613	1.6589	
L11	4,164,191	971,267	4.2874	
L1	5,094	1,611,819	0.0032	2
L2	6,145	1,658,872	0.0037	
L3	10,207	1,658,337	0.0062	
L4	18,928	1,586,681	0.0119	
L5	34,678	1,570,039	0.0221	
L6	73,600	1,534,700	0.0480	
L7	163,872	1,558,445	0.1052	
L8	368,404	1,454,277	0.2533	
L9	875,082	1,321,236	0.6623	
L10	1,914,969	1,189,553	1.6098	
L11	4,264,798	885,407	4.8168	
L1	4,728	1,689,815	0.0028	3
L2	6,595	1,680,195	0.0039	
L3	10,841	1,677,259	0.0065	
L4	18,867	1,670,998	0.0113	
L5	35,859	1,696,961	0.0211	
L6	72,444	1,694,990	0.0427	
L7	169,019	1,543,636	0.1095	
L8	386,871	1,510,771	0.2561	
L9	897,343	1,315,416	0.6822	
L10	1,994,610	1,225,289	1.6279	
L11	4,063,788	944,247	4.3037	

mammalian cell lines or tissues is normally in a range between 0.6×10^{-2} and 1.0×10^{-2} m⁵dC/dN, modifications that are less abundant can deviate by factors of hundreds or thousands, depending on the biological background of the sample (e.g., organism, culture conditions).

Figure 7 shows typical quantities of m⁵dC, hm⁵dC, f⁵dC, ca⁵dC and 8oxodG for iNGNs obtained 8 d after differentiation, for untransfected HEK293T cells and HEK293T cells that were transfected with a plasmid coding for Tet3 and incubated for an additional 24 h. For all the cell lines shown, 2 µg of gDNA was used per technical replicate for the digestion. As m⁵dC is a very abundant DNA modification, only very little amounts of gDNA are needed for detection and quantification, and 2 µg is by far sufficient to exceed the LLOQ and the LOD_{practical}. Also, for hm⁵dC, for which the typical amount is in the range of 10^{-4} hm⁵dC/dN, 2 µg of gDNA is more than enough for the quantification, even in cell lines that have typically low quantities of this modification, such as HEK293T cells. However, for less abundant modifications, here f⁵dC and ca⁵dC, 2 µg is often not enough to meet the LLOQ or LOD_{practical} criterion. Although the detected f⁵dC levels in iNGNs and untransfected HEK293T cells were above the LLOQ, these results cannot be used for further interpretation, because in this case the background signal was too high and the samples therefore did not meet the LOD_{practical} criterion. For ca⁵dC, the abundance was above neither the LOD_{practical} nor the LLOQ. One possible way to overcome this problem is to use more gDNA for the quantification. If you are interested only in a certain modification, you can also try to optimize the UHPLC–MS/MS system for

Table 8 | Calculations for an exemplary m⁴dC calibration curve

Calibration level	Mean (A/A*)	s.d. (A/A*)	%s.d. (A/A*)	Backfit (invalid)	Accuracy (invalid)	Backfit	Accuracy
L1	0.0029	0.0002	7.7	0.0010	35	0.0025	
L2	0.0039	0.0002	5.7	0.0030	76	0.0038	105
L3	0.0063	0.0002	2.4	0.0069	110	0.0064	99
L4	0.0116	0.0003	2.8	0.0148	128	0.0115	101
L5	0.0212	0.0008	3.8	0.0306	144	0.0218	97
L6	0.0453	0.0026	5.8	0.0622	137	0.0424	107
L7	0.1066	0.0025	2.3	0.1254	118	0.0835	
L8	0.2560	0.0026	1.0	0.2519	98	0.1658	
L9	0.6644	0.0168	2.5	0.5047	76	0.3305	
L10	1.6322	0.0248	1.5	1.0103	62	0.6597	
L11	4.4693	0.3010	6.7	2.0216	45	1.3181	

**Fig. 8 |** Exemplary calibration curve for m⁴dC. Three technical replicates were measured per data point. The mean is plotted and error bars represent s.d.

this modification; however, this might impair the detection of other modifications. It is not possible to avoid completely the generation of 8oxodG during the whole sample preparation process, and cells always show a low background of oxidative damage. However, the 8oxodG quantity (unless cells were put under (oxidative) stress) should not exceed the low 10^{-6} 8oxodG/dN range.

Evaluation of calibration curve

Table 7 shows an exemplary evaluation of the calibration curve of m⁴dC as described in the Procedure. Consideration of all data points for linear regression would result in the following equation:

$$y = 0.50101x - 9.60772 \times 10^{-4}.$$

Applying a so-called backfit (Step 23) results in the values shown in Table 8, column 5. These values then must be divided by the mean for A/A* (Table 8, column 2) and multiplied by 100 to provide the accuracy of the equation (Table 8, column 6) as a percentage. As described in Step 23, the accuracy must be between 80 and 120%, but with consideration of all data points, e.g., levels, the accuracy varies highly and does not meeting the requirements. Therefore Step 23 must be applied.

In this case, Step 24 had to be performed multiple times to yield a calibration curve consisting only of data points with sufficient accuracy and therefore a valid linear equation (Fig. 8). Calibration levels 2-6 (see Table 8, columns 7 and 8) provide the lower and upper limits of quantification for n of 18.9

and 302 fmol, respectively, with the lower and upper limits for A/A^* in the range between 0.0042 and 0.0451.

Ethics statement

Tissue material was provided by the animal facility of the Department of Pharmacy at Ludwig-Maximilians-Universität München, which is approved by the government of Upper Bavaria.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary.

References

1. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
2. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
3. Pfaffeneder, T. et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
4. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
5. He, Y.-F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
6. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Gene Dev.* **25**, 2436–2452 (2011).
7. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).
8. Langemeijer, S. M. C. et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* **41**, 838–842 (2009).
9. Ko, M. et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
10. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).
11. Pfaffeneder, T. et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–581 (2014).
12. Bachman, M. et al. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (2015).
13. Iwan, K. et al. 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nat. Chem. Biol.* **14**, 72–78 (2018).
14. Raiber, E.-A. et al. 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–49 (2014).
15. Song, C. X. et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
16. 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).
17. 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).
18. Hill, P. W. S., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324–333 (2014).
19. Yu, M. et al. Base-resolution detection of N(4)-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite sequencing. *Nucleic Acids Res.* **43**, e148 (2015).
20. Ehrlich, M. et al. DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. *Nucleic Acids Res.* **13**, 1399–1412 (1985).
21. Arber, W. & Dussoix, D. Host specificity of DNA produced by *Escherichia coli*: I. Host controlled modification of bacteriophage λ . *J. Mol. Biol.* **5**, 18–36 (1962).
22. Wu, T. P. et al. DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
23. Huang, W. et al. Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. *RSC Adv.* **5**, 64046–64054 (2015).
24. Koziol, M. J. et al. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30 (2015).
25. Liu, J. et al. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat. Commun.* **7**, 13052 (2016).
26. Schiffrers, S. et al. Quantitative LC-MS provides no evidence for m⁶dA or m⁴dC in the genome of mouse embryonic stem cells and tissues. *Angew. Chem. Int. Ed. Engl.* **56**, 11268–11271 (2017).
27. Wang, D., Kreuzer, D. A. & Essigmann, J. M. Mutagenicity and repair of oxidative DNA damage: insights from studies using defined lesions. *Mutat. Res.* **400**, 99–115 (1998).

28. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214 (2003).
29. Lunec, J. ESCODD: European standards committee on oxidative DNA damage. *Free Radical Res.* **29**, 601–608 (1998).
30. Murray, K. K. et al. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure Appl. Chem.* **85**, 1515–1609 (2013).
31. Oakeley, E. J. DNA methylation analysis: a review of current methodologies. *Pharmacol. Ther.* **84**, 389–400 (1999).
32. Taghizadeh, K. et al. Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nat. Protoc.* **3**, 1287–1298 (2008).
33. Annesley, T. M. Ion suppression in mass spectrometry. *Clin. Chem.* **49**, 1041–1044 (2003).
34. Buhrman, D. L., Price, P. I. & Rudewiczcor, P. J. Quantitation of SR 27417 in human plasma using electrospray liquid chromatography–tandem mass spectrometry: a study of ion suppression. *J. Am. Soc. Mass Spectrom.* **7**, 1099–1105 (1996).
35. Brückl, T., Globisch, D., Wagner, M., Müller, M. & Carell, T. Parallel isotope-based quantification of modified tRNA nucleosides. *Angew. Chem. Int. Ed. Engl.* **48**, 7932–7934 (2009).
36. Münzel, M. et al. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
37. Globisch, D. et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
38. Pearson, D. et al. LC-MS based quantification of 2'-ribosylated nucleosides Ar(p) and Gr(p) in tRNA. *Chem. Commun.* **47**, 5196–5198 (2011).
39. Globisch, D. et al. Systems-based analysis of modified tRNA bases. *Angew. Chem. Int. Ed. Engl.* **50**, 9739–9742 (2011).
40. Wetzel, C. & Limbach, P. A. Global identification of transfer RNAs by liquid chromatography–mass spectrometry (LC-MS). *J. Proteomics* **75**, 3450–3464 (2012).
41. Schiesser, S. et al. Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
42. Schröder, A. S. et al. Synthesis of a DNA promoter segment containing all four epigenetic nucleosides: 5-methyl-, 5-hydroxymethyl-, 5-formyl-, and 5-carboxy-2'-deoxycytidine. *Angew. Chem. Int. Ed. Engl.* **53**, 315–318 (2014).
43. Thienpont, B. et al. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature* **537**, 63–68 (2016).
44. Rahimoff, R. et al. 5-Formyl- and 5-carboxydeoxycytidines do not cause accumulation of harmful repair intermediates in stem cells. *J. Am. Chem. Soc.* **139**, 10359–10364 (2017).
45. Turowski, M. et al. Deuterium isotope effects on hydrophobic interactions: the importance of dispersion interactions in the hydrophobic phase. *J. Am. Chem. Soc.* **125**, 13836–13849 (2003).
46. Kanao, E. et al. Isotope effects on hydrogen bonding and CH/CD– π interaction. *J. Phys. Chem.* **122**, 15026–15032 (2018).
47. Liu, S. & Wang, Y. Mass spectrometry for the assessment of the occurrence and biological consequences of DNA adducts. *Chem. Soc. Rev.* **44**, 7829–7854 (2015).
48. Lentini, A. et al. A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nat. Methods* **15**, 499–504 (2018).
49. Shrivastava, A. & Gupta, V. B. HPLC: isocratic or gradient elution and assessment of linearity in analytical methods. *J. Adv. Sci. Res.* **3**, 12–20 (2012).
50. Schellinger, A. P. & Carr, P. W. Isocratic and gradient elution chromatography: a comparison in terms of speed, retention reproducibility and quantitation. *J. Chromatogr. A* **1109**, 253–266 (2006).
51. Zhang, J. J. et al. Analysis of global DNA methylation by hydrophilic interaction ultra high-pressure liquid chromatography tandem mass spectrometry. *Anal. Biochem.* **413**, 164–170 (2011).
52. Contreras-Sanz, A. et al. Simultaneous quantification of 12 different nucleotides and nucleosides released from renal epithelium and in human urine samples using ion-pair reversed-phase HPLC. *Purinergic Signal.* **8**, 741–751 (2012).
53. Yonekura, S., Iwasaki, M., Kai, M. & Ohkura, Y. High-performance liquid chromatographic determination of guanine and its nucleosides and nucleotides in biospecimens with precolumn fluorescence derivatization using phenylglyoxal. *Anal. Sci.* **10**, 247–251 (1994).
54. Giel-Pietraszuk, M. et al. Quantification of 5-methyl-2'-deoxycytidine in the DNA. *Acta Biochim. Pol.* **62**, 281–286 (2015).
55. Pitt, J. J. Principles and applications of liquid chromatography–mass spectrometry in clinical biochemistry. *Clin. Biochem. Rev.* **30**, 19–34 (2009).
56. Niessen, W. M. A. State-of-the-art in liquid chromatography–mass spectrometry. *J. Chromatogr. A* **856**, 179–197 (1999).
57. Williamson, L. N. & Bartlett, M. G. Quantitative liquid chromatography/time-of-flight mass spectrometry. *Biomed. Chromatogr.* **21**, 567–576 (2007).
58. Payne, A. H. & Glish, G. L. Tandem mass spectrometry in quadrupole ion trap and ion cyclotron resonance mass spectrometers. *Methods Enzymol.* **402**, 109–148 (2005).

59. Tozuka, Z. et al. Strategy for structural elucidation of drugs and drug metabolites using (MS)*n* fragmentation in an electrospray ion trap. *J. Mass Spectrom.* **38**, 793–808 (2003).
60. Ens, W. & Standing, K. G. Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules. *Methods Enzymol.* **402**, 49–78 (2005).
61. Commission Decision 2002/657/EC. *Implementation of Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results* (Brussels, Official Journal of the European Communities, 2002).
62. Zhao, Z. & Zhang, F. Sequence context analysis in the mouse genome: single nucleotide polymorphisms and CpG island sequences. *Genomics* **87**, 68–74 (2006).
63. Kellner, S. et al. Oxidation of phosphorothioate DNA modifications leads to lethal genomic instability. *Nat. Chem. Biol.* **13**, 888–894 (2017).
64. Kellner, S. et al. Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **42**, e142 (2014).
65. Amouroux, R. et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 (2016).
66. Busskamp, V. et al. Rapid neurogenesis through transcriptional activation in human stem cells. *Mol. Syst. Biol.* **10**, 760 (2014).
67. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
68. Collins, A. R., Cadet, J., Möller, L., Poulsen, H. E. & Viña, J. Are we sure we know how to measure 8-oxo-7,8-dihydroguanine in DNA from human cells? *Arch. Biochem. Biophys.* **423**, 57–65 (2004).
69. ESCODD. Comparative analysis of baseline 8-oxo-7,8-dihydroguanine in mammalian cell DNA, by different methods in different laboratories: an approach to consensus. *Carcinogenesis* **23**, 2129–2133 (2002).
70. Ravanat, J.-L. et al. Cellular background level of 8-oxo-7,8-dihydro-2'-deoxyguanosine: an isotope based method to evaluate artefactual oxidation of DNA during its extraction and subsequent work-up. *Carcinogenesis* **23**, 1911–1918 (2002).

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft for funding through SFB749, SFB1032, SPP1784 and CA275-11/1. Further support is acknowledged from the Excellence Cluster CiPSM (Center for Integrated Protein Science). F.R.T. thanks the Boehringer Ingelheim Fonds for her PhD fellowship. S.K. thanks the Fonds der Chemischen Industrie for the Liebig Fellowship. We thank T. Pfaffeneder for early method development and helpful input. We thank S. Michalakis (Department of Pharmacy, Ludwig-Maximilians-Universität München) for providing mouse tissues.

Author contributions

F.R.T., S.S. and K.I. designed and developed the protocol. S.K. provided expertise on mass spectrometry. F.S. developed the DNA isolation protocol. M.M. designed graphics. T.C. supervised and designed the studies. All authors participated in writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41596-018-0094-6>.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published online: 17 December 2018

Related links

Key references using this protocol

- Pfaffeneder, T. et al. *Nat. Chem. Biol.* **10**, 574–581 (2014): <https://www.nature.com/articles/nchembio.1532>
Iwan, K. et al. *Nat. Chem. Biol.* **14**, 72–78 (2018): <https://www.nature.com/articles/nchembio.2531>
Schiffers, S. et al. *Angew. Chem. Int. Ed. Engl.* **56**, 11268–11271 (2017): <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201700424>
Rahimoff, R. et al. *J. Am. Chem. Soc.* **139**, 10359–10364 (2017): <https://pubs.acs.org/doi/10.1021/jacs.7b04131>

3.2 Synthesis of (*R*)-Configured 2'-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides

Arne S. Schröder, Olga Kotljarova, Edris Parsa, Katharina Iwan, Nada Raddaoui, and Thomas Carell*

Prolog

Die Erforschung der Funktionen der neuen epigenetischen Basen ist eine große Herausforderung, vor allem da die oxidierten Derivate wie fdC und cadC durch die sogenannte Basenexzisionsreparatur (BER) ausgeschnitten werden können und somit nur in sehr geringen Mengen in der DNA zu detektieren sind. Daher benötigt man Nukleosid-Analoga, die nicht durch BER repariert werden können. Eine 2'-F Substitution der Nukleoside verhindert die Hydrolyse der glykosidischen Bindung durch Glykosylasen, die für die BER verantwortlich sind. Die Einführung des F-Substituenten an der 2' Position kann sowohl in *arabino*- als auch in *ribo*-Stellung erfolgen. Eine erste Verfütterungsstudie von 2'-(*S*)-F-dC (*arabino*) und 2'-(*R*)-F-dC (*ribo*) an murine embryonale Stammzellen (mESC) zeigte, dass ersteres zu toxisch ist. Aufgrund dessen wurden die Phosphoramidite der Nukleosid-Analoga 2'-(*R*)-F-mdC, 2'-(*R*)-F-hmdC, 2'-(*R*)-F-fdC und 2'-(*R*)-F-cadC ausgehend von 2'-(*R*)-F-dC synthetisiert. Eine anschließende Studie mit der Methyltransferase M.SssI mit einem F-dC-substituierten und einem natürlichen Oligonukleotid zeigte, dass die 2'-F Substitutionen keinen Effekt auf die biochemische Methylierung haben. Der Vergleich eines F-fdC-Oligonukleotids mit einem fdC-haltigen Oligonukleotid in einer Studie zur Aktivität der Glykosylase (hTDG) zeigte den erwarteten Effekt: Der fdC-haltige Doppelstrang wurde gespalten, während der F-fdC-haltige Doppelstrang intakt blieb. Der 2'-F Substituent schützt das Nukleosid-Analoga effektiv vor der BER und ist somit das ideale Werkzeug, um den epigenetischen Metabolismus ohne den Einfluss der BER zu untersuchen.

Autorenbeitrag

Für die massenspektrometrische Untersuchung der fluorierten Nukleoside wurde eine erste Messmethode (UHPLC-MS/MS) von mir entwickelt. Die DNA der murinen embryonalen Stammzellen wie auch die der Oligonukleotide der Methylierungsstudie wurden von mir zu Nukleosiden verdaut. Anschließend wurden diese mit der entwickelten UHPLC-MS/MS Methode von mir gemessen und ausgewertet.

Lizenz

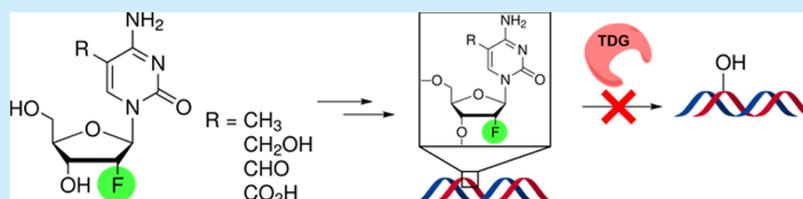
Reprinted with permission from Schröder, A.S. *et al.* Synthesis of (*R*)-Configured 2'-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides. *Organic Letters* **18**, 4368-4371 (2016). Copyright 2016 American Chemical Society.

Synthesis of (*R*)-Configured 2'-Fluorinated mC, hmC, fC, and caC Phosphoramidites and Oligonucleotides

Arne S. Schröder, Olga Kotljarova, Edris Parsa, Katharina Iwan, Nada Raddaoui, and Thomas Carell*

Center for Integrated Protein Science, Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstraße 5-13, 81377 Munich, Germany

Supporting Information



ABSTRACT: Investigation of the function of the new epigenetic bases requires the development of stabilized analogues that are stable during base excision repair (BER). Here we report the synthesis of 2'-(*R*)-fluorinated versions of the phosphoramidites of 5-methylcytosine (mC), 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and 5-carboxycytosine (caC). For oligonucleotides containing 2'-(*R*)-F-fdC, we show that these compounds cannot be cleaved by the main BER enzyme thymine-DNA glycosylase (TDG).

Fluorine is an element that is used in medicinal chemistry to replace H atoms in pharmaceutically active molecules with astonishing effects. Fluorine substitution stabilizes molecules to extend their lifetimes in the bloodstream, and often it increases the affinities of molecules for their biological targets by increasing their lipophilicities.¹ In nucleoside chemistry, for example, fluorination of dC at the 2' position creates molecules like gemcitabine (**1**), which are used as antimetabolites in modern cancer therapy.² The 2'-F substitution has several effects. Most importantly, a 2'-(*R*)-configuration as in 2'-(*R*)-F-dC (**2**) stabilizes the C3'-endo conformation of the ribose sugar so that the base becomes RNA-like.³ A fluorine at C2' also blocks the activity of glycosylases, thereby stabilizing the base during base excision repair (BER).⁴ We are currently investigating the chemistry that occurs at the nucleoside 2'-deoxycytidine (dC, **3**) that leads to the formation and removal of the methylated and subsequently oxidized epigenetic dC derivatives 5-methyl- (mdC, **4**), 5-hydroxymethyl- (hmdC, **5**), 5-formyl- (fdC, **6**), and 5-carboxy-2'-deoxycytidine (cadC, **7**) (Figure 1).⁵ Nucleosides 5–7 are products of consecutive enzymatic oxidation of **4** by the action of ten-eleven-translocation enzymes (Tet enzymes), which use molecular oxygen and α -ketoglutarate to perform the oxidation chemistry.⁶ Current data suggest that fdC and cadC are removed from the genome by BER via the enzyme thymine-DNA glycosylase (TDG).^{5d,7} Other data predict that the bases may undergo some kind of deformylation/decarboxylation reaction, which would convert fdC and cadC directly back into the canonical base dC.⁸ In order to distinguish these processes, it is important to have tool molecules that cannot be repaired by BER. This would allow one to decipher chemical processes at fdC and cadC beyond BER. Here we report the synthesis of

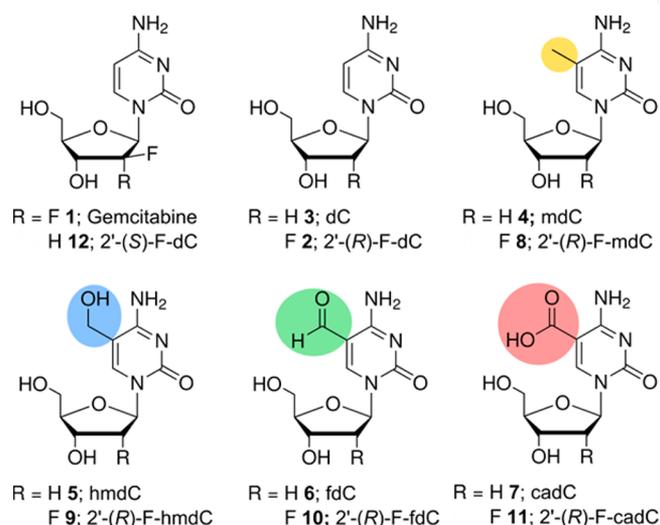


Figure 1. Overview of epigenetically relevant nucleosides and 2'-fluoro nucleosides that are important in this context.

the 2'-(*R*)-fluorinated versions of mdC (**8**), hmdC (**9**), fdC (**10**), and cadC (**11**). We have developed phosphoramidite building blocks for the incorporation of these bases into DNA strands, and we show that these nucleosides are indeed stable during BER. With the plan in mind to investigate epigenetic processes directly in the genome of stem cells, we realized that the 2'-*arabino*-configured compound 2'-(*S*)-F-dC (**12**) might be too toxic. Indeed, when we evaluated the toxicity of the *ribo*-

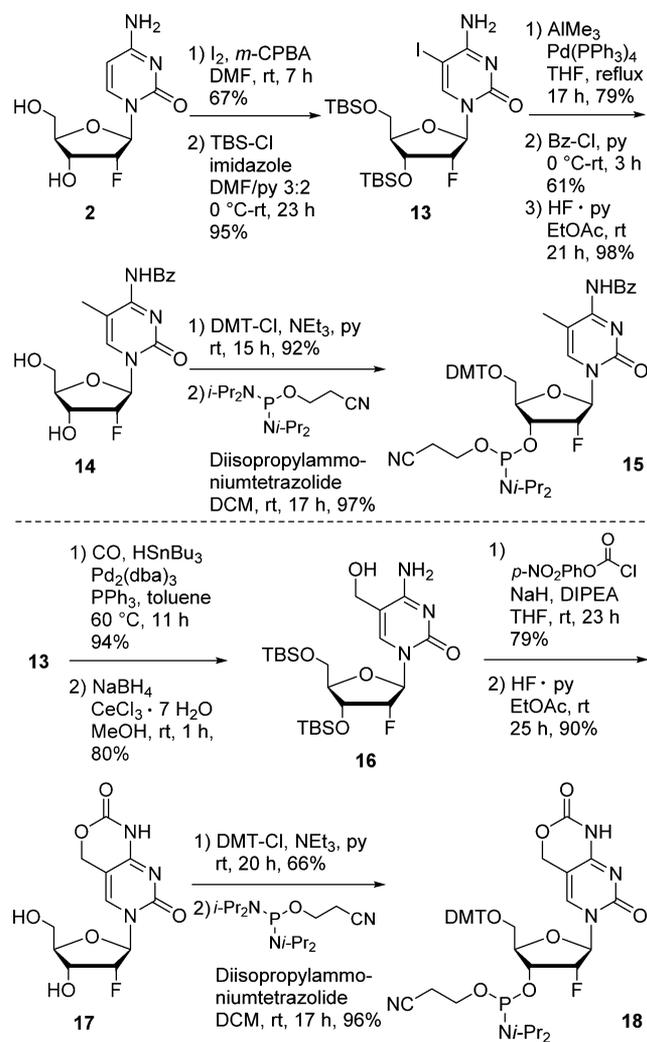
Received: July 20, 2016

Published: August 19, 2016

configured compound 2'-(R)-F-dC against **12** in stem cells (see the [Supporting Information](#)), we noted a strongly reduced toxicity for 2'-(R)-F-dC. This is already interesting because it is believed that the 2'-(S)-F configuration has a much smaller impact on the overall DNA structure.^{3,9} Our stem cell data are, however, in full agreement with toxicity studies in rats and woodchucks showing that feeding of 2'-(R)-F-dC at up to 500 mg kg⁻¹ day⁻¹ is possible without considerable toxicity effects.¹⁰

For the synthesis of the 2'-(R)-F-xdC nucleosides and phosphoramidites **15**, **18**, **21**, and **24** (see [Scheme 1](#)), we

Scheme 1. Synthesis of 2'-(R)-F-mdC and 2'-(R)-F-hmdC Phosphoramidite Building Blocks **15 and **18****



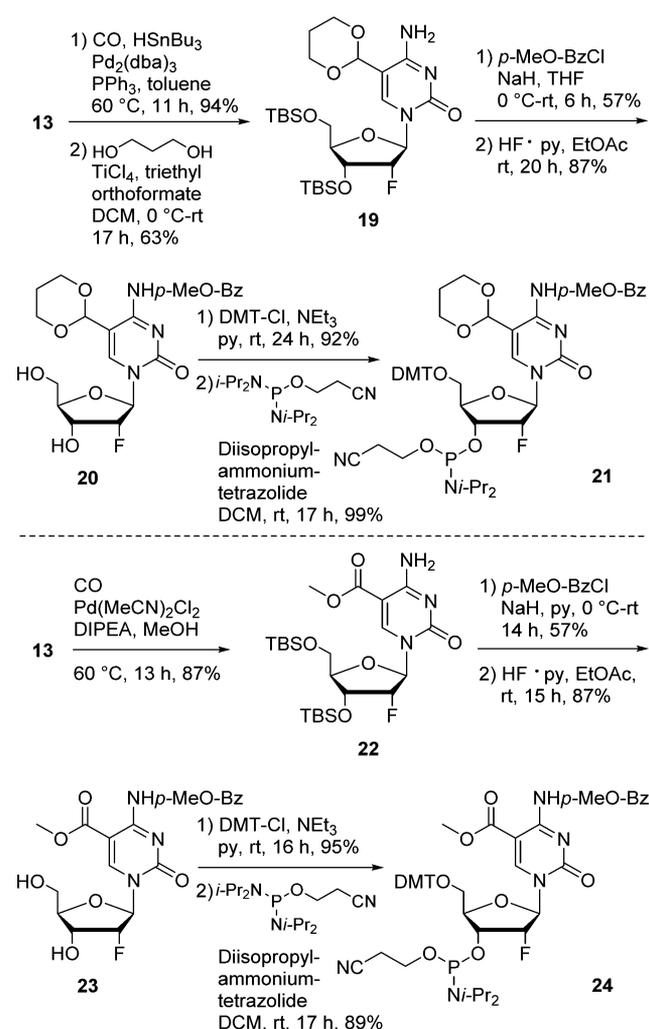
started with 2'-(R)-F-dC (**2**), which was iodinated at C5 with elemental iodine and *m*-CPBA.¹¹ Subsequent silylation yielded TBS-protected 5-iodo-2'-(R)-F-dC **13**. The needed methylation was best carried out under Kumada conditions with trimethylaluminum.¹² This furnished the 2'-(R)-F-mdC compound in 79% yield. Notably, the use of other methyl-transferring agents such as MeMgCl resulted in a 1:1 mixture of methylated and dehalogenated products. We believe that the exocyclic amine requires complete deprotonation to avoid a 1,3-proton shift from the exocyclic amine to the Pd-activated C5-position. Further protection with BzCl and silyl deprotection with Olah's reagent furnished 2'-(R)-F-mdC derivative

14, which was converted into the 2'-(R)-F-mdC phosphoramidite building block **15** using standard procedures.¹³

For the synthesis of the 2'-(R)-F-hmdC phosphoramidite **18**, we started from intermediate **13**. Carbonylative Stille coupling with tributyltin hydride and reduction of the formyl group under Luche conditions yielded 2'-(R)-F-hmdC derivative **16**.¹⁴ The exocyclic amine together with the hydroxyl group was protected as a carbamate using *p*-nitrophenyl chloroformate.¹⁵ Efficient conversion required full deprotonation of both functional groups with NaH prior to addition of the protecting reagent. Final silyl deprotection, DMT protection, and synthesis of the hmdC phosphoramidite building block **18** with Bannwarth's reagent furnished the 2'-(R)-F-hmC phosphoramidite in high yield (34% over six steps from **13**).

Regarding 2'-(R)-F-fdC phosphoramidite building block **21**, we performed a carbonylative Stille coupling reaction of **13** with tributyltin hydride (see [Scheme 2](#)). Subsequent masking of

Scheme 2. Synthesis of 2'-(R)-F-fdC and 2'-(R)-F-cadC Phosphoramidite Building Blocks **21 and **24****



the formyl group as a 1,3-dioxane unit with 1,3-propanediol and TiCl₄ as the activating Lewis acid provided compound **19**. For the protection of the exocyclic amine, we chose *p*-MeOC₆H₄COCl as recently reported.¹⁶ The electron-pushing methoxy unit strongly enhances the stability of the amine protecting group during solid-phase DNA synthesis, and this is

A ODN4 (target):

5' - d(AGC TGT CCA TCG CTC **AXG** TAC AGA GCT G) - 3'
 X = fdC / 2'-(R)-F-fdC

ODN5 (complement):

3' - d(T TCG ACA GGT AGC GAG TGC ATG TCT CGA CTT) - 5'

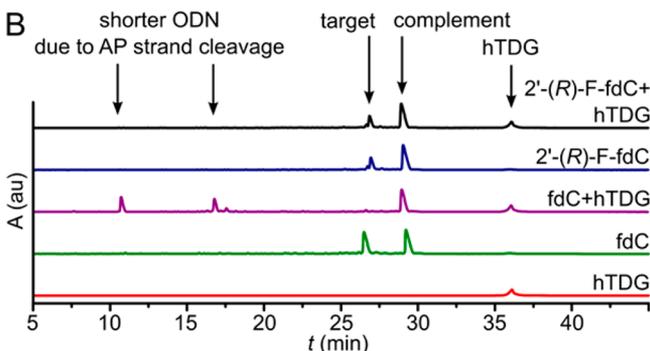


Figure 4. (A) Sequence of synthesized ODN4 and ODN5 for the hTDG glycosylation assay. (B) Reversed-phase HPL chromatogram of the hTDG glycosylation assay.

containing 2'-(R)-F-mdC, 2'-(R)-F-hmdC, 2'-(R)-F-fdC, and 2'-(R)-F-cadC in high yields and quality. Furthermore, we showed that a 2'-(R)-F label on fdC blocks the activity of the critical TDG enzyme, thus inhibiting base excision repair of this base. The 2'-(R)-F label is consequently the ideal tool for analysis of the epigenetic metabolism beyond base excision.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.orglett.6b02110](https://doi.org/10.1021/acs.orglett.6b02110).

Synthesis of all compounds and oligonucleotides and details of the assays (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Thomas.Carell@lmu.de. Website: <http://www.carellgroup.de>.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Nadine Hinkel, Korbinian Krieger, Stefan Marchner, and Robert Rampmaier (all Ludwig-Maximilians-Universität München) for practical assistance. Furthermore, we thank the Deutsche Forschungsgemeinschaft (SFB749), the Fonds der Chemischen Industrie (predoctoral fellowship for A.S.S.), and the Excellence Cluster (CiPS^M, EXC114 and GRK 2062) for support.

■ REFERENCES

- (1) (a) Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A. *J. Med. Chem.* **2015**, *58*, 8315–8359. (b) Purser, S.; Moore, P. R.; Swallow, S.; Gouverneur, V. *Chem. Soc. Rev.* **2008**, *37*, 320–330.
- (2) Parker, W. B. *Chem. Rev.* **2009**, *109*, 2880–2893.
- (3) (a) Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171–1202. (b) Williams, A. A.; Darwanto, A.; Theruvathu, J. A.; Burdzy, A.; Neidigh, J. W.; Sowers, L. C. *Biochemistry* **2009**, *48*, 11994–12004. (c) Berger, I.

Tereshko, V.; Ikeda, H.; Marquez, V. E.; Egli, M. *Nucleic Acids Res.* **1998**, *26*, 2473–2480.

(4) (a) Dai, Q.; Lu, X.; Zhang, L.; He, C. *Tetrahedron* **2012**, *68*, 5145–5151. (b) Schaerer, O. D.; Verdine, G. L. *J. Am. Chem. Soc.* **1995**, *117*, 10781–10782. (c) Schärer, O. D.; Kawate, T.; Gallinari, P.; Jiricny, J.; Verdine, G. L. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 4878–4883.

(5) (a) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929–930. (b) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930–935. (c) 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. (d) 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.

(6) (a) Loenarz, C.; Schofield, C. J. *Chem. Biol.* **2009**, *16*, 580–583. (b) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300–1303.

(7) (a) Maiti, A.; Drohat, A. C. *J. Biol. Chem.* **2011**, *286*, 35334–35338. (b) Raiber, E.-A.; Beraldi, D.; Ficiz, G.; Burgess, H. E.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, *13*, R69. (c) Zhang, L.; Lu, X.; Lu, J.; Liang, H.; Dai, Q.; Xu, G.-L.; Luo, C.; Jiang, H.; He, C. *Nat. Chem. Biol.* **2012**, *8*, 328–330.

(8) (a) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeyer, C.; Truss, M.; Carell, T. *Angew. Chem., Int. Ed.* **2012**, *51*, 6516–6520. (b) Liutkevičiūtė, Z.; Kriukienė, E.; Ličytė, J.; Rudytė, M.; Urbanavičiūtė, G.; Klimašauskas, S. *J. Am. Chem. Soc.* **2014**, *136*, 5884–5887. (c) 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) Lee, S.; Bowman, B. R.; Ueno, Y.; Wang, S.; Verdine, G. L. *J. Am. Chem. Soc.* **2008**, *130*, 11570–11571.

(10) (a) Richardson, F. C.; Tennant, B. C.; Meyer, D. J.; Richardson, K. A.; Mann, P. C.; McGinty, G. R.; Wolf, J. L.; Zack, P. M.; Bendele, R. A. *Toxicol. Pathol.* **1999**, *27*, 607–617. (b) Kierdaszuk, B.; Krawiec, K.; Kazimierczuk, Z.; Jacobsson, U.; Johansson, N. G.; Munch-petersen, B.; Eriksson, S.; Shugar, D. *Nucleosides Nucleotides* **1999**, *18*, 1883–1903.

(11) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671–5673.

(12) Hirota, K.; Kitade, Y.; Kanbe, Y.; Isobe, Y.; Maki, Y. *Synthesis* **1993**, *1993*, 213–215.

(13) Bannwarth, W.; Trzeciak, A. *Helv. Chim. Acta* **1987**, *70*, 175–186.

(14) Luche, J. L. *J. Am. Chem. Soc.* **1978**, *100*, 2226–2227.

(15) Sammet, B. *Synlett* **2009**, *2009*, 3050–3051.

(16) Schröder, A. S.; Steinbacher, J.; Steigenberger, B.; Gnerlich, F. A.; Schiesser, S.; Pfaffeneder, T.; Carell, T. *Angew. Chem., Int. Ed.* **2014**, *53*, 315–318.

(17) (a) Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chem. - Eur. J.* **2011**, *17*, 13782–13788. (b) Nomura, Y.; Hagino, N.; Ueno, Y.; Matsuda, A. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2811–2816.

(18) (a) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859–1862. (b) Sinha, N. D.; Biernat, J.; McManus, J.; Köster, H. *Nucleic Acids Res.* **1984**, *12*, 4539–4557.

(19) Hashimoto, H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X. *Nucleic Acids Res.* **2012**, *40*, 10203–10214.

(20) Bennett, M. T.; Rodgers, M. T.; Hebert, A. S.; Ruslander, L. E.; Eisele, L.; Drohat, A. C. *J. Am. Chem. Soc.* **2006**, *128*, 12510–12519.

3.3 2'-(R)-Fluorinated mC, hmC, fC and caC triphosphates are substrates for DNA polymerases and TET-enzymes

A. S. Schröder, E. Parsa, K. Iwan, F. R. Traube, M. Wallner, S. Serdjukow and T. Carell*

Prolog

In der vorausgegangenen Publikation wurde bereits gezeigt, dass die 2'-F Substitution des dC-Nukleosids keinen Einfluss auf die Aktivität der Methyltransferase M.Sss1 *in vitro* hat, während die höher oxidierte Spezies 2'-F-fdC von hTDG nicht entfernt werden kann und eine von BER Prozessen unbeeinflusste Studie zulässt. In dieser Arbeit wurde untersucht, ob die 2'-fluorierten Nukleoside auch Substrate für DNA Polymerasen und TET-Enzyme sind. Hierfür wurden zunächst die 2'-(R)-fluorierten mdC, hmdC, fdC und cadC Triphosphate synthetisiert. Der Einbau sollte enzymatisch erfolgen. Als mögliches Templat wurde ein Fragment der *octamer-binding transcription factor 4 (Oct4)*-Promotorsequenz gewählt. Diese Sequenz ist bekannt für eine hohe Dichte an epigenetisch relevanten Nukleosiden, da es viele CpG-Dinukleotide enthält, was sie aber oft zu einem für die Polymerasekettenreaktion (PCR) schwer zugänglichen Templat macht. Mit allen 2'-(R)-F-xdCTPs konnte eine Amplifikation der *Oct4*-Promotorsequenz mittels PCR erzielt werden. Um zu überprüfen, ob 2'-(R)-F-mdC auch ein Substrat der TET-Enzyme ist, wurde anschließend eine *in vitro* Studie durchgeführt. Hierfür wurde TET1cd in HEK293T Zellen überexprimiert, isoliert und mit dem 2'-(R)-F-mdC enthaltenen *Oct4*-Promotor Fragment für drei Stunden bei 37 °C inkubiert. Als Positivkontrolle wurde das gleiche Experiment mit der *Oct4*-Promotorsequenz und natürlichem mdC durchgeführt. Die auf den Verdau folgende Analyse *via* UHPLC-MS/MS zeigte, dass auch das fluorierte mdC ein Substrat der TET-Enzyme ist. Allerdings ist die Aktivität der TET-Enzyme durch den F-Substituenten etwas verringert. Diese fluorierten Nukleosid-Analoga sind somit gute Substrate für DNA-Polymerasen, TET-Enzyme und somit hervorragend für weitere Studien des epigenetischen Metabolismus geeignet.

Autorenbeitrag

Die bereits entwickelte UHPLC-MS/MS Methode wurde für diese Studie von mir weiterentwickelt, so dass es nun durch externe Quantifizierung möglich war, die Mengen der entstandenen Nukleoside zu bestimmen. Für die Messung und Quantifizierung der natürlichen Nukleoside wurde die von *Toni Pfaffeneder* entwickelte UHPLC-MS/MS Methode verwendet. Alle Proben wurden von mir verdaut und ausgewertet.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags.



Cite this: *Chem. Commun.*, 2016, 52, 14361

Received 14th September 2016,
Accepted 25th November 2016

DOI: 10.1039/c6cc07517g

www.rsc.org/chemcomm

2'-(*R*)-Fluorinated mC, hmC, fdC and caC triphosphates are substrates for DNA polymerases and TET-enzymes†

A. S. Schröder,‡ E. Parsa,‡ K. Iwan, F. R. Traube, M. Wallner, S. Serdjukow and T. Carell*

A deeper investigation of the chemistry that occurs on the newly discovered epigenetic DNA bases 5-hydroxymethyl-(hmC), 5-formyl-(fdC), and 5-carboxy-deoxycytidine (cadC) requires chemical tool compounds, which are able to dissect the different potential reaction pathways in cells. Here we report that the 2'-(*R*)-fluorinated derivatives F-hmC, F-fdC, and F-cadC, which are resistant to removal by base excision repair, are good substrates for DNA polymerases and TET enzymes. This result shows that the fluorinated compounds are ideal tool substances to investigate potential C–C bond cleaving reactions in the context of active demethylation.

While all cells of a multicellular organism have an identical DNA sequence, their morphology and function differ to a great extent (*i.e.* neurons *vs.* adenocytes). This is possible because these cells have different sets of genes in active and passive states.¹ This programming of the individual genomes occurs on the DNA level by the chemical modification of deoxycytidine (dC, **1**), which is potentially methylated to give 5-methyldeoxycytidine (mdC, **2**). In the years 2009^{2,3} and 2011^{4,5} it was shown that mdC is additionally oxidized by the action of TET-enzymes⁶ to give 5-hydroxymethyl-(hmC, **3**), 5-formyl-(fdC, **4**), and 5-carboxy-deoxycytidine (cadC, **5**). Although the exact function of these oxidized deoxycytidine derivatives is not clear, it is believed that they play a fundamental role during the epigenetic programming of the genome that leads to activated and passivated genes.⁷ Although it is unclear how this is mechanistically achieved we have learned that fdC and cadC are substrates of the base excision repair (BER) process.^{8–10} The glycosylase TDG is able to recognize fdC and cadC (but not hmC) upon which cleavage of the glycosidic bond occurs that transforms fdC and cadC into abasic sites (**6**), which are repaired *via* the insertion of “fresh” dC to give finally a demethylated position as shown in Fig. 1.

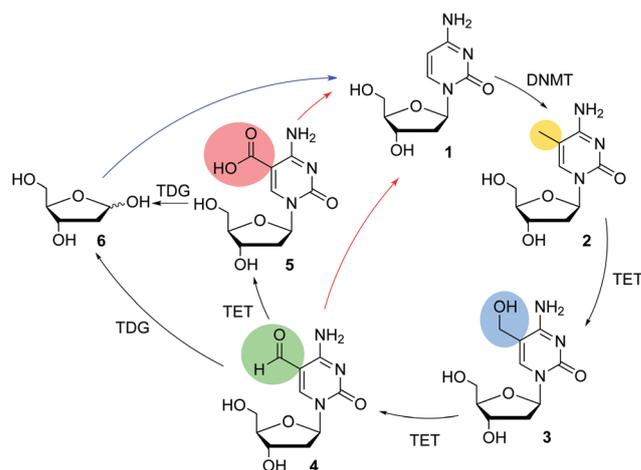


Fig. 1 Putative active demethylation pathways. DNMT: DNA methyltransferase, TET: ten-eleven-translocation methylcytosine dioxygenase, TDG: thymine DNA glycosylase. Red arrows: deformylation/decarboxylation, blue arrow: BER based active demethylation.

From a chemical point of view a direct deformylation of fdC or decarboxylation of cadC are attractive alternative pathways (C–C bond cleavage pathway, red arrows in Fig. 1) that would allow the direct conversion of fdC and cadC into dC, without the need to create potentially harmful abasic sites.^{11–13} These sites are especially dangerous when they are generated on both strands in the duplex because then harmful double strand breaks are generated. In order to study the potential C–C bond cleavage pathway we need tool substances that are TET substrates but resist repair by BER. Recently we reported that 2'-fluorinated versions of fdC and cadC are BER resistant.¹⁴ Here, we report that the triphosphates of F-mdC (**7**), F-hmC (**8**), F-fdC (**9**) and F-cadC (**10**), in which the 2'-center is (*R*)-configured, are good substrates for various DNA polymerases and that this property can be used to generate long DNA strands containing multiple 2'-(*R*)-fluorinated mdCs, hmCs, fdCs, and cadCs using only slightly adapted PCR protocols. We finally show that the 2'-fluorinated compounds are also good substrates for the

Center for Integrated Protein Science, Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstraße 5-13, 81377 Munich, Germany. E-mail: Thomas.carell@lmu.de; Web: www.carellgroup.de

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cc07517g

‡ These authors contributed equally.

TET enzymes and this together makes them ideal tool substances to study active demethylation *via* the putative C–C bond cleavage.

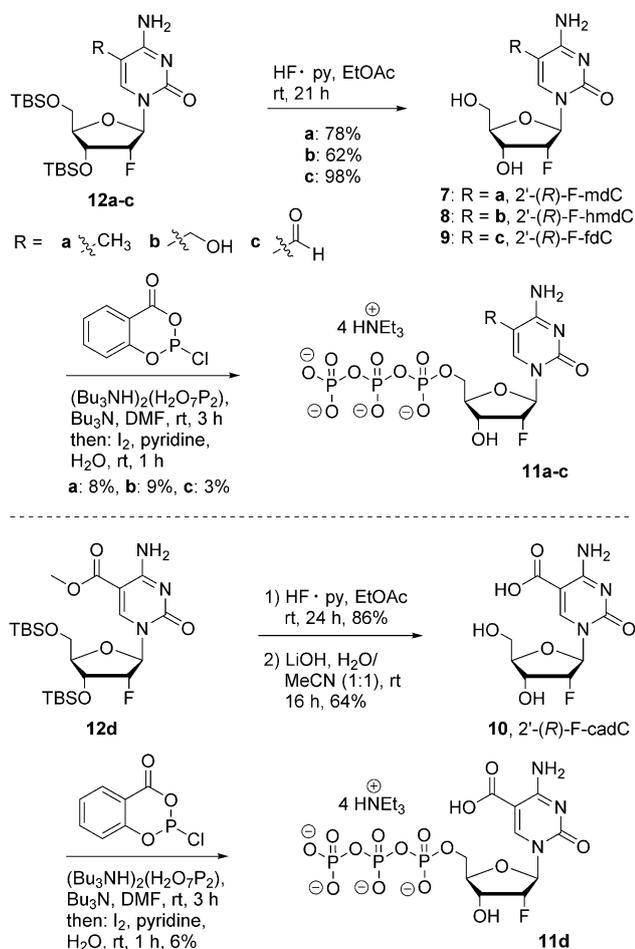
For the synthesis of the 2'-(*R*)-F-xdC triphosphates **11a–d** (see Scheme 1), we started with the corresponding, silyl protected 2'-(*R*)-F-xdC nucleosides **12a–d**.¹⁴ Silyl deprotection of **11a–c** with Olah's reagent furnished the 2'-(*R*)-fluorinated nucleosides of mdC (**7**), hmdC (**8**) and fdC (**9**) in good yields between 62% and 98%. Regarding the nucleoside 2'-(*R*)-F-cadC (**10**), we first deprotected the silyl groups prior to saponification with LiOH in H₂O/MeCN (1 : 1). Using Eckstein conditions, it was possible to convert the nucleosides **7–10** into the corresponding 2'-(*R*)-fluorinated triphosphate mdCTP (**11a**), hmdCTP (**11b**), fdCTP (**11c**) and cadCTP (**11d**) in one-pot-reactions.^{15–17} After extensive purification by preparative HPL chromatography, we obtained yields in the range of 3–9%. These yields are remarkable, particularly in light of the fact that no protecting groups were used, *e.g.* on the benzylic hydroxyl group of 2'-(*R*)-F-hmdC (**8**).

We next examined if the triphosphates are able to function as substrates for DNA polymerases. For this purpose, we started with primer extension studies to screen for suitable polymerases (see ESI†).^{18,19} Remarkably, all triphosphates with the exception

of 2'-(*R*)-F-cadCTP (**11d**) were accepted by the DNA-polymerase *Phusion*. This is surprising, because *Phusion* exhibits robust proof-reading activity and as such may hinder incorporation of unnatural triphosphates. The here examined triphosphates have not only a chemical group on the base but feature in addition a fluorine atom at C2'. The result suggests that the 2'-(*R*)-F derivatives mimic the natural situation quite closely.¹⁷ The fluorine atom is bioisosteric to a H-atom²⁰ and it seems that its (*R*)-configuration, which most likely leads to a C3'-*endo* conformation of the sugar pucker does not hinder DNA polymerase based incorporation.^{21–24} The incorporation of 2'-(*R*)-F-hmdCTP (**11b**) and 2'-(*R*)-F-fdCTP (**11c**) was furthermore possible with the polymerase *KOD-XL*, and here the yield were slightly higher. Incorporation of 2'-(*R*)-F-cadCTP (**11d**) was finally achieved with the polymerase *Therminator*. We believe that the problems associated with this base are caused by the carboxylic acid at the 5-position of the base, which carries a negative charge.¹⁷

With the 2'-(*R*)-F-xdCTP (**11a–d**) and the knowledge of which polymerases to use at hand, we next searched for appropriate conditions for the PCR (see Fig. 2). As a template, we selected a fragment of the OCT4 promoter sequence (see Fig. 2A). This sequence is known for its high density of epigenetically relevant nucleosides and, due to the high amount of CpG units it is usually considered to be a difficult template for PCR. The primers for the PCR were designed to yield an 81 base pair product containing 14–15 2'-(*R*)-F-xdC bases depending on the primer. 4–5 dCs are present in the primer and these are of course not exchanged during PCR. For the reaction, we fully replaced the dC triphosphate by the corresponding 2'-(*R*)-F-xdCTP. Hence, full length PCR products can only be formed if the appropriate 2'-(*R*)-F-xdCTP is accepted and incorporated by the polymerase. Forward and reverse primer were annealed to the template at 55 °C. The elongation of the primers was best performed at 72 °C for only 25 seconds. For complete extension of the primer, we extended the final elongation time to 5 min. The experimental results of the PCR are shown in Fig. 2B. For gel electrophoresis analysis, we had to use a tris-borate buffer system instead of tris-acetate to avoid “smearing” of the 2'-(*R*)-F-cadC product caused by the additional carboxyl groups present at this base.

Remarkably, we obtained full length PCR products for all 2'-(*R*)-F-xdCTPs showing again that a 2'-(*R*)-F-substitution hardly affects the procession of the polymerase. Further proof of the correct incorporation of 2'-(*R*)-F-xdCTP was gained by LC-MS/MS analysis. To this end, the PCR products were fully digested to the nucleoside level. The sugar phosphate backbone was first cleaved with nuclease S1 and *Antarctic phosphatase*, resulting in the oligo- and 5'-monophosphates as well as nucleosides. Further hydrolysis down to the nucleoside level was realized with snake venom phosphodiesterase I.^{25–28} The resulting nucleoside mixture was subsequently analyzed by UHPLC-MS/MS (QQQ) (see Fig. 2C and D). As an example, the UV-trace of the digested 2'-(*R*)-F-fdC PCR product is shown. The clean chromatogram shows only the expected nucleosides dA, dT, dG, dC (from the primer) and the 2'-(*R*)-F-fdC compound. The quantification of the 2'-(*R*)-F-nucleosides using exact calibration curves of the synthesized nucleosides confirms the incorporation. Most importantly,



Scheme 1 Synthesis of 2'-(*R*)-F-substituted triphosphates of mdC (**11a**); 2'-(*R*)-F-mdCTP), hmdC (**11b**); 2'-(*R*)-F-hmdCTP), fdC (**11c**); 2'-(*R*)-F-fdCTP) and cadC (**11d**); 2'-(*R*)-F-cadCTP).

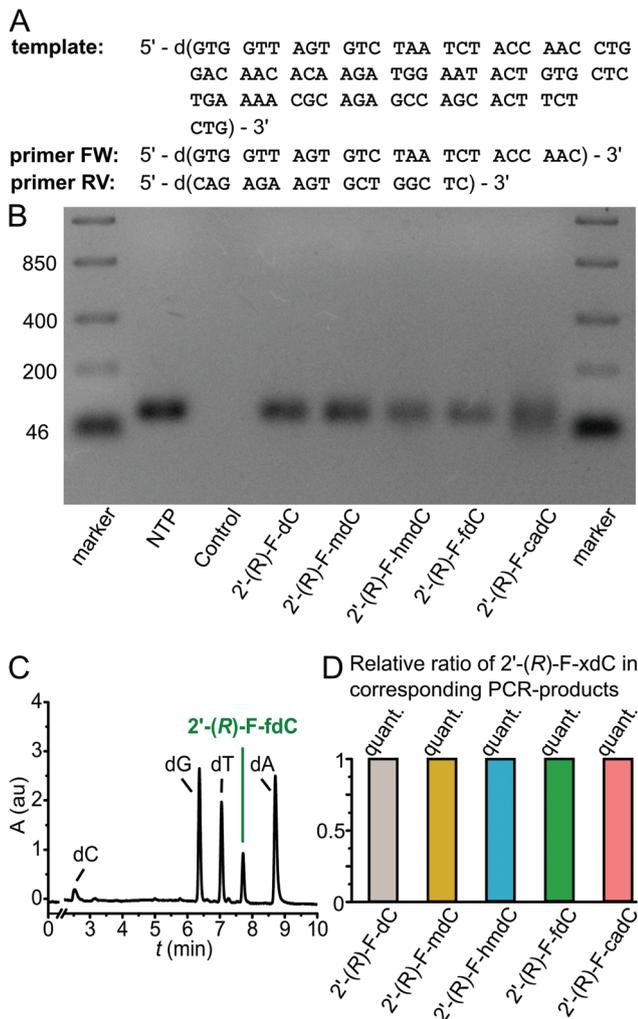


Fig. 2 Synthesis of 2'-(R)-F-xdC containing OCT4 promoter fragments by PCR. (A) Sequence of the template and primers used for the PCR (FW: forward, RV: reverse). (B) Result of the PCR analyzed by gel electrophoresis. The retention of the band corresponds to the expected PCR products. (Control: PCR with all NTP, but without a polymerase. 2'-(R)-F-xdC: PCR with all NTP except dCTP but with corresponding 2'-(R)-F-xdC.) (C) UV trace of UHPLC-chromatogram after enzymatic digestion of the 2'-(R)-F-dC PCR-product. (D) Quantification data of 2'-(R)-F-xdC PCR-products after enzymatic digestion and UHPLC-MS/MS analysis.

we did not detect any side products during the PCR reaction. Overall, the data show, that the 2'-(R)-F-xdCTPs can be incorporated despite the 2'-(R)-fluoro label *via* PCR into long oligonucleotides.

We next turned our attention to the question, if 2'-(R)-F-xdCs can be oxidized by TET enzymes despite the presence of the 2'-(R)-fluoro substituent. To this end, we overexpressed fused constructs of GFP and the catalytic domain of TET1 (TET1cd) in HEK293T cells and isolated functional TET1cd with the help of agarose beads coated with anti-GFP-antibodies. To obtain sufficient amounts of the enzyme we treated the cells with sodium butyrate. This compound acts as an HDAC inhibitor which leads to increased protein expression.²⁹ High concentrations of benzonase and rigorous washing yielded pure and functional TET1cd. This was confirmed by incubation with

an ODN containing mdC in which mdC was efficiently oxidized (see ESI[†]).

We then added the 81 basepair long OCT4 promoter fragment containing 14–15 2'-(R)-F-mdCs, depending on forward or reverse strand, respectively. After incubating the PCR-product with TET1cd in a reducing buffer (see ESI[†]) for 3 h at 37 °C, we isolated the fragment, digested it down to the nucleoside level as described above and analyzed the obtained nucleoside mixture using UHPLC-MS/MS (QQQ). As shown in Fig. 3B the MS-trace showed clearly that the 2'-(R)-F-mdC starting material was not only oxidized to 2'-(R)-F-hmdC but also to the higher oxidized species 2'-(R)-F-fdC and 2'-(R)-F-cadC. Exact quantification data show that TET1cd oxidation produced 7.9% 2'-(R)-F-hmdC, 3.3% 2'-(R)-F-fdC and 0.2% 2'-(R)-F-cadC. Again, no side products like deamination to 2'-(R)-F-dT or background C–C bond cleavage, which would provide 2'-(R)-F-dC, were detected. In order to study if the TET protein oxidizes the 2'-(R)-F-mdC with an efficiency comparable to the non-fluorinated mdC, we added the same 81 basepair long OCT4 promoter fragment containing mdC instead of 2'-(R)-F-mdC to pure and functional TET1cd. After digestion and LC-MS/MS analysis, we detected now only fdC and cadC but no hmdC showing that the non-fluorinated mdC is the slightly better substrate (see ESI,† Table S2). However, we see for mdC and 2'-F-mdC oxidation up to the fdC and cadC level. This shows that the 2'-(R)-fluoro analogs can report TET activity, although the F-atom does reduce the TET activity to some extent.

In summary, here we show that the 2'-(R)-F triphosphates of the epigenetically relevant nucleosides mdC, hmdC, fdC and cadC can be efficiently incorporated into long oligonucleotides

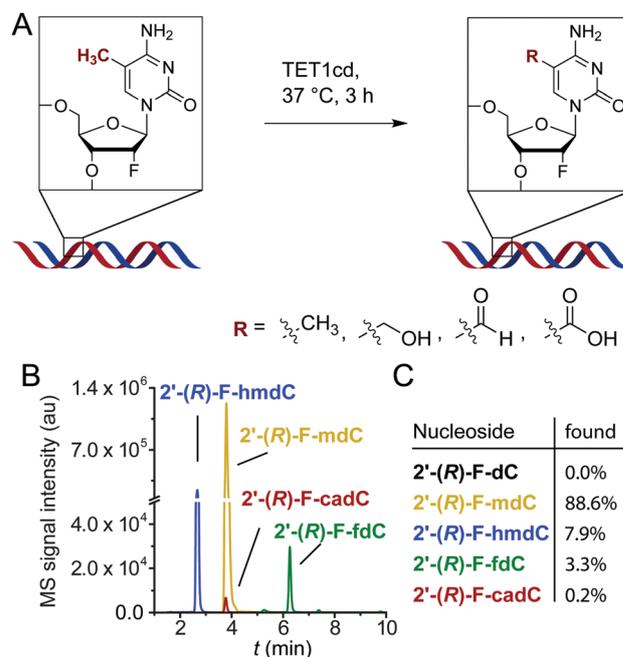


Fig. 3 TET1cd oxidation assay. (A) Scheme of the TET1cd dependent oxidation reaction. (B) UHPLC-MS/MS-trace of 2'-(R)-F-xdC nucleosides after enzymatic digestion of the DNA fragment. (C) Quantification data of the product after complete enzymatic digestion and UHPLC-MS/MS (QQQ) analysis.

using PCR. This provides oligonucleotides of sufficient length for detailed mechanistic studies. Importantly, while the 2'-(R)-F substituent blocks the BER-based removal of fdC and cadC it is a good substrate for the TET enzymes.

We thank Jessica Furtmeier for practical assistance (Ludwig-Maximilians-Universität München). Furthermore, we thank the Deutsche Forschungsgemeinschaft (SFB749, SFB1032, CA275-8/5) for financial support. Additional support was obtained from the Fonds der Chemischen Industrie (predoctoral fellowship for A. S. S.), the Boehringer Ingelheim Fonds (predoctoral fellowship for F. R. T.), the Excellence Cluster (CiPS^M, EXC114), the DFG GRK2062 and the Marie Curie Training and Mobility network (Clickgene) is acknowledged.

Notes and references

- 1 A. Avgustinova and S. A. Benitah, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 643–658.
- 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 and A. Rao, *Science*, 2009, **324**, 930–935.
- 3 S. Kriaucionis and N. Heintz, *Science*, 2009, **324**, 929–930.
- 4 T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeyer and T. Carell, *Angew. Chem., Int. Ed.*, 2011, **50**, 7008–7012.
- 5 S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, *Science*, 2011, **333**, 1300–1303.
- 6 C. Loenarz and C. J. Schofield, *Chem. Biol.*, 2009, **16**, 580–583.
- 7 A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell and S. Michalakis, *Cell Rep.*, 2015, **11**, 283–294.
- 8 A. Maiti and A. C. Drohat, *J. Biol. Chem.*, 2011, **286**, 35334–35338.
- 9 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 and G.-L. Xu, *Science*, 2011, **333**, 1303–1307.
- 10 E.-A. Raiber, D. Beraldi, G. Ficiz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik and S. Balasubramanian, *Genome Biol.*, 2012, **13**, 1–11.
- 11 S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld and T. Carell, *J. Am. Chem. Soc.*, 2013, **135**, 14593–14599.
- 12 Z. Liutkevičiūtė, E. Kriukienė, J. Ličytė, M. Rudytė, G. Urbanavičiūtė and S. Klimašauskas, *J. Am. Chem. Soc.*, 2014, **136**, 5884–5887.
- 13 S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeyer, M. Truss and T. Carell, *Angew. Chem., Int. Ed.*, 2012, **51**, 6516–6520.
- 14 A. S. Schröder, O. Kotljarova, E. Parsa, K. Iwan, N. Raddaoui and T. Carell, *Org. Lett.*, 2016, **18**, 4368–4371.
- 15 J. Ludwig and F. Eckstein, *J. Org. Chem.*, 1989, **54**, 631–635.
- 16 J. Caton-Williams, M. Smith, N. Carrasco and Z. Huang, *Org. Lett.*, 2011, **13**, 4156–4159.
- 17 B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder and T. Carell, *Org. Lett.*, 2013, **15**, 366–369.
- 18 T. Ono, M. Scalf and L. M. Smith, *Nucleic Acids Res.*, 1997, **25**, 4581–4588.
- 19 C. G. Peng and M. J. Damha, *Can. J. Chem.*, 2008, **86**, 881–891.
- 20 R. S. Rowland and R. Taylor, *J. Phys. Chem.*, 1996, **100**, 7384–7391.
- 21 M. L. Sinnott, *Chem. Rev.*, 1990, **90**, 1171–1202.
- 22 A. A. Williams, A. Darwanto, J. A. Theruvathu, A. Burdzy, J. W. Neidigh and L. C. Sowers, *Biochemistry*, 2009, **48**, 11994–12004.
- 23 I. Berger, V. Tereshko, H. Ikeda, V. E. Marquez and M. Egli, *Nucleic Acids Res.*, 1998, **26**, 2473–2480.
- 24 S. Lee, B. R. Bowman, Y. Ueno, S. Wang and G. L. Verdine, *J. Am. Chem. Soc.*, 2008, **130**, 11570–11571.
- 25 T. Ando, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1966, **114**, 158–168.
- 26 M. Rina, C. Pozidis, K. Mavromatis, M. Tzanodaskalaki, M. Kokkinidis and V. Bouriotis, *Eur. J. Biochem.*, 2000, **267**, 1230–1238.
- 27 T. Nihei, G. L. Cantoni and With the technical assistance of Rachelle Rothenberg, *J. Biol. Chem.*, 1963, **238**, 3991–3998.
- 28 A. S. Schröder, J. Steinbacher, B. Steigenberger, F. A. Gnerlich, S. Schiesser, T. Pfaffeneder and T. Carell, *Angew. Chem., Int. Ed.*, 2014, **53**, 315–318.
- 29 M. G. Riggs, R. G. Whittaker, J. R. Neumann and V. M. Ingram, *Nature*, 1977, **268**, 462–464.

3.4 5-Formylcytosine to cytosine conversion by C-C bond cleavage *in vivo*

Katharina Iwan,² René Rahimoff,² Angie Kirchner,² Fabio Spada,² Arne S. Schröder, Olesea Kosmatchev, Shqiponja Ferizaj, Jessica Steinbacher, Edris Parsa, Markus Müller & Thomas Carell*

(² geteilte Erstautorenschaft)

Prolog

Wie die aktive Demethylierung in der DNA verläuft, ist eine viel diskutierte Frage. Bisher sind zwei Wege vorstellbar: Das Ausschneiden der höher oxidierten Nukleoside (fdC und cadC) durch die an der Basenexzisionsreparatur beteiligten Glykosylasen oder der direkte C-C Bindungsbruch in Form einer Deformylierung von fdC oder einer Decarboxylierung des cadCs. Beide möglichen Wege würden zu einem unmodifizierten dC führen. Im Gegensatz zur BER hat ein direkter C-C Bindungsbruch einen großen Vorteil, denn er führt nicht zu einem potentiell gefährlichen Strangbruch. In dieser Studie wurde gezeigt, dass es möglich ist, isotopenmarkiertes fdC wie auch 2'-(*R*)-F-fdC in das Genom von Säugetierzellen zu inkorporieren. Der Metabolismus dieser fdC-Derivate in der DNA kann durch UHPLC-MS/MS Messungen nachverfolgt werden. In diesen Experimenten konnte quantitativ gezeigt werden, dass nach der Inkorporation in das Genom, eine Deformylierungsreaktion von fdC zu dC stattfindet. Diese C-C Bindungsbruchreaktion führt zu einer direkten Umwandlung von fdC zu einem unmodifizierten dC ohne Entstehung einer abasischen Stelle bzw. eines Strangbruches.

Autorenbeitrag

Die bereits entwickelte UHPLC-MS/MS Methode wurde für diese Experimente von mir wie folgt weiterentwickelt: Zum einen wurden die natürlichen Cytidin-Modifikationen in die Messmethode integriert und zum anderen wurde eine interne Quantifizierungsmethode für F-fdC, F-dC und F-mdC entwickelt. Die DNA-Proben wurden von mir verdaut, gemessen und ausgewertet. Die Interpretation der erhaltenen Daten wurde von den Erstautoren, zusammen mit *Thomas Carell*, gemeinsam vorgenommen. Eine detaillierte Auflistung der Autorenbeiträge ist im nachfolgenden Manuskript zu finden.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags.

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 cytidine 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

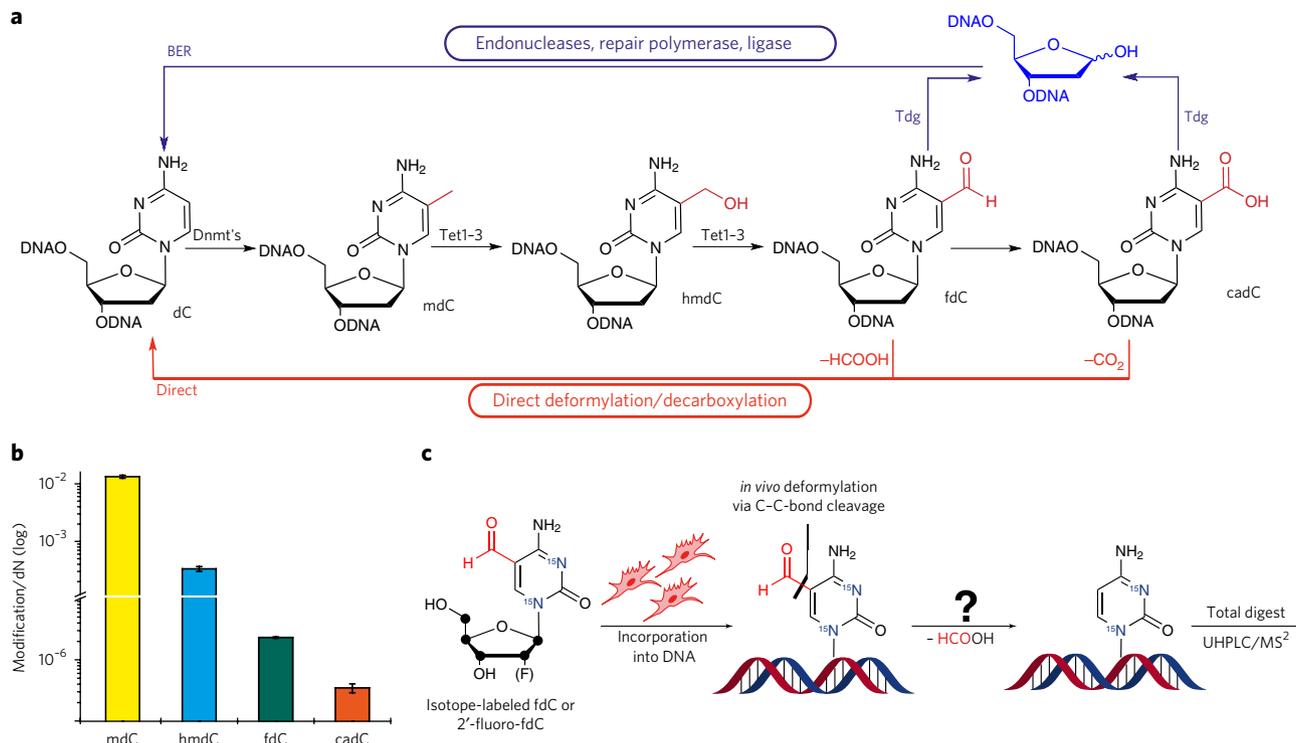


Figure 1 | Isotope tracing studies. (a) Suggested pathways of active demethylation. Thymine-DNA glycosylase (Tdg)-based cleavage of the glycosidic bond of 5-formyl-dC (fdC) and 5-carboxy-dC (cadC) results in an abasic site, which initiates a BER process that leads to the replacement of fdC and cadC by 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]fdC (**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

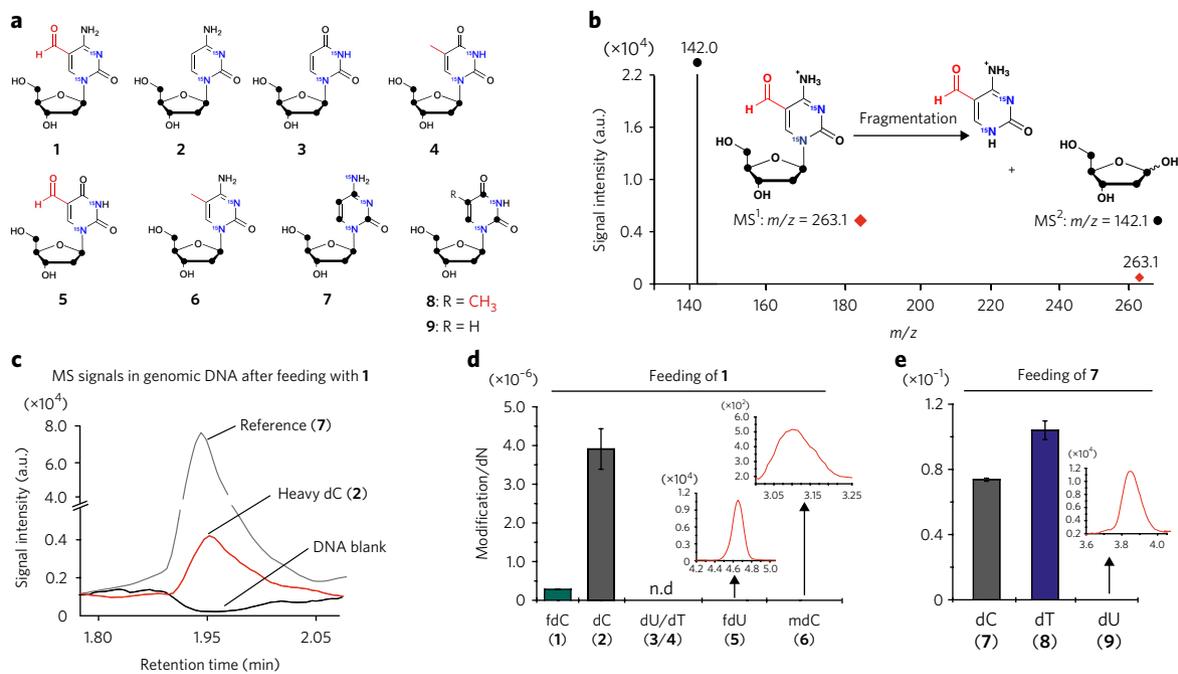


Figure 2 | Conversion of isotopically labeled fdC into dC in mESCs. (a) Overview of the compounds that may be detected after feeding of **1** or **7** to mESCs. (b) Feeding of **1** to mESCs results in incorporation of the isotopologue into the genomic DNA, as proven by its fingerprint MS transition. 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 μM, 1.0 μM or 2.5 μM 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 μM **10** for 3 d, we detected around 3% (±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 μM 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

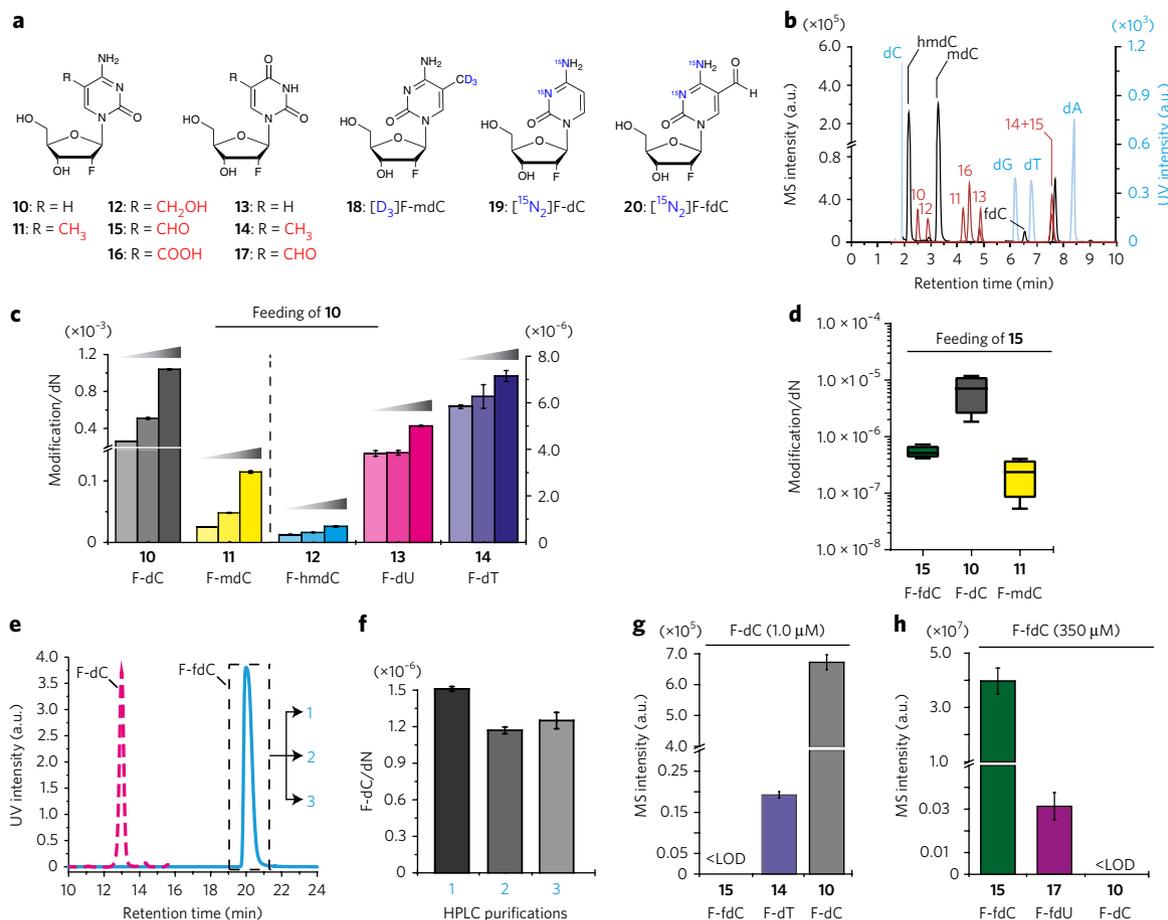


Figure 3 | F-fdC is converted into F-dC within the genome. (a) Chemical structures of 2'-fluorinated dC and dU derivatives that were investigated and internal standards used. (b) Resulting UV and MS traces of the nucleosides under investigation. Light blue, canonical bases; black, natural dC derivatives; 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 deformylation 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 deformylation.

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% ($\pm 0.3\%$) 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

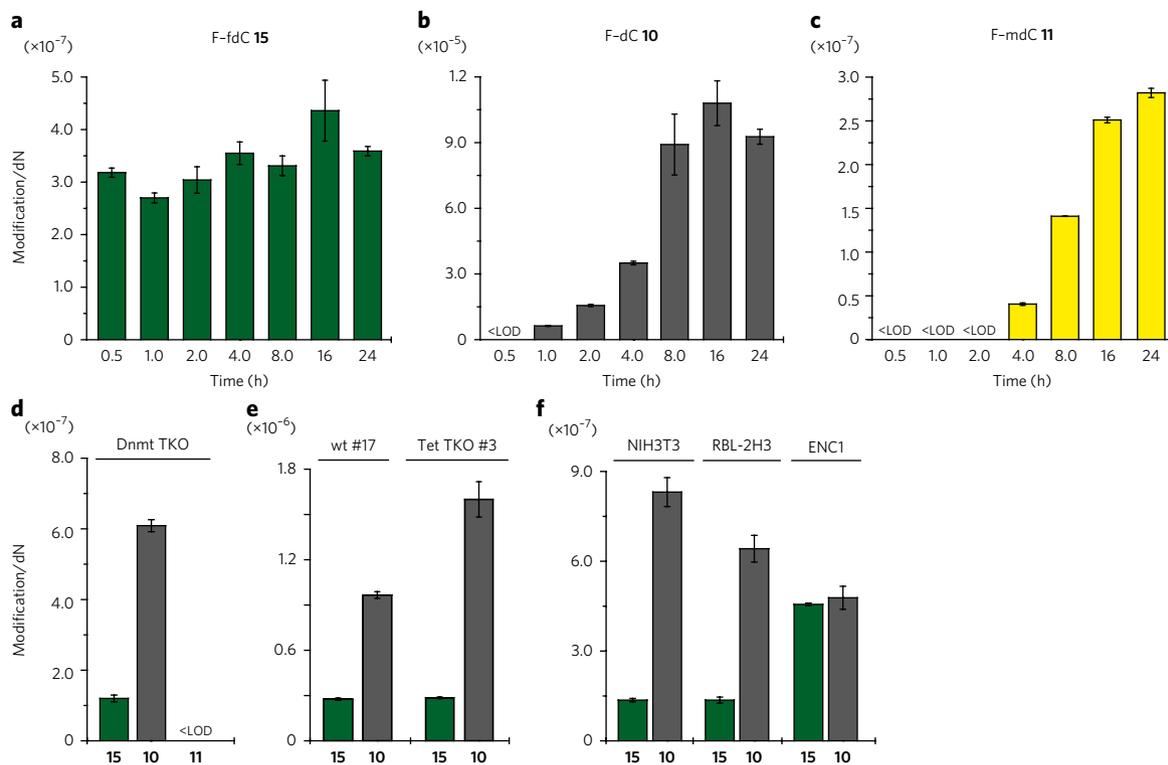


Figure 4 | Demodification of 2'-fluorinated fdC is a rapid process, does not require Dnmt or Tet enzymes and occurs also in somatic cell types.

(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 defluorinated 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 bound peroxy 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**, e13567 (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).
- Branco, M.R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2011).
- Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
- Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **11**, 555–557 (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).
- Liutkeviciūtė, 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.

ONLINE METHODS

Chemical synthesis. Synthetic schemes, detailed procedures and characterization of synthesized products can be found in the **Supplementary Note**. Unless noted otherwise, all reactions were performed using flame- or oven-dried glassware under an atmosphere of nitrogen. Compounds **7** (B.A.C.H. UG) and **10** (Carbosynth) were commercially available. **15** was synthesized as previously described in the literature³². Identities of these compounds were confirmed by NMR and LC-MS/MS. Molsieve-dried solvents were used from Sigma-Aldrich, and chemicals were bought from Sigma-Aldrich, TCI, Carbolution and Carbosynth. Technical grade solvents were distilled before extraction or chromatography of compounds. Reaction controls were performed using TLC Plates from Merck (Merck 60 F254), flash-column chromatography purifications were performed on a Merck Geduran Si 60 (40–63 μM). Visualization of the developed 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. Infrared (IR) spectroscopic measurements were performed on a PerkinElmer 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.

Cell culture. Basal medium for mESC culture was DMEM high glucose containing 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, 100 $\mu\text{g}/\text{mL}$ streptomycin, 1 \times MEM Non-essential Amino Acid Solution and 0.1 mM β -mercaptoethanol (all from Sigma). All mESC lines were maintained in their naïve state on gelatin-coated plates by supplementing basal medium with 1,000 U/mL LIF (ORF Genetics), GSK3 inhibitor CHIR99021 at 3 μM and Mek inhibitor PD0325901 1 μM ("2i"). Metabolic labeling experiments with fluorine- or isotope-labeled nucleosides were performed by plating mESCs in priming conditions consisting of basal mESC medium supplemented with 3 μM CHIR99021 and Wnt pathway inhibitor IWR1-endo at 2.5 μM as previously reported⁴⁴. Under these conditions primed cells remained pluripotent for at least seven days as determined by epifluorescence with an Oct4-YFP knock-in cell line⁴⁵. Priming and labeling was performed for 3 d. Over this period naturally occurring genomic mdC and hmdC (**Fig. 1b**) reached levels similar to those recently reported for epiblast-like cells, which are regarded as the closest *in vitro* counterpart to noncommitted post-implantation epiblast^{46,47}. All inhibitors were purchased from Selleckchem. Dnmt TKO J1 mESCs were described in ref. 48. Two independent sets of Tet TKO and respective wt mESC lines were used: wt #17 and Tet TKO #3 were reported in ref. 49 and wt #4 and Tet TKO #29 were described in ref. 50. J1 mESCs are from the 129/SvJae strain, while all Tet TKO and corresponding wt mESC lines are from mixed genetic backgrounds.

The time-course experiment was performed by culturing J1 mESCs under priming conditions for 48 h. The medium was exchanged to priming medium containing 350 μM F-fdC and cells were harvested after 0.5, 1.0, 2.0, 4.0, 8.0, 16 and 24 h, as described.

RBL-2H3, HeLa, NIH3T3 and Neuro-2a cells were cultured in DMEM high glucose containing 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin, 100 $\mu\text{g}/\text{mL}$ streptomycin. CHOK1 cells were maintained in DMEM/F12 supplemented as reported above for the other somatic cell lines. ENC1 neural stem cells were cultured as previously described⁵¹. Cells were exposed to labeled nucleosides for 4 (RBL-2H3 and NIH3T3), 5 (CHO-K1), 6 (Neuro-2a) and 7 d (HeLa and ENC1).

Labeled nucleosides were added to the culture medium at the following concentrations: F-fdC (**15**), 350 μM ; F-dC (**10**), 0.5, 1.0 and 2.5 μM ; [¹³C]₂[¹⁵N]₂ fdC (**1**), 50 μM ; [¹³C]₂[¹⁵N]₂dC (**7**), purchased from B.A.C.H. UG), 100 μM .

Isolation of genomic DNA. Cultures were washed with PBS and lysed by adding RLT buffer (Qiagen) containing 400 μM each of 2,6-di-tert-butyl-4-methylphenol (BHT) and desferoxamine mesylate (DM) directly to the plates. Isolation of genomic DNA was performed with Zymo-Spin V, V-E or IIC-XL columns according to the instruction of the ZR-Duet DNA/RNA MiniPrep Kit (Zymo Research) with the following modifications. DNA was sheared by bead

milling in 2 mL microfuge tubes using one 5-mm diameter stainless steel bead per tube and a MM400 bead mill (Retsch) set at 30 Hz for 1 min. Lysates were then loaded onto spin columns and the bound material was first incubated for 15 min with Genomic Lysis Buffer (Zymo Research) supplemented with 0.2 mg/mL RNase A (Qiagen). After washing genomic DNA fragments were eluted with water containing 0.4 μM of each BHT and DM.

DNA digestion. 0.5–10 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 nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 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 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for 3 h at 37 °C or overnight and then stored at –20 °C. Prior to UHPLC-MS/MS analysis, samples were filtered by using an AcroPrep Advance 96 filter plate 0.2 μm Supor (Pall Life Sciences).

LC/MS-MS analysis of DNA samples. Quantitative UHPLC-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. Prior to every measurement series, external calibration curves were measured to quantify the levels of the F-nucleosides (**Supplementary Fig. 12**). Additionally, (R)-2'-[¹⁵N₂]F-dC (**19**), (R)-2'-[¹⁵N₂]F-hmdC (**37**) and (R)-2'-[¹⁵N₂]F-fdC (**20**) were used to validate the resulting peaks by co-injection. For exact quantification of fluorinated nucleosides also internal quantification with stable isotope dilution techniques for F-fdC, F-dC and F-mdC were developed (**Supplementary Fig. 13**). Natural nucleosides were quantified with the stable isotope dilution technique. An improved method, based on earlier published work^{26,28,52–54} was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run¹¹. 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/250 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 1–4**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 μm , 2.1 mm \times 150 mm) at 35 °C using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 mL/min: 0–4 min; 0–3.5% (v/v) MeCN; 4–7.9 min; 3.5–5% MeCN; 7.9–8.2 min; 5–80% MeCN; 8.2–11.5 min; 80% MeCN; 11.5–12 min; 80–0% MeCN; 12–14 min; 0% MeCN. 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 39 μL .

Quantification of nucleosides. Prior to every sample set, calibration curves to quantify all fluorine labeled nucleosides were measured under the same conditions and settings. All calibration curves are valid within the range of 1–500 fmol with five measuring points and measured as technical triplicates. **Supplementary Fig. 12** shows representative calibration curves for all Fluoro-nucleosides used for the quantification.

To obtain the internal calibration curves for exact quantification, we analyzed each standard, namely (R)-2'-[¹⁵N₂]F-fdC (**20**; $n = 205$ fmol), (R)-2'-[¹⁵N₂]F-dC (**19**; $n = 793$ fmol) and (R)-2'-[D₃]F-mdC (**18**; $n = 461$ fmol), in comparison to the corresponding nonlabeled nucleoside with constant concentration. Technical triplicates were measured, and the linear regression was applied using Origin 6.0 (Microcal). Therefore, the ratio of the area under the curve of unlabeled nucleoside (A) to the labeled standard (A*) was plotted against the ratio of the amount of unlabeled nucleoside (n) to the labeled one (n*) (see **Supplementary Fig. 13**). Acceptable precision (<20% relative s.d. within each triplicate) and accuracy (80%–120%) was achieved for all three calibration curves. The accuracy is calculated as the ratio of the measured to the calculated ratios of the areas (A/A*) under the curves in percent. The ratios of the areas (A/A*) can be calculated by using the linear equations for the corresponding ratio of amount (n/n*).

The lower limit of detection was defined as the detected amount, which is three times higher than the blank response (LOD). The lower limit of detection (LLOQ) and the upper limit of detection (ULOQ) are the lowest and the highest amounts (n) and the ratio of the amounts (A/A*) fulfilling the requirements of the corresponding linear equation, respectively.

Nucleoside stability test. Compounds **1** and **15** were incubated at 100 μM in mESC culture medium at 37 °C and 5% CO_2 for 3 d. For the recovery of the nucleosides Supel-Select SPE HLB cartridges from Sigma-Aldrich were used. Prior to use, the cartridges were equilibrated with MeOH, followed by acidified H_2O (with HCl to pH = 4). The pH of the samples was adjusted to 4, and the acidic solution was loaded on the cartridges. After washing with 10 mL of H_2O , the cartridges were dried *in vacuo*. The nucleosides were eluted with MeOH/MeCN (1:1), evaporated to dryness via speedvac and resuspended in H_2O .

Oligonucleotide stability test. An oligonucleotide (6.9 pmol) containing one F-fdC (28-mer) was incubated in mESC culture medium at 37 °C and 5% CO_2 for 3 d. For the recovery of the oligonucleotide, Oligo Clean & Concentrator from Zymo Research was used according to the manual. The resulting oligonucleotide was dissolved in H_2O and digested as described for genomic DNA.

Extraction of nucleoside/nucleotide soluble pools. J1 mESCs were plated under priming conditions (as described above) for 3 d. The culture medium was supplemented with 1.0 μM F-dC (**10**), 50 μM [$^{13}\text{C}_5$][$^{15}\text{N}_2$]fdC (**1**) or 350 μM F-fdC (**15**). Cells were washed twice with PBS (Sigma-Aldrich), harvested by trypsinization and pelleted by centrifugation for 3 min at 300g. 500 μL ice-cold 50% (v/v) acetonitrile was added dropwise to the pellet and vortexed⁵⁵. The mixture was incubated on ice for 10 min. The insoluble fraction was then separated from the soluble pool by centrifugation for 10 min at 21,000 $\times g$ at 0 °C. The supernatant was removed and used for nucleoside isolation. The soluble fraction containing the nucleosides was dried by lyophilization and metabolites were purified using Supel-Select SPE HLB cartridges (as described in the nucleoside stability test) before UHPLC-MS/MS analysis.

Life Sciences Reporting Summary. Further information on experimental design and reagents is available in the **Life Sciences Reporting Summary**.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

44. Kim, H. *et al.* Modulation of β -catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat. Commun.* **4**, 2403 (2013).
45. Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development* **135**, 909–918 (2008).
46. Shirane, K. *et al.* Global landscape and regulatory principles of DNA methylation reprogramming for germ cell specification by mouse pluripotent stem cells. *Dev. Cell* **39**, 87–103 (2016).
47. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
48. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
49. Hu, X. *et al.* Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14**, 512–522 (2014).
50. Dawlaty, M.M. *et al.* Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* **29**, 102–111 (2014).
51. Liu, N. *et al.* Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS One* **8**, e62755 (2013).
52. Cao, H. & Wang, Y. Collisionally activated dissociation of protonated 2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives. *J. Am. Soc. Mass Spectrom.* **17**, 1335–1341 (2006).
53. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
54. Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).
55. Dietmair, S., Timmins, N.E., Gray, P.P., Nielsen, L.K. & Krömer, J.O. Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. *Anal. Biochem.* **404**, 155–164 (2010).

3.5 ALKBH5-Induced Demethylation of Mono- and Dimethylated Adenosine

Timm T. Ensfelder,[‡] Matthias Q. Kurz,[‡] Katharina Iwan,[‡] Simon Geiger, Sarah Matheisl, Markus Müller, Roland Beckmann, Thomas Carell*

([‡] geteilte Erstautorenschaft)

Prolog

RNA enthält eine große Anzahl von modifizierten Nucleosiden. Am häufigsten sind in allen RNA-Spezies mono-methylierte oder vielfach methylierte Nucleoside vorhanden. Ein besonders hoher Gehalt an methylierten Nucleosiden wird in *transfer* RNA (tRNA) nachgewiesen, aber auch *ribosomale* RNA (rRNA) und *messenger* RNA (mRNA) weisen solche nicht-kanonischen Nucleoside auf. Grundsätzlich existieren von allen kanonischen Nucleosiden methylierte Derivate. Sie alle haben unterschiedliche Funktionen. So können zum Beispiel die Methylierung von Adenosin zu N⁶-Methyladenosin (m⁶A) und die anschließende Demethylierung die Homöostase der mRNA regulieren. Es wird angenommen, dass m⁶A hauptsächlich durch die α -Ketoglutarat-abhängige Oxidase ALKBH5 demethyliert wird. In dieser Studie wird gezeigt, dass ALKBH5 auch das dimethylierte Adenosin m⁶₂A, eine nicht-kanonische Modifikation, die in ribosomaler RNA vorkommt, demethylieren kann.

Autorenbeitrag

Für diese Studie wurde von mir eine neuartige UHPLC-MS/MS Methode zur Messung und Quantifizierung der beiden RNA Modifikationen m⁶A und m⁶₂A entwickelt. Dafür wurden die Proben von mir verdaut, mit dieser Methode gemessen und ausgewertet. Die Interpretation der erhaltenen Daten wurde von den Erstautoren, zusammen mit *Markus Müller* und *Thomas Carell*, gemeinsam vorgenommen.

Lizenz

Kopie der Publikation mit Erlaubnis des Verlags.



Cite this: DOI: 10.1039/c8cc03980a

 Received 18th May 2018,
Accepted 4th July 2018

DOI: 10.1039/c8cc03980a

rsc.li/chemcomm

ALKBH5-induced demethylation of mono- and dimethylated adenosine†

 Timm T. Ensfelder,‡^a Matthias Q. Kurz, ^{‡a} Katharina Iwan, ^{‡a} Simon Geiger,^a Sarah Matheisl,^b Markus Müller, ^a Roland Beckmann^b and Thomas Carell ^{*a}

RNA contains methylated A-base derivatives. A methylation to m⁶A and then demethylation regulate homeostasis in mRNA. It is assumed that m⁶A is mainly demethylated by the α -ketoglutarate dependent oxidase ALKBH5. Here we show that ALKBH5 also demethylates the dimethylated adenosine m⁶₂A, which is a non-canonical base present in ribosomal RNA.

RNA contains a large number of modified nucleosides.^{1,2} Most abundant throughout all RNA species are mono-methylated or multiply methylated nucleosides.² Particularly high levels of methylated nucleosides are detected in transfer RNA (tRNA), but also ribosomal RNA (rRNA) and messenger RNA (mRNA) feature such non-canonical nucleosides.^{1,3} Basically, for all canonical nucleosides methylated derivatives exist.¹ They all serve different functions.⁴ The methylated RNA nucleoside m⁶A was first reported in mRNA in 1974^{5,6} and is among the best studied modifications within the set of methylated adenosines (Fig. 1).⁷ Others^{8–14} and we¹⁵ have shown that the nucleoside recruits and repels proteins to control the lifetime of the corresponding mRNA. As such the nucleoside is directly involved in regulating mRNA homeostasis.¹⁵ Particularly exciting is the observation that the nucleoside m⁶A can be demethylated by α -ketoglutarate dependent oxidases,¹⁶ which opens the possibility that methylation and demethylation establish a new layer of regulation to facilitate transcriptional control. In addition to m⁶A, ribosomal RNA also contains the double methylated species m⁶₂A, where two methyl groups are both situated at the exocyclic N⁶-heteroatom. Because the methyl groups are connected to a heteroatom, this oxidation would allow oxidative double demethylation of m⁶₂A in contrast to m²A and m⁸A, where the methyl

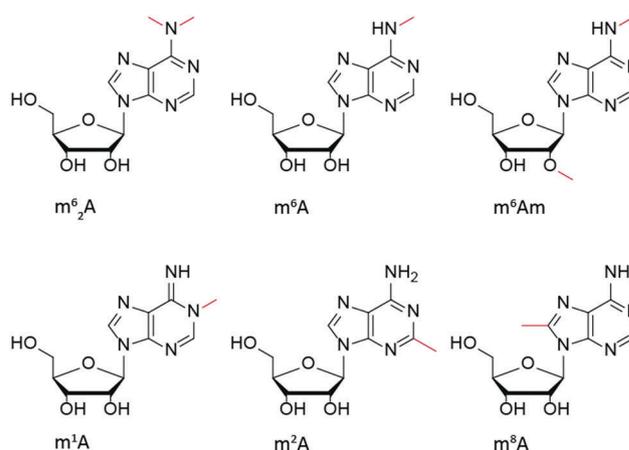


Fig. 1 Selection of naturally occurring methylated adenosines in RNA.

group resides on a carbon. Such sequential demethylations performed by a single enzyme are known: for example, histone demethylases are able to demethylate mono-, di- and even trimethylated lysines.^{17,18} Here, we investigated if the Fe²⁺-, α -ketoglutarate-dependent ALKBH5, which is known to demethylate m⁶A, would be able to demethylate naturally occurring m⁶₂A as well. In contrast to m⁶A, however, m⁶₂A is predominately, if not exclusively, found in a conserved sequence context of the small ribosomal subunit.^{2,3} For this study, we first investigated the demethylation of small RNA strands containing m⁶A and m⁶₂A in different sequence contexts with ALKBH5_{66–292} *in vitro*. Therefore, we synthesized the phosphoramidites of m⁶A and m⁶₂A and subsequently used them for solid phase RNA synthesis (SPOS) of oligoribonucleotides (ORN) 1–4 (Fig. 2). In this study we inserted m⁶A and m⁶₂A into their natural sequence context, *i.e.*, 5'-GGm⁶ACU-3' for m⁶A and 5'-GUGm⁶₂A-m⁶₂ACUU-3' for m⁶₂A. To address potential sequence bias during the enzymatic studies, we also inserted m⁶A into the sequence of m⁶₂A and *vice versa*. In addition to the investigation with pure oligonucleotides, the study was furthermore complemented with the putative demethylation of m⁶₂A in intact ribosomes and in human ribosomal rRNA.

^a Center for Integrated Protein Science, Department of Chemistry, Ludwig-Maximilians-Universität München, Butenandtstraße 5-13, 81377 Munich, Germany. E-mail: Thomas.carell@lmu.de; Web: www.carellgroup.de

^b Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Straße 25, 81377 Munich, Germany

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8cc03980a

‡ Equal contributions.

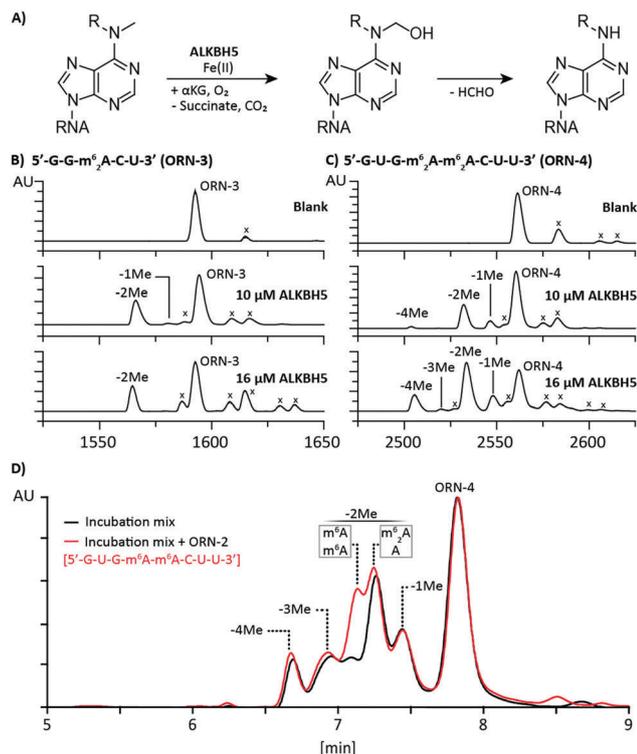


Fig. 2 (A) Oxidation scheme of methylated adenine by ALKBH5 with R being a hydrogen or a methyl group. (B and C) MALDI-TOF spectra of ORN 3–4 and their demethylation products ($-x\text{Me}$) at different concentrations of ALKBH5_{66–292}. The sodium signals in the spectra are marked with an “x”. (D) HPL chromatogram of the incubation of ORN-4 with 16 μM ALKBH5_{66–292} (black line) and with a spike of ORN-2 (red line).

The His₆-tagged protein human ALKBH5_{66–292} enzyme was overexpressed in *E. coli* BL21(DE3) from a pNIC28–Bsa4-ALKBH5_{66–292} plasmid¹⁹ using a modified procedure compared to the literature²⁰ data (see the ESI†) and isolated in one step *via* Ni-NTA affinity chromatography. This procedure yielded the protein in a purity of 90% (ESI,† Fig. S1). The buffer was changed to the assay buffer (50 mM Tris HCl, pH 7.5, 15 mM KCl). Stock solutions of diammonium-Fe(II)-sulfate and α -ketoglutarate were prepared fresh for every assay and added to a final concentration of 300 μM . ALKBH5_{66–292} concentrations in the assay buffer were either 10 or 16 μM . The corresponding oligonucleotides were held constant at 10 μM . Exclusion of oxygen was not needed. The solution was incubated at 37 °C for about 1 h (see the ESI†). For analysing the demethylation reaction, we desalted the assay solution using a Millipore MF™ VSWP-membrane (0.025 μm) and measured the molecular weight of the oligonucleotides using MALDI-TOF mass spectrometry. The obtained data are depicted in Fig. 2.

As expected, we observed complete demethylation of m⁶A in both sequence contexts ORN-1 and ORN-2 (ESI,† Fig. S2A and B). This is interesting because it shows that the oxidation of m⁶A, putatively to the corresponding N⁶-hydroxymethyl-A followed by hydrolysis of the hemiaminal (Fig. 2A), has no observable sequence bias under our conditions. In addition, we see that the demethylation proceeds also at two directly adjacent m⁶A to completion.

When we analysed the demethylation with the m⁶₂A strands ORN-3 and ORN-4 (Fig. 2B and C), we obtained more complex spectra. We could indeed detect demethylation but with a significantly reduced efficiency. When m⁶₂A is placed in the context of the m⁶A consensus sequence, the formation of the mono-demethylated ORN-3 at 10 μM enzyme concentration is seen, furthermore the fully demethylated ORN-3 (-2Me) is clearly visible (Fig. 2B). The conversion is about 40%, estimated based on the observed signal intensities. Demethylation in the rRNA context, where two directly adjacent m⁶₂A occur, is shown in Fig. 2C. At 10 μM and an incubation time of 1 h, mono-, di-, and tetrademethylations are observed. At a slightly higher enzyme concentration, we detected all expected species with the loss of one–four methyl groups. The loss of two methyl groups can yield two different species, either one with m⁶A–m⁶A or one with m⁶₂A–A. To address this issue, we further performed LC-MS analysis of the ORN-4 assay mixture with and without a spike of ORN-2. Fig. 2D shows that both species occur and the m⁶₂A–A product is the favoured one. In summary, the data show that ALKBH5_{66–292} is able to fully demethylate an m⁶₂A–m⁶₂A sequence *via* oxidation of all four methyl groups. Interestingly, the mono- and tri-demethylated ORNs do not accumulate. We interpret this effect as an indication that the first demethylation step is rate limiting, while the demethylation of m⁶A proceeds quickly.

We next wanted to study if double-demethylation of m⁶₂A is also possible on intact rRNA. To this end, we isolated the ribosomes from HEK293T cells using ultracentrifugation followed by rRNA isolation. The obtained ribosomes and rRNA are more complex than the synthesized oligonucleotides because they contain one m⁶A per subunit and in addition the m⁶₂A–m⁶₂A sequence. These m⁶₂A positions are highly conserved in all species with only very few exceptions.²¹ For the analysis, due to the large size of the RNA strand, we digested the RNA after the assay to the individual nucleosides and measured all four canonical nucleosides plus the methylated adenosines m⁶A and m⁶₂A *via* UHPLC-MS/MS using our described procedure (see the ESI†). In order to obtain absolute quantitative data, we synthesized the isotopologues of m⁶A and m⁶₂A that were used for internal calibration (see the ESI†). The obtained quantitative MS-data are compiled in Fig. 3. In the control studies performed without enzymes, we clearly detected the m⁶A and m⁶₂A nucleosides at the expected ratios (two of each per rRNA). Upon addition of ALKBH5_{66–292} in two different concentrations, we see a clear and strong reduction of the m⁶A signal, showing that in principle ALKBH5_{66–292} is able to demethylate m⁶A also in ribosomal RNA. Because the sequence context is different, this result argues again that the activity of ALKBH5_{66–292} has no observable sequence preference. One has to note that in these experiments ALKBH5_{66–292} is present in great excess over the used large rRNA. Interestingly, however, m⁶A levels are not reduced to zero which we would expect based on the observed full demethylation with the smaller RNA strands. For m⁶₂A, we also see clearly reduced levels, showing that ALKBH5_{66–292} demethylated m⁶₂A also in rRNA. It is interesting to note that at 16 μM enzyme concentration, in contrast to the ORN, the resulting m⁶A values are higher than those observed at

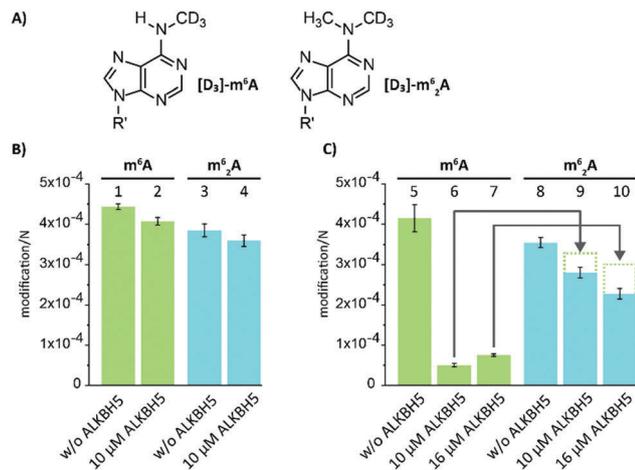


Fig. 3 Quantification data of m^6A and m^6_2A upon incubation with ALKBH5_{66–292} at different concentrations. (A) For quantification, isotope labelled standards of m^6A and m^6_2A were used. R' represents the ribose sugar moiety. Panel (B) represents the amount in ribosomes and (C) the amount in rRNA. The dashed lines shown in (C) indicate the sum of residual m^6A and m^6_2A .

10 μM , despite lower m^6_2A levels. We hypothesize that the observed m^6A levels are the hemi-demethylated intermediates that are obtained when ALKBH5_{66–292} oxidizes and demethylates m^6_2A . The sum of residual m^6A and m^6_2A levels never exceeds the starting values of m^6_2A (Fig. 3D), showing that the full demethylation of m^6_2A occurs and that all the original m^6A was likely demethylated as well.

ALKBH5 is an α -ketoglutarate-dependent oxidase that was shown to demethylate m^6A in mRNA. We investigated here if the enzyme is able to demethylate also the double-methylated m^6_2A , which is known to occur in the small subunit of the ribosome. We also wanted to know if the enzyme is in addition able to demethylate m^6A in rRNA. We show that ALKBH5 demethylates m^6A in basically all sequence contexts, when used at an equimolar ratio or higher, even when two m^6A bases are adjacent to each other. It also demethylates m^6_2A both in rRNA and in small ORNs. Here, in general demethylation occurs and the mono-demethylated intermediate (ORN3/ORN4 -1Me; Fig. 2D and E) is detectable. This is very interesting because it shows that the enzyme releases its substrate after the first demethylation step. Since the demethylation of the mono-methylated compound m^6A is much more efficient compared to m^6_2A , one would expect direct conversion of m^6_2A into adenosine if the substrate is not released. We speculate that the substrate release after the first demethylation could be required to reload another α -ketoglutarate co-substrate into the active site. It is unclear if the demethylation of m^6_2A ALKBH5_{66–292} is a process that occurs in nature. We suspect that the small amount of observed demethylation at approximately 100-fold excess of the enzyme relative to the added rRNA speaks against this possibility. Indeed, when we performed the demethylation studies with assembled ribosomes, demethylation did not occur. This makes demethylation of intact ribosomes unlikely, but it still leaves the possibility that demethylation occurs during rRNA maturation

similar to the insertion of non-canonical nucleosides into mRNA upon mRNA maturation.

We thank Felix Hagelskamp and the Kellner Lab (LMU Munich) for sharing unpublished protocols. We thank Christina Schneider for applying this protocol to help us with the LC-MS measurements. We thank Michael McDonough, Christopher Schofield and Dong Zhang for kindly providing us with the pNIC28–Bsa4-ALKBH5_{66–292} plasmid. Furthermore, we thank the Deutsche Forschungsgemeinschaft (SFB749, SFB1032, GRK2062 and SPP1784) for financial support. Additional support was obtained from the Excellence Cluster CIPSM.

Conflicts of interest

There are no conflicts to declare.

Notes and references

- P. Boccaletto, M. A. Machnicka, E. Purta, P. Piątkowski, B. Bagiński, T. K. Wirecki, V. de Crécy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm and J. M. Bujnicki, *Nucleic Acids Res.*, 2018, **46**, D303–D307.
- M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki and H. Grosjean, *Nucleic Acids Res.*, 2013, **41**, D262–D267.
- D. Piekna-Przybylska, W. A. Decatur and M. J. Fournier, *Nucleic Acids Res.*, 2008, **36**, D178–D183.
- G. Jia, Y. Fu and C. He, *Trends Genet.*, 2013, **29**, 108–115.
- R. Desrosiers, K. Friderici and F. Rottman, *Proc. Natl. Acad. Sci. U. S. A.*, 1974, **71**, 3971–3975.
- R. P. Perry and D. E. Kelley, *Cell*, 1974, **1**, 37–42.
- B. S. Zhao, I. A. Roundtree and C. He, *Nat. Rev. Mol. Cell Biol.*, 2016, **18**, 31–42.
- W. Xiao, S. Adhikari, U. Dahal, Y.-S. Chen, Y.-J. Hao, B.-F. Sun, H.-Y. Sun, A. Li, X.-L. Ping, W.-Y. Lai, X. Wang, H.-L. Ma, C.-M. Huang, Y. Yang, N. Huang, G.-B. Jiang, H.-L. Wang, Q. Zhou, X.-J. Wang, Y.-L. Zhao and Y.-G. Yang, *Mol. Cell*, 2016, **61**, 507–519.
- D. P. Patil, C.-K. Chen, B. F. Pickering, A. Chow, C. Jackson, M. Guttman and S. R. Jaffrey, *Nature*, 2016, **537**, 369–373.
- X. Wang, B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi and C. He, *Cell*, 2015, **161**, 1388–1399.
- K. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S.-B. Qian and S. R. Jaffrey, *Cell*, 2015, **163**, 999–1010.
- X. Wang, Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan and C. He, *Nature*, 2014, **505**, 117–120.
- K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason and S. R. Jaffrey, *Cell*, 2012, **149**, 1635–1646.
- D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek and G. Rechavi, *Nature*, 2012, **485**, 201–206.
- R. R. Edupuganti, S. Geiger, R. G. H. Lindeboom, H. Shi, P. J. Hsu, Z. Lu, S.-Y. Wang, M. P. A. Baltissen, P. W. T. C. Jansen, M. Rossa, M. Müller, H. G. Stunnenberg, C. He, T. Carell and M. Vermeulen, *Nat. Struct. Mol. Biol.*, 2017, **24**, 870–878.
- G. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C.-M. Huang, C. J. Li, C. B. Vågbo, Y. Shi, W.-L. Wang, S.-H. Song, Z. Lu, R. P. G. Bosmans, Q. Dai, Y.-J. Hao, X. Yang, W.-M. Zhao, W.-M. Tong, X.-J. Wang, F. Bogdan, K. Furu, Y. Fu, G. Jia, X. Zhao, J. Liu, H. E. Krokan, A. Klungland, Y.-G. Yang and C. He, *Mol. Cell*, 2013, **49**, 18–29.
- K. Yamane, C. Toumazou, Y.-I. Tsukada, H. Erdjument-Bromage, P. Tempst and Y. Zhang, *Cell*, 2006, **125**, 483–495.
- S. M. Kooistra and K. Helin, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 297–311.
- W. Aik, J. S. Scotti, H. Choi, L. Gong, M. Demetriades, C. J. Schofield and M. A. McDonough, *Nucleic Acids Res.*, 2014, **42**, 4741–4754.
- F. W. Studier, *Protein Expression Purif.*, 2005, **41**, 207–234.
- Z. Xu, H. C. O'Farrell, J. P. Rife and G. M. Culver, *Nat. Struct. Mol. Biol.*, 2008, **15**, 534–536.

4 Unveröffentlichte Arbeiten

4.1 Studie zur biologischen Funktion der Hypermodifikation ms^2i^6A

4.1.1 Prolog

Wie in Abschnitt 1.2 beschrieben, ist die RNA im Gegensatz zur DNA sehr viel stärker modifiziert und auch die strukturelle Vielfalt bekannter Ribonukleotide ist erheblich größer. Besonders interessant sind hierbei die sogenannten Hypermodifikationen. Diese sind Nucleoside mit langen Seitenketten oder Nucleoside die einen veränderten Heterozyklus aufweisen. Zwei dieser hypermodifizierten Nucleoside sind N^6 -Isopentenyladenosin (i^6A) und N^6 -Isopentenyl-2-thiomethyladenosin (ms^2i^6A) (Abb. 10 A). In humanen Zellen beginnt ausgehend von Adenosin die zweistufige Biosynthese von ms^2i^6A mit der Einführung der Isopentenyl-Gruppe durch das Enzym TRIT1. Das so entstandene i^6A wird anschließend durch CDK5RAP1 thiomethyliert. Aufgrund der komplexen Biosynthese sind hypermodifizierte Nucleoside nur in wenigen RNA Spezies zu finden, wie zum Beispiel an der Position 37 oder an der Wobble Position der Antikodonschleife der tRNA. Bisher wurden außerhalb von tRNA überwiegend einfache Modifikationen entdeckt. Obwohl man lange Zeit dachte, dass die mRNA fast ausschließlich kanonische Nucleoside enthält, ist mittlerweile bekannt, dass neben der m^7G Cap-Struktur auch m^6A , m^1A und Pseudouridin vorkommen. Ob auch Hypermodifikationen in anderen RNA-Spezies als tRNA vorkommen können, ist bisher noch fraglich.

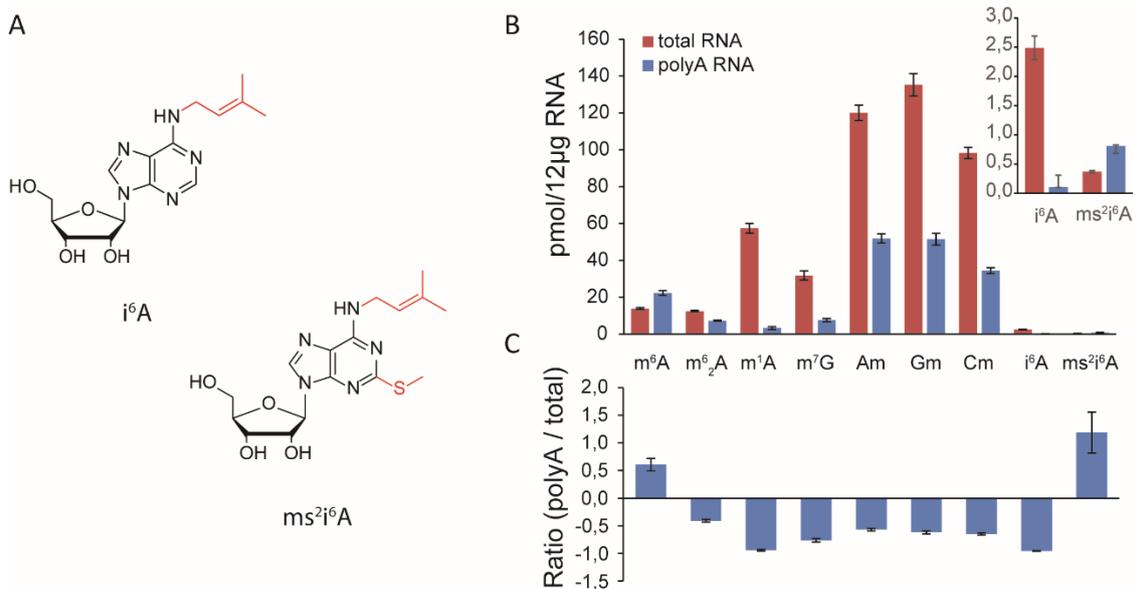


Abbildung 10: A): Die Strukturen der beiden hypermodifizierten Nucleoside i^6A und ms^2i^6A . B): Quantifizierungsdaten verschiedener Purine in Gesamt- und PolyA-RNA aus HeLa-Zellen. C): Berechnung der Verhältnisse der jeweiligen Modifikation von PolyA-RNA zur Gesamt-RNA.^[66]

Dieser Frage ist bereits *Veronika Reiter* in ihrer Dissertation nachgegangen. Die Abbildungen 10 B und C zeigen die Quantifizierung von Purinen der Gesamt- und PolyA-RNA aus HeLa Zellen. Die PolyA-RNA wurde über die Bindung an Oligo-dT-Stränge angereichert. Beide RNA-Spezies wurden anschließend verdaut und über eine hochauflösende LC-MS Methode (Orbitrap XL) mit Isotopenverdünnungstechnik quantifiziert. Erst über die zusätzliche Berechnung des Verhältnisses der jeweiligen Modifikation in PolyA-RNA zur Gesamt-RNA wird deutlich, dass genau zwei dieser Purin-Nukleoside in der PolyA-RNA angereichert vorliegen- zum einen das bereits in mRNA bekannte m⁶A und zum anderen die Hypermodifikation ms²i⁶A. Wie in der Abbildung 10 B gezeigt, verhält sich der direkte Vorläufer i⁶A genau gegenläufig dazu. Dieses Ergebnis ist der erste Hinweis darauf, dass auch ein hypermodifiziertes Nukleosid in mRNA bzw. PolyA-RNA vorkommen kann und sollte in dieser Arbeit weiter erforscht werden.

4.1.2 Ergebnis und Diskussion

Im Zuge dieser Arbeit wurde eine neue Quantifizierungsmethode basierend auf UHPLC-MS/MS Messungen mittels der Isotopenverdünnungstechnik entwickelt. Im Grunde stellt diese Methode eine Verbesserung zu der zuvor angewandten LC-MS Methode mit der Orbitrap XL dar. Der Nachteil an dieser Messmethode war, dass mindestens 12 µg RNA verdaut und gemessen werden mussten, um einen validen Datensatz für die Modifikation $ms^{2i}A$ zu generieren (siehe Abb. 10 B, C). Die Methode wurde nun für den sensitiveren Massendetektor Triplequadropol entwickelt. Da es noch keine RNA-Methode für dieses Gerät gab, musste auch die Messmethode bzw. Quantifizierung der kanonischen RNA-Basen bewerkstelligt werden. Hierfür wurden UV-Kalibrationsgeraden für die vier kanonischen RNA-Nukleoside aufgenommen (Abb. 11 und Tabelle 1), denn im Gegensatz zur DNA ist die Zusammensetzung der RNA-Nukleoside nicht zwangsläufig bekannt bzw. abhängig von der RNA-Spezies. Deshalb müssen alle vier kanonischen Nukleoside über die UV-Spur quantifiziert werden.

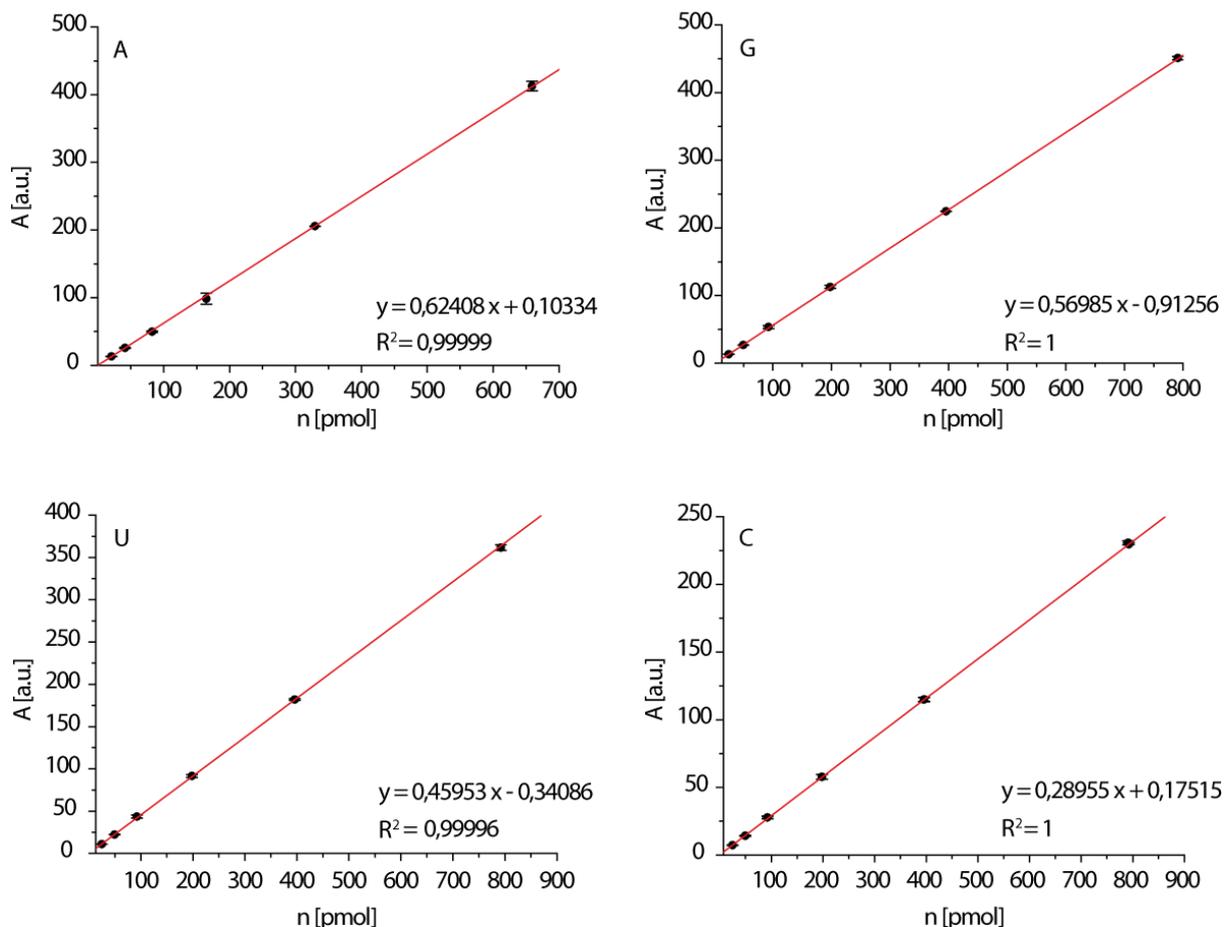


Abbildung 11: Kalibrationsgeraden der vier kanonischen RNA-Nukleoside mit der jeweiligen Geradengleichung und dem Bestimmtheitsmaß R^2 .

Tabelle 1: Die UV-Kalibrationsgeraden der vier kanonischen Nucleoside mit der jeweiligen Geradengleichung und den gültigen Bereichen.

Nucleosid	Geradengleichung	n (ULOQ) [pmol]	n (LLOQ) [pmol]
C	$y=0.28955x+0.17515$	791	24,7
U	$y=0.45953x-0.34086$	792	24,7
G	$y=0.56985x-0.91256$	792	24,7
A	$y=0.62408x-0.10334$	659	20,6

Da die beiden Modifikationen i^6A und $ms^{2i^6}A$ Derivate des Adenosins sind, wäre auch eine Berechnung und Abbildung der Daten pro Adenosin denkbar, würde aber bei dem Vergleich von verschiedenen RNA-Spezies durch z.B. den PolyA-Schwanz Einfluss auf die absoluten Werte nehmen. Zusätzlich zu den kanonischen Nucleosiden wurden anschließend die internen Kalibrationsgeraden der beiden Modifikationen $ms^{2i^6}A$ (Abb. 12 A, B und Tabelle 2) und i^6A (Abb. 12 C, D und Tabelle 2) aufgenommen.

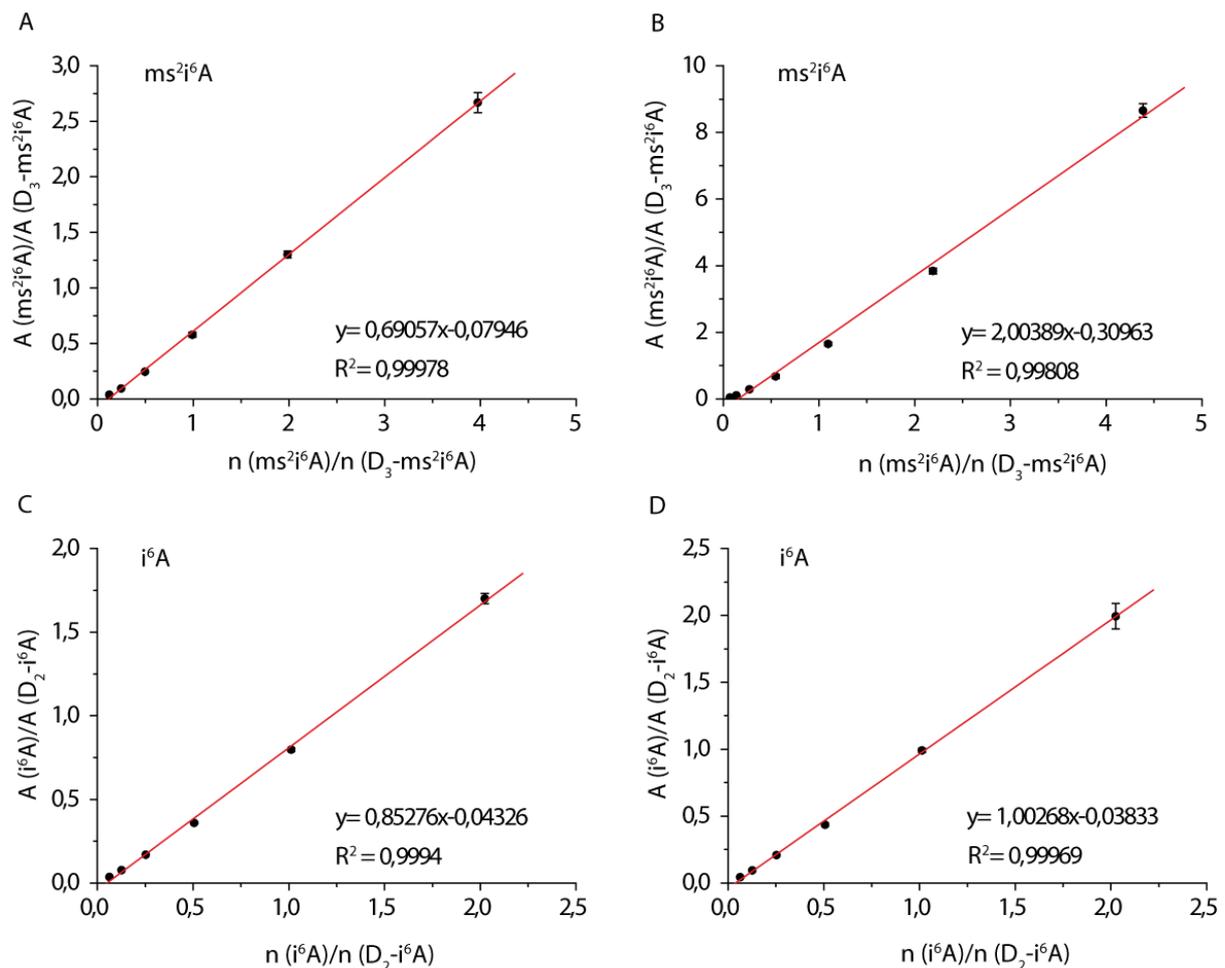


Abbildung 12: Die internen Kalibrationsgeraden für die exakte Quantifizierung der Modifikationen $ms^{2i^6}A$ (siehe A) und B)) und i^6A (siehe C) und D)) mit Geradengleichung und Bestimmtheitsmaß R^2 .

Es sind jeweils zwei Kalibrationsgeraden pro Modifikation, da Kalibrationsgeraden nur in ihrem jeweiligen Bereich (Stoffmenge und Verhältnis der Intergrale) valide sind und es nicht möglich war, eine Kalibrationsgerade über einen größeren Bereich aufzunehmen. Diese Bereiche sind in der Tabelle 2 aufgelistet.

Tabelle 2: Die internen Kalibrationsgeraden der beiden Nukleoside ms^2i^6A und i^6A mit ihren jeweiligen Geradengleichungen und den gültigen Bereichen.

Nukleosid	ULOQ [pmol]	LLOQ [pmol]	ULOQ A/A*	LLOQ A/A*
ms^2i^6A (A)	0,641	0,0401	2,66	0,079
ms^2i^6A (B)	3,04	0,189	8,47	0,239
i^6A (C)	0,101	0,00634	1,68	0,064
i^6A (D)	0,403	0,0250	1,99	0,080

Mit diesen Kalibrationsgeraden, einer passenden Chromatographie über die UHPLC und einem geeigneten RNA-Verdauprotokoll konnte mit dieser Messmethode die Menge an benötigter RNA von 12 μg auf 1 μg gesenkt werden. Im Hinblick auf die Isolation von PolyA-RNA ist das ein großer Vorteil, denn nur ca. 1 - 2,5 % der isolierten Gesamt-RNA können als PolyA-RNA gewonnen werden. In der Abbildung 13 ist ein TBE-Agarose-Gel der beiden RNA Präparationen gezeigt. Die RNA-Präparationen wurden unabhängig von der jeweiligen Zelllinie auf die gleiche Weise durchgeführt.

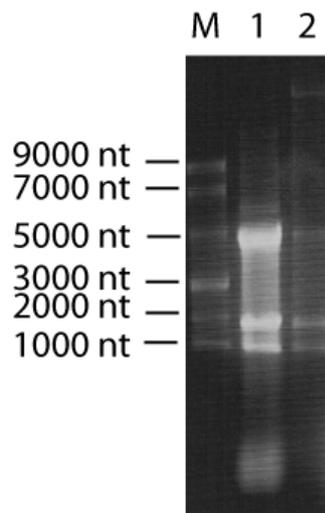


Abbildung 13: Vergleich der verschiedenen RNA Präparationen aus HeLa-Zellen auf einem 1,5 % TBE-Agarose-Gel. M: *ssRNA ladder*; 1: Gesamt-RNA (4,4 μg); PolyA-RNA (599 ng).

Für die Isolation der Gesamt-RNA wurde das Trizol-Reagenz verwendet, dessen Funktionsweise auf einer Phenol-Chloroform-Extraktion des Zelllysats beruht. Die einphasige Lösung enthält neben Phenol auch Guanidinisothiocyanat (GTC), welches die Zellen lysiert und gleichzeitig RNasen und andere enthaltene Enzyme inhibiert. Die anschließende Zugabe von Chloroform und Zentrifugation führt zur Phasentrennung in drei verschiedenen Phasen. In der oberen wässrigen Phase ist die RNA gelöst, während die Proteine in der unteren organischen Phase vorkommen und die DNA vor allem in der Interphase vorliegt. Die RNA wird aus der wässrigen Phase mittels Isopropanol gefällt. Da diese Methode aber nicht zwischen den einzelnen RNA-Spezies unterscheiden kann, ist sie nur für die Isolation der Gesamt-RNA geeignet.

Für die Isolation der PolyA-RNA wurde das *Oligotex mRNA Midi Kit* von *Qiagen* verwendet. Die Isolation der PolyA-RNA beginnt mit der bereits isolierten Gesamt-RNA. Diese wird auf sogenannte *Spin Columns* aufgetragen, welche Polystyrol-Latex Partikel mit dC₁₀T₃₀ Oligonukleotiden enthalten, die kovalent an der Oberfläche gebunden sind. Dadurch wird eine effiziente Bindung aller polyadenylierten RNAs gewährleistet. Da sich insbesondere mRNA durch ihren PolyA-Schwanz auszeichnet, geht man davon aus, dass es sich bei der isolierten PolyA-RNA vor allem um mRNA handelt. Die Präparation der PolyA-RNA zeigt auf dem Agarose-Gel (Abb. 13) den charakteristischen „Schmier“ in einem Bereich von 500 – 6000 nt. Im Vergleich zur isolierten Gesamt-RNA sind jedoch die Banden korrespondierend zur 28 S rRNA im Bereich von 5000 nt (4712 nt) und zur 18 S rRNA bei etwa 2000 nt (1874 nt) nur schwach ausgeprägt. Die Banden, die kleineren RNA Fragmenten zuzuordnen sind (tRNA (50-100 nt), 5,8 S rRNA (160 nt), 5 S rRNA (120 nt)), sind bei Präparation der PolyA-RNA im Vergleich zur Präparation der Gesamt-RNA verschwunden. Daraus ist zu erkennen, dass zwar eine Anreicherung der gewünschten PolyA-RNA gelungen ist, innerhalb der Präparation jedoch keine vollständige Separation von anderen RNA-Spezies erzielt werden konnte.

Sowohl von der Gesamt-RNA als auch von der PolyA-RNA der jeweiligen Zellkultur wurden drei biologische Replikate analysiert. Je biologischem Replikat wurden drei technische Replikate verdaut und mit der sensitiveren UHPLC-MS/MS Methode gemessen. In der Abbildung 14 sind die Quantifizierungsergebnisse der beiden Modifikationen ms²ⁱ⁶A und i⁶A pro N (alle Ribonukleoside) abgebildet. In allen Proben wurden diese beiden Modifikationen, wenn auch zum Teil in sehr geringen Mengen, nachgewiesen. Die Mengen an ms²ⁱ⁶A und i⁶A in Gesamt-RNA sind in allen Zellkulturen in etwa gleich (blaue Säulen in Abb. 14 A, B). Im Gewebe, z.B. der Leber, hingegen sind die Mengen an ms²ⁱ⁶A ungefähr zehn Mal so groß. Dieses Ergebnis könnte man durch die viel größere Menge an tRNA in Geweben erklären, welche wiederum stärker modifiziert vorliegen als die in Zellkulturen.^[164] Auch die PolyA-RNA Messungen enthalten ms²ⁱ⁶A, die Mengen an i⁶A sinken im Vergleich zur Gesamt-RNA um ca. eine Zehnerpotenz durch die Anreicherung von PolyA-RNA. Das ist ein typisches Ergebnis für

eine Modifikation, die vor allem in tRNA vorkommt. Ganz im Gegenteil dazu steigt die Menge an ms^2i^6A durch die Anreicherung von PolyA-RNA um mindestens das Doppelte an. Nur die PolyA-RNA Anreicherungen der Maus-Gewebe (Abb. 14 C) liefert nicht denselben Anreicherungsfaktor von ms^2i^6A wie in den Zellkulturproben. Der Grund hierfür ist wieder die große Menge an stark modifizierter tRNA. Aber dadurch, dass die Menge an ms^2i^6A weniger deutlich durch die Isolation der PolyA-RNA absinkt als die Menge von i^6A , könnte man davon ausgehen, dass die Hypermodifikation ms^2i^6A in PolyA-RNA vorkommt.

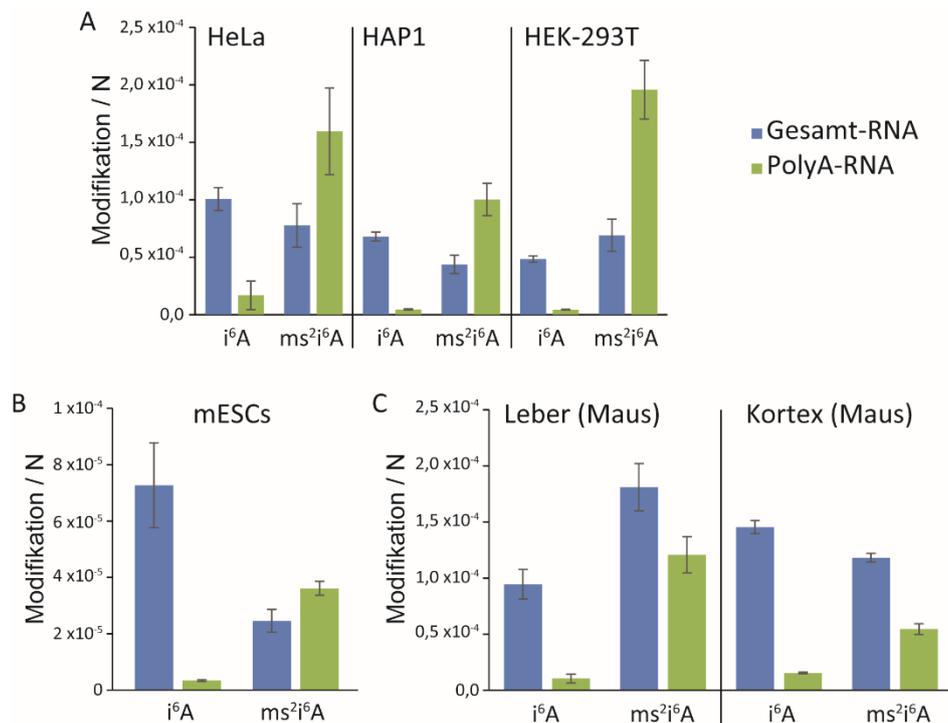


Abbildung 14: A): Quantifizierungsdaten der beiden Modifikationen i^6A , ms^2i^6A in Gesamt- und PolyA-RNA der HeLa-, HAP1- und HEK293T-Zelllinie. B): Quantifizierungsdaten der beiden Modifikationen i^6A , ms^2i^6A in Gesamt- und PolyA-RNA in mESCs. C): Quantifizierungsdaten der beiden Modifikationen i^6A , ms^2i^6A in Gesamt- und PolyA-RNA aus Maus-Gewebe (Leber und Kortex).

Wie in Abbildung 13 erkennbar, ist noch immer rRNA in der Probe enthalten. Aus diesem Grund sollte zusätzlich ein *GeneRead rRNA Depletion Kit* von Qiagen verwendet werden. Dieses erlaubt die selektive Abreicherung von rRNA. Das Kit verwendet spezifische DNA-Oligonukleotide, die mit den rRNA Sequenzen (18 S, 28 S, 5 S, 5,8 S, 12 S und 16 S) hybridisieren können. Das resultierende RNA:DNA Hybrid kann über einen Antikörper erkannt werden. Der entstandene Komplex aus Antikörper und RNA:DNA Hybrid wird an Protein G *Beads* effizient gebunden und diese Beads werden von der Probe abgetrennt, wodurch die rRNA abgereichert wird. Diese stufenweise Aufreinigung sollte eine noch bessere Reinheit der PolyA-RNA liefern. Die Abbildung 15 zeigt wie zuvor einen Anstieg der ms^2i^6A -Menge durch die 1. PolyA-RNA Anreicherung, dann jedoch eine Abnahme nach dem 2. PolyA-RNA Anreicherungsschritt und eine weitere Abnahme nach der rRNA Abreicherung. Aus diesem

Experiment wird deutlich wie schwierig es ist, reine PolyA-RNA zu isolieren, dadurch, dass aber auch nach all diesen Schritten noch eine messbare Menge an $ms^{2i6}A$ vorhanden ist, könnte man schließen, dass es tatsächlich in PolyA-RNA vorkommt. Und da es sich bei PolyA-RNA vor allem um polyadenylierte mRNA handelt, könnte man davon ausgehen, dass $ms^{2i6}A$ die erste Hypermodifikation in mRNA darstellt.

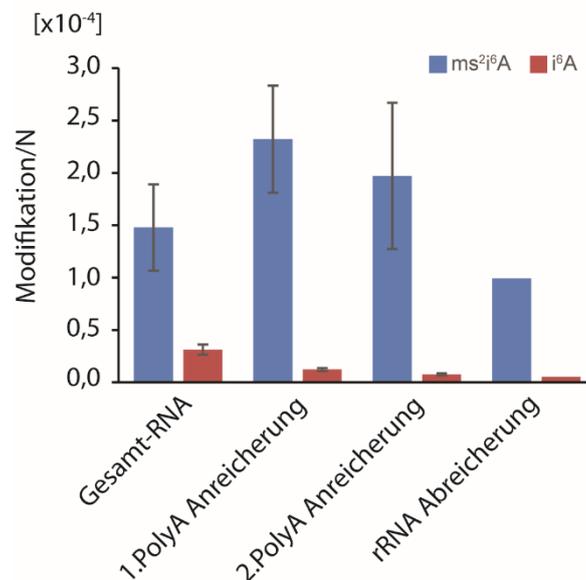


Abbildung 15: Quantifizierungsergebnisse der Modifikationen $ms^{2i6}A$ und i^6A in mESC im Verlauf einer stufenweisen Aufreinigung von Gesamt-RNA über eine 1. PolyA-RNA Anreicherung, einer 2. PolyA-RNA Anreicherung und einer anschließenden rRNA Abreicherung.

Da Veronika Reiter neben der PolyA-RNA auch eine Anreicherung der $ms^{2i6}A$ -Mengen in Präparationen von RNA < 200 nt gefunden hat und diese Fraktionen neben der tRNA auch miRNA und Intron Fragmente enthält, wurde gemutmaßt, dass eine Beteiligung von $ms^{2i6}A$ im Splicingprozess existiert. Dazu wurde ein Experiment mit dem Splicing Inhibitor Isoginkgetin durchgeführt. Dieser Inhibitor verhindert die Bildung des U4/U5/U6-Ribonukleoproteins und führt dadurch zu einer Anhäufung des prä-spliceosomalen A-Komplex, d.h. zu einer höheren Konzentration der prä-mRNA.^[165] Je nach Position der Modifikation in der PolyA-RNA könnte es entweder zu einer Anreicherung oder einer Abreicherung des $ms^{2i6}A$ -Gehalts in der isolierten PolyA-RNA kommen. Es wäre vorstellbar, dass sich $ms^{2i6}A$ in Introns und/oder im untranslatierten Bereich befindet, denn dort könnte die Sekundärstruktur *Hairpin*-artige Strukturen stabilisieren. Damit hätte $ms^{2i6}A$ auch dort eine ähnliche Funktion wie in der tRNA. Vor dem eigentlichen Experiment muss die Effizienz des Splicing Inhibitors Isoginkgetin bestimmt werden. Hierfür wurden die HAP1-Zellen für 6 h mit 50 μ M Isoginkgetin (gelöst in 0,1 % DMSO) inkubiert. Die Negativkontrolle wurde auch mit 0,1 % DMSO behandelt, um auf mögliche Lösungsmittelleffekte normieren zu können. Nach der Isolation der Gesamt-RNA

wurden die Proben mit DNaseI behandelt, um eventuelle DNA-Rückstände zu entfernen. Anschließend wurde die RNA in cDNA umgeschrieben und es wurde eine qPCR durchgeführt. Hierfür wurden ein *forward Primer* und zwei *reverse Primer* verwendet. Der eine *reverse Primer* ist komplementär zu einer Intronsequenz und der andere zur darauffolgenden Exonsequenz. Über die in Abbildung 16 gezeigte Formel lässt sich so das Verhältnis von prä-mRNA zu mRNA bestimmen.^[166] Durch die Isoginkgetin Behandlung konnte die relative Expression der prä-mRNA um 50 % gesteigert werden.

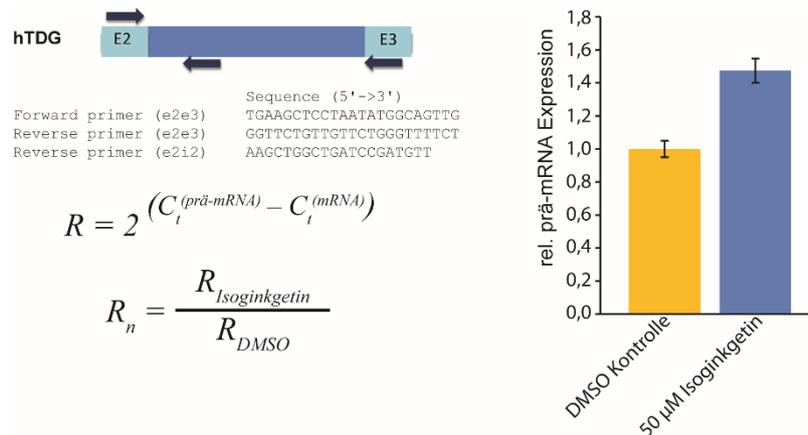
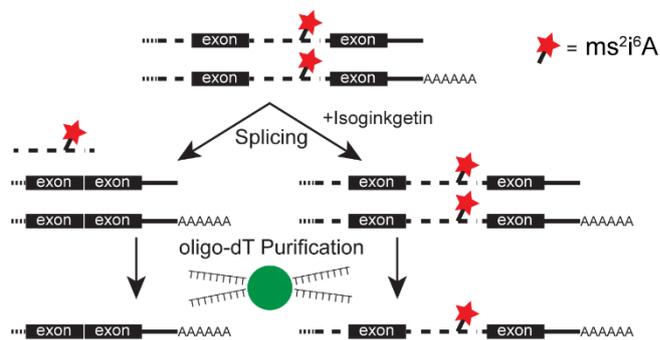


Abbildung 16: Berechnung der relativen prä-mRNA Expression über die gezeigte Formel durch die Verwendung von zwei verschiedenen *reversen* Primern, wobei der eine Primer komplementär zur Intronsequenz und der andere komplementär zur Exonsequenz ist. Die relative prä-mRNA Expression der DMSO Kontrolle ist in gelb gezeigt und wird mit der Isoginkgetin-behandelten Probe (blau) verglichen. Diese enthält bis zu 50 % mehr prä-mRNA als die Negativkontrolle.

Um nun mehr über die mögliche Beteiligung von $ms^{2i^6}A$ im Splicingprozess und dessen Position herauszufinden, wurde die PolyA-RNA der Isoginkgetin behandelten sowie die der Negativkontrolle, isoliert und verdaut. Die Funktionsweise dieses Experiments ist schematisch in der Abbildung 17 A gezeigt. Die darauffolgende Quantifizierung des Gehalts an $ms^{2i^6}A$ und i^6A mittels UHPLC-MS/MS zeigte, dass die Inhibition des Splicings zu einer kleinen, aber signifikanten Erhöhung der Menge an $ms^{2i^6}A$ und i^6A führte (Abb. 17 B). Die Behandlung durch Isoginkgetin führte sowohl zu einer Erhöhung der relativen Expression von prä-mRNA als auch zu einem höheren Gehalt an $ms^{2i^6}A$ und i^6A . Das ist kein Beweis, dass die Modifikation $ms^{2i^6}A$ das Splicing beeinflusst. Viel mehr deutet es darauf hin, dass die Mengen an $ms^{2i^6}A$ durch die Zugabe des Splicing Inhibitors beeinflusst werden. Ein Grund könnte zum Beispiel auch sein, dass $ms^{2i^6}A$ vor allem in Introns oder übersprungenen Exons vorkommt, die durch das Splicing für gewöhnlich entfernt werden. Da auch die Menge an i^6A erhöht wird, könnte man eher einen sekundären Effekt des Splicing Inhibitors für die Zunahme der beiden Modifikationen verantwortlich machen. Da es sich bei der Modifikation i^6A um keine Modifikation der mRNA handelt (siehe Abb. 10), ist noch immer fraglich, ob $ms^{2i^6}A$ tatsächlich in mRNA vorkommt.

A



B

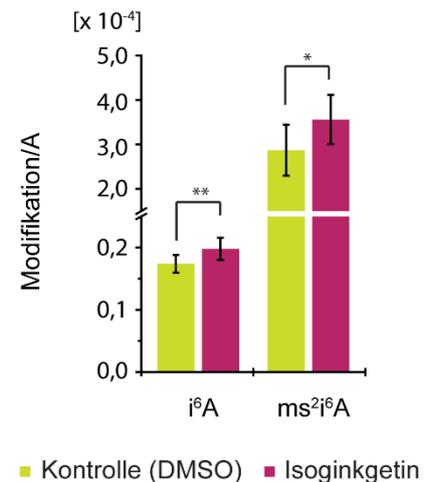


Abbildung 17: A): Schematische Darstellung der Funktionsweise die diesem Experiment zugrunde liegt. Es wird überprüft ob die Zugabe des Splicing Inhibitors Isoginkgetin Auswirkungen auf die Menge an i⁶A und ms²i⁶A durch die Zunahme von prä-mRNA hat. B): Quantifizierungsdaten von i⁶A und ms²i⁶A in PolyA-RNA aus Zellen, die entweder mit DMSO (gelb) oder mit Isoginkgetin (pink) behandelt wurden.

Da die vorangegangenen Experimente noch keine endgültigen Schlüsse zuließen, wurde entschieden den Phänotyp von ms²i⁶A Knockout Zellen genauer zu untersuchen, um Hinweise auf die Funktion dieser Hypermodifikation zu erzielen. Hierfür wurde eine CDK5RAP1-Knockout-Zelllinie bei der Firma *Haplogen* bestellt wurde. Wie bereits in 1.2.1 beschrieben, ist CDK5RAP1 das humane Ortholog zum bakteriellen MiaB Protein, welches für die Umwandlung von i⁶A zu ms²i⁶A verantwortlich ist.^[82] Die Knockout-Zelllinie wurde mit Hilfe des CRISPR/Cas9 Systems in der HAP1 Zelllinie generiert. HAP1 Zellen sind humane haploide Zellen, die aus KBM-7 Zellen abstammen.^[167] In das CDK5RAP1-Gen wurde eine Deletion von 5 bp in das Exon 3 eingeführt, wodurch das Leseraster stromabwärts verschoben wurde und so in einem verfrühten Stoppkodon (PTC: *premature termination codon*) resultieren sollte. Die daraus entstehende verkürzte mRNA wird über den sogenannten *nonsense-mediated mRNA decay pathway* abgebaut und das Gen somit nicht translatiert. Dieser Kontrollmechanismus eukaryotischer Zellen verhindert eine Expression verkürzter und nicht funktionsfähiger Proteine.^[168] Vom Hersteller *Haplogen* wird jedoch nur die eingeführte Mutation garantiert. Ob diese Mutation in der erhaltenen Zelllinie tatsächlich zu einem Knockout des Gens, das für das CDK5RAP1 Enzym codiert, führt, muss daher vor der Verwendung der Zelllinie im Rahmen weiterer Experimente verifiziert werden. Eine Möglichkeit der Verifikation ist die Reverse Transkriptase-Polymerase-Kettenreaktion (RT-PCR). Hierfür wurde zunächst die isolierte Gesamt-RNA aus den HAP1- und CDK5RAP1 Knockout-Zellen mit Hilfe einer Reversen Transkriptase in cDNA umgeschrieben. Die anschließende PCR wurde mit einem Primer-Paar durchgeführt, das spezifisch für eine bestimmte Region im CDK5RAP1-Gen ist. Die eingeführte Mutation sollte im Falle eines tatsächlich vorliegenden Gen-Knockouts zum Abbau

der mRNA führen und demnach keine Amplifikation der cDNA möglich sein. Die nach der RT-PCR erhaltene cDNA beider Zelllinien bzw. die Banden auf dem 1,5 % TAE-Agarose-Gel sahen nahezu identisch aus (Abb. 18 A), was gegen eine CDK5RAP1-Knockout Mutante spricht. Jedoch kann der *nonsense-mediated mRNA decay pathway* nur dann initiiert werden, wenn sich das vorzeitige Stoppkodon mindestens 50 – 55 Basenpaare (in 5'-3' Richtung) vor der letzten Exon-Exon-Brücke befindet.^[168] Anstatt im Folgenden eine quantitative RT-PCR durchzuführen, wurde direkt eine phänotypische Analyse angestrebt, um den Gen-Knockout nachzuweisen. Hierfür wurde sich zu Nutzen gemacht, dass durch das Fehlen des Proteins keine Umsetzung von i^6A zu ms^2i^6A stattfinden kann, sodass in der RNA der Zellen bei einem vorliegenden Gen-Knockout kein ms^2i^6A vorhanden sein dürfte.

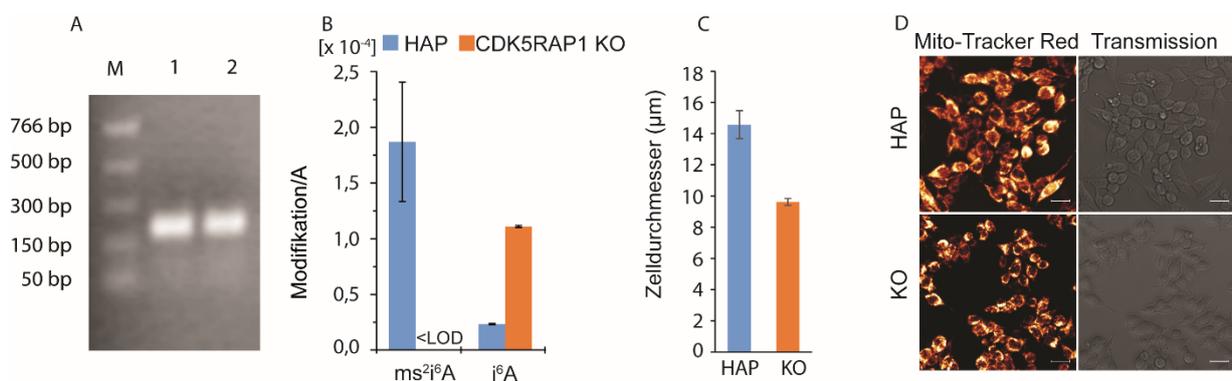


Abbildung 18: Auswirkungen des CDK5RAP1 Knockouts auf die Zellen. A): 1,5 % TAE-Agarose-Gel der erhaltenen amplifizierten cDNA. M: PCR ladder 1: cDNA aus HAP1 Zellen 2: cDNA aus CDK5RAP1-KO-Zellen. B): Quantifizierungsdaten der beiden Modifikationen i^6A , ms^2i^6A in der HAP1-Wildtyp-Zelllinie und der CDK5RAP1-KO-Zelllinie. C): Messung des Durchmessers der beiden Zelllinien. D): Mikroskopische Untersuchung der Mitochondrien beider Zelllinien durch die Färbung der Mitochondrien mit *MitoTracker Red* (links) und im Durchlicht (rechts). Maßstab: 20 μm .

Der Modifikationsgehalt der Gesamt-RNA beider Zelllinien, HAP1 und CDK5RAP1 Knockout, wurde quantifiziert und wie in der Abbildung 18 B gezeigt, enthält der Knockout kein ms^2i^6A , dafür größere Menge des biosynthetischen Vorläufers i^6A . Dieses Ergebnis zeigt deutlich, dass die eingeführte Mutation tatsächlich zum Knockout des Gens, kodierend für das CDK5RAP1 Enzym, geführt hat. Diese Zelllinie wurde auf weitere phänotypische Unterschiede untersucht, wie zum Beispiel die Zellgröße. Hier zeigte sich (siehe Abb. 18 C), dass die CDK5RAP1 KO Zellen eine um 34 % reduzierte Zellgröße aufweisen. Da ms^2i^6A vor allem in Mitochondrien gefunden wurde, wurde in einem weiteren Experiment die Morphologie der Mitochondrien verglichen. Hierfür wurden die Mitochondrien beider Zelllinien mit dem *Mito Tracker Red* eingefärbt (Abb. 18 D). Es zeigten sich dabei keine auffälligen morphologischen Unterschiede der Mitochondrien, die für die geringere Zellgröße verantwortlich sein könnten. Allerdings sei darauf hingewiesen, dass für eine Aussage zur Funktionsfähigkeit der gebildeten

Mitochondrien andere Marker verwendet werden müssten, die z.B. Membranpotentiale anzeigen, deren Einsatz jedoch deutlich komplizierter ist.

Im Folgenden wurden Mitochondrien dieser HAP1 Zelllinie durch eine differentielle Dichtegradientenzentrifugation gewonnen. Die darin enthaltenen Modifikationen $ms^{2i^6}A$ und i^6A wurden in dieser mitochondrialen RNA quantifiziert (Abb. 19) Die quantifizierten $ms^{2i^6}A$ -Mengen unterliegen dabei großen Schwankungen und sind deshalb nicht vereint gezeigt. Trotzdem wird sichtbar, dass ungefähr 4-mal so viel $ms^{2i^6}A$ wie i^6A in der RNA der Mitochondrien vorkommt. Es ist also davon auszugehen, dass $ms^{2i^6}A$ eine größere Rolle in mitochondrialer RNA einnimmt. Doch auch durch dieses Experiment ist die explizite Herkunft der Hypermodifikation $ms^{2i^6}A$ noch nicht geklärt.

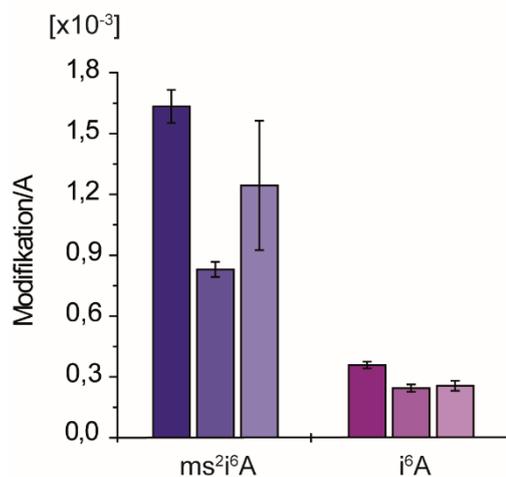


Abbildung 19: Quantifizierungsergebnisse der Modifikationen $ms^{2i^6}A$ und i^6A in RNA aus den Mitochondrien der HAP1 Zelllinie. Die jeweiligen Balken des Diagramms stellen eine biologische Probe dar, welche als technisches Triplikat gemessen wurde.

Nach Betrachtung aller Ergebnisse muss man sagen, dass es noch immer fraglich ist, ob $ms^{2i^6}A$ tatsächlich die erste Hypermodifikation in mRNA ist und welche Funktionen diese Modifikation dort ausübt. *Fakruddin et al.* fanden erst kürzlich heraus, dass CDK5RAP1 spezifisch an der Position 37 in vier verschiedenen mitochondrialen tRNAs i^6A zu $ms^{2i^6}A$ konvertiert. Nach dem Auslösen dieser mitochondrialen tRNAs konnte in den Zellen kein $ms^{2i^6}A$ mehr detektiert werden. Deshalb ist davon auszugehen, dass $ms^{2i^6}A$ eine mitochondriale tRNA Modifikation ist und in anderen RNA-Spezies nicht vorkommt.^[169] Die Frage ist nur, wie es sein kann, dass eine Modifikation, die nur in mitochondrialer tRNA vorkommt, durch eine PolyA-Anreicherung signifikant angereichert werden kann. Ein Grund könnte sein, dass mt-tRNAs, die $ms^{2i^6}A$ enthalten, relativ schwache tertiäre Wechselwirkungen zwischen der D-Schleife und der T-Schleife des Kleeblatts ausbilden und somit viel flexibler sind. Durch die denaturierenden Bedingungen der PolyA-Anreicherung könnte es zum Beispiel zu einer unspezifischen Bindung der A-reichen Sequenz der mt-tRNA an die Oligo-dT

Sequenz kommen.^[169] Des Weiteren ist bekannt, dass auch mt-tRNAs polyadenyliert werden können. Vermutlich bedingt durch ihren bakteriellen Ursprung ist diese Polyadenylierung, anderes als bei der nuklearen mRNA, ein Signal für den Abbau von strukturell und funktionell abnormalen mitochondrialen tRNAs.^[170] Möglicherweise könnten auch diese polyadenylierten mitochondrialen tRNAs durch die PolyA-Anreicherung angezogen werden.

Zusammengefasst muss man sagen, dass es viele experimentelle Hindernisse gibt, wenn erforscht werden soll, ob eine spezielle Modifikation auch in mRNA vorkommt. Eine Möglichkeit wäre, die mRNA nicht über ihren PolyA-Schwanz anzureichern, sondern sie spezifisch über die 5'-Cap-Struktur zu isolieren.

4.1.3 Zusammenfassung der Projektbeiträge

Die UHPLC-MS/MS Methodenentwicklung zur Messung der ms²i⁶A- und i⁶A-Mengen erfolgte durch mich. Die Zellkultur, Isolation und Quantifizierung der Daten aus Zellkultur führte ich durch, die Isolation der RNA-Proben aus Gewebe und der Mitochondrien aus den HAP1 Zellen meine damalige Masterstudentin *Sarah Schiffers*. Die Charakterisierung der CDK5RAP1-Knockout-Zelllinie erfolgte durch mich in enger Zusammenarbeit mit *Angie Kirchner*, genauso wie die Experimente zum alternativen Splicing. *Markus Müller* war für die Erstellung fast aller Abbildungen in diesem Abschnitt verantwortlich.

4.1.4 Material und Methoden

4.1.4.1 Zelllinien

Kultur der HeLa/HEK293T Zellen

HeLa-Zellen wurden in Dulbecco's Modified Eagle Medium (DMEM, Sigma-Aldrich), HEK293T-Zellen in RPMI-1640 (Sigma-Aldrich) mit jeweils 10 % *Fetal Bovine Serum* (Invitrogen), 4 mM L-Alanyl-L-Glutamin (200 mM, Sigma) und 1x Penicillin-Streptomycin (100x, Sigma-Aldrich) bei 37 °C in wassergesättigter und 5 % CO₂-angereicherter Atmosphäre kultiviert. Die Zellen wurden alle zwei bis drei Tage mit einem Verdünnungsverhältnis von 1:5 bzw. 1:10 passagiert. Die Zellen wurden mit 0,25 % Trypsin-EDTA für 5 min bei 37 °C geerntet, bei 300 g für 5 min pelletiert und mit DPBS-Puffer (Sigma-Aldrich) gewaschen, bevor sie in neue Kulturflaschen ausgesät wurden.

Kultur der HAP Zellen

HAP1-Zellen und CDK5RAP1-Knockout-Zellen wurden bei Haplogen (Horizon Genomics) bestellt. Die Deletion des CDK5RAP1-Gens in HAP1-Zellen wurde durch die CRISPR-Cas9 Technik erzeugt. Diese Zellen wurden in Iscove's Modified Dulbecco's Medium (IMDM, Sigma-Aldrich) mit 10 % *Fetal Bovine Serum* (Invitrogen), 4 mM L-Alanyl-L-Glutamin (200 mM, Sigma) und 1x Penicillin-Streptomycin (100x, Sigma-Aldrich) bei 37 °C in wassergesättigter und 5 % CO₂-angereicherter Atmosphäre kultiviert. Die Zellen wurden alle zwei bis drei Tage mit einem Verdünnungsverhältnis von 1:6 bzw. 1:12 passagiert.

Die adhärenenten Zellen wurden im IMDM-Wachstumsmedium bis zu 80 % Konfluenz kultiviert, dann wurde das Medium entfernt und mit DPBS (Sigma-Aldrich) gewaschen. Anschließend wurden die Zellen mit TrypLE-Express (Gibco) trypsiniert und bei 37 °C und 5 % CO₂ für 5 min inkubiert. Dieses Reagenz zur Ablösung wurde durch die Zugabe von dem Wachstumsmedium inaktiviert. Zur Langzeitlagerung in flüssigem Stickstoff wurden die Zellen in IMDM-Wachstumsmedium resuspendiert und eine gleiche Menge (v/v) Gefriermedium mit 20 % DMSO, 20 % FBS und 60 % IMDM-Wachstumsmedium über eine Minute tropfenweise zugegeben. Die Mischung wurde 15 min bei Raumtemperatur inkubiert und anschließend in Kryoröhrchen aliquotiert und bei -176°C eingefroren.

4.1.4.2 RNA Isolation

RNA Isolation aus Maus-Gewebe

Gewebe von 3 Monate alten Mäusen wurde isoliert und sofort in flüssigem Stickstoff eingefroren (LMU München, Departement Pharmazie, AK Biel). Die gefrorenen Gewebeproben wurden gewogen und mit 1 mL TriFast (Peqlab) versetzt. Die Homogenisierung erfolgte für 1 min bei 30 Hz mit einem TissueLyser (Qiagen). Das Homogenat wurde verdünnt und zusätzlich drei- bis fünfmal homogenisiert, bis das Endvolumen 1 mL TriFast (Peqlab) pro 50 mg Gewebe betrug. Nach der Homogenisierung wurde die RNA-Isolation gemäß Herstellerprotokoll durchgeführt. Das getrocknete RNA-Pellet wurde in einem geeigneten Volumen an Reinstwasser resuspendiert und die PolyA-RNA-Anreicherung wurde mit dem Oligotex mRNA Midi Kit (Qiagen) nachfolgendem Verfahren durchgeführt.

Gesamt-RNA Isolation aus Zellkultur

Präparation verschiedener RNA-Typen: Isolierung der Gesamt-RNA. Für Standardexperimente wurden Zellen in einem 175 cm² Kulturgefäß bis zu 80-90 % Konfluenz kultiviert. Die Gesamt-RNA wurde mit TriFast (Peqlab) oder TRI-Reagenz (Sigma-Aldrich) gewonnen. Die Zellen wurden direkt mit 1,1 mL TRI-Reagenz pro 10 cm² Oberfläche lysiert. Nach der Zugabe wurde das Zelllysat mehrmals durch eine Pipette auf und ab pipettiert, um ein homogenes Lysat zu erhalten und anschließend für 5 min bei Raumtemperatur inkubiert. Pro 1,1 mL TRI-Reagenz wurden 0,2 mL CH₃Cl zugegeben, das Gemisch wurde stark geschüttelt, für 15 min bei Raumtemperatur inkubiert und anschließend zentrifugiert (9500 x g, 35 min, 4 °C). Die wässrige Phase wurde abgenommen und mit Isopropanol (0,5 mL pro 1,1 mL TRI-Reagenz) versetzt. Die RNA wurden über Nacht bei 0 °C ausgefällt und mit 21130 x g für 60 min bei 4 °C zentrifugiert. Der Überstand wurde verworfen und das Pellet wurde zweimal mit 1 mL 75 %igem Ethanol gewaschen. Das getrocknete Pellet wurde in RNase-freiem Wasser resuspendiert.

PolyA-RNA Isolation aus Zellkultur

PolyA-RNA aus HeLa-Zellen wurde durch Hybridisierung an biotinylierten Phosphono Peptid Nukleinsäuren-Analoga (polyT-gripNA) und anschließender Bindung an Streptavidin-Beads mit dem mTRAP-Maxi Kit (Active Motif) gereinigt. Die PolyA-RNA aller anderen Zelllinien wurde mit dem Oligotex mRNA Midi Kit (Qiagen) nach Herstellerangaben isoliert.

rRNA Abreicherung

Aus der isolierten Gesamt-RNA wurde eine doppelte PolyA-RNA Anreicherung mit dem Oligotex mRNA Midi Kit (Qiagen) vorgenommen und anschließend eine rRNA-Abreicherung mit dem GeneRead rRNA depletion Kit (Qiagen) und Bereinigung mit dem RNeasy MinElute CleanUp Kit (Qiagen) nach Herstellerangaben durchgeführt.

Bestimmung der RNA Konzentration

Die getrocknete RNA wurde in einem angemessenen Volumen an Reinstwasser gelöst und vereint. Die Konzentration wurde anschließend im linearen Bereich des NanoDrop UV/Vis Spectrophotometers (Peqlab) bestimmt.

4.1.4.3 Agarose Gel-Elektrophorese

Um die verschiedenen RNA Präparationen hinsichtlich ihrer Reinheit und Konzentration zu prüfen und miteinander zu vergleichen, wurden Proben davon auf einem 1,5 % TBE-Agarose-Gel aufgetrennt. Hierzu wurden 0,75 g Agarose in 50 mL 1 x TBE Puffer aufgekocht, bis die Agarose vollständig gelöst war und die Lösung zum Abkühlen stehen gelassen. Anschließend wurden 7 µL des verwendeten Farbstoffs (Ethidiumbromid (1 mg/mL)/PeqGreen) zugefügt und die Lösung zum Aushärten für ca. 30 min in eine horizontale Gelkammer gegossen. RNA Proben der zu untersuchenden Präparationen wurden mit 2 x RNA Auftragspuffer versetzt, für 2 min auf 90 °C erhitzt und nach weiteren 2 min auf Eis in die Geltaschen geladen. Als Marker wurde *ssRNA ladder* der Firma NEB verwendet. Dazu wurden 2 µL des Markers mit 8 µL 2 x RNA Auftragspuffer versehen und entsprechend den anderen Proben vor dem Beladen auf das Gel erhitzt und abgekühlt. Das Gel wurde für 50 min bei 110 V in 0,5 x TBE Puffer laufen gelassen.

Zur Analyse der cDNA wurde ein 1,5 TAE-Agarose-Gel verwendet. Das Gel wurde entsprechend den obigen Angaben zu TBE-Agarose-Gelen mit 0,75 g Agarose in 50 mL 1 x TAE und 7 µL Ethidiumbromid hergestellt. Als Marker diente der *PCR ladder* der Firma NEB von dem 7,5 µL auf das Gel geladen wurden. Die nach der PCR enthaltenen Proben (20 µL) wurden mit je 5 µL DNA Auftragspuffer gemischt und ebenso auf das Gel geladen. Anschließend wurde das Gel bei einer Spannung von 90 V für 55 min mit 0,5 x TAE Puffer laufen gelassen.

Die Banden wurden mit Hilfe des IDA Imager der Firma Raytest unter UV-Licht bei einer Anregungswellenlänge von 260 nm oder mit Hilfe der E-Box VX5 der Firma Vilber Lourmat unter Blaulicht visualisiert.

4.1.4.4 Enzymatischer Verdau

Es wurden 1-2 µg der genomischen RNA in 35 µL ddH₂O wie folgt verdaut: Eine wässrige Lösung (7,5 µL) aus 480 µM ZnSO₄, die 42 U Nuklease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic Phosphatase (New England BioLabs) und spezifische Mengen an isotoopenmarkierten internen Standards enthält, wurde für 3 h bei 37 °C inkubiert. Nach der Zugabe einer [Na]₂-EDTA Lösung (520 µM, 7,5 µL), die 0,2 U Snake Venom Phosphodiesterase (*Crotalus adamanteus*, USB corporation) enthält, wurde die Probe erneut für 3 h bei 37 °C inkubiert. Anschließend wurde die Probe bei -20 °C gelagert. Bevor sie durch die UHPLC-MS/MS Methode analysiert wurde, wurde sie durch eine AcroPrep Advance 96 Filterplatte (0.2 µm Supor, Pall Life Sciences) filtriert.

4.1.4.5 UHPLC-MS/MS Analyse

Die quantitative UHPLC-MS/MS-Analyse von verdauten RNA-Proben wurde mit einem Agilent 1290 UHPLC-System mit UV-Detektor und einem Agilent 6490 Triple Quadrupol-Massenspektrometer in Verbindung mit der stabilen Isotopenverdünnungstechnik durchgeführt. Eine verbesserte Methode, basierend auf den bereits veröffentlichten Arbeiten^[65, 82, 171] wurde entwickelt, um gleichzeitig alle Nukleoside in einem einzigen Analyselauf zu quantifizieren. Die quellenabhängigen Parameter waren: Gastemperatur 80 °C, Gasfluss 15 L/min (N₂), Nebulizer 30 psi, *Sheath Gas* Temperatur 275 °C, *Sheath Gas* Fluss 11 L/min (N₂), Kapillarspannung 2.500 V im positiven Ionenmodus und *Nozzle* Spannung 500 V. Die Fragmentorspannung betrug 380 V. Delta EMV wurde auf 500 (positiver Modus) eingestellt. Substanz-spezifische Parameter sind in Tabelle 3 zusammengefasst. Die Chromatographie wurde mit einer Poroshell 120 SB-C8-Säule (Agilent, 2,7 µm, 2,1 mm x 150 mm) bei 30 °C unter Verwendung eines Gradienten aus Wasser und MeCN, die jeweils 0,0085 % (v/v) Ameisensäure enthalten, bei einer Flussrate von 0,35 mL/min durchgeführt: 0→6 min; 0→6% (v/v) MeCN; 6→7 min; 6→40 % (v/v) MeCN; 7→10,8 min; 40→80 % MeCN; 10,8→20 min; 80 % MeCN; 20→ 20,8 min, 80 %→0 % (v/v) MeCN, 20,8→22 min; 0 % MeCN. Der Eluent wurde bis zu 1,1 min und nach 14 min durch ein Valco-Ventil in den Abfall umgeleitet. Der Autosampler wurde auf 4 °C gekühlt. Das Injektionsvolumen betrug 39 µL.

Tabelle 3: Substanz-abhängige LC-MS/MS Parameter für die Analyse der RNA. CE: *Collision Energy*, CAV: *Collision Cell Accelerator Voltage*, EMV: *Electron Multiplier Voltage*. Die Nukleoside werden im positiven Modus gemessen. ($[M+H]^+$ Spezies im *ion selected reaction monitoring mode* (SRM))

Verbindung	Precursor Ion (m/z)	MS1 Auflösung	Product Ion (m/z)	MS2 Auflösung	Dwell time [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 2,4-14 min								
[D ₃]-ms ² i ⁶ A	385,17	Wide	253,13	Wide	60	25	5	Positiv
ms ² i ⁶ A	382,15	Wide	250,11	Wide	60	25	5	Positiv
[D ₂]-i ⁶ A	338,18	Wide	206,14	Wide	60	15	5	Positiv
i ⁶ A	336,20	Wide	204,12	Wide	60	15	5	Positiv

Bestimmung des Extinktionskoeffizienten von ms²i⁶A

Für die nachfolgende absolute interne Quantifizierung von Nukleosiden muss die genaue Konzentration der isotoopenmarkierten internen Standards bekannt sein. Um die Konzentrationen zu messen, müssen zuerst die Extinktionskoeffizienten bestimmt werden. Für diese Bestimmung wurde ms²i⁶A, sowie das Referenz-Nukleosid U, in D₂O/D₆-DMSO (20:80) gelöst. Die Absorption der beiden Lösungen wurde als technisches Quintett auf dem NanoDrop 1000 Spektrophotometer für die λ_{\max} jedes Nukleosids gemessen. Anschließend wurde eine 1:1-Mischung als technisches Triplikat aus ms²i⁶A und U hergestellt. Nach ¹H-NMR-Messungen dieser Gemische ergaben sich spezifische Verhältnisse der Stoffmenge n (ms²i⁶A) zu der Stoffmenge n (U). ¹H-NMR-Spektren wurden auf Bruker Avance III HD 800 MHz aufgenommen. Durch Anwendung des Lambert-Beerschen Gesetzes kann man die tatsächliche Konzentration c (U) der Lösung mit dem in der Literatur bekannten Extinktionskoeffizienten berechnen (ϵ_{λ}). Mit dem Verhältnis von n (U) zu n (ms²i⁶A) kann man die Konzentrationen c (ms²i⁶A) in jeder Lösung berechnen. Bei erneuter Anwendung des Lambert-Beerschen Gesetzes ergibt sich der gewünschte Extinktionskoeffizient des Nukleosids ϵ_{λ} (ms²i⁶A) bei einer bestimmten Wellenlänge λ (siehe Tabelle 4).

Tabelle 4: Die Extinktionskoeffizienten für ms²i⁶A und i⁶A.

Nukleosid	$\epsilon_{\lambda_{\max}}$ [L·mol ⁻¹ ·cm ⁻¹]	λ_{\max} [nm]
ms ² i ⁶ A	14917	283
i ⁶ A	19559	268

Quantifizierung der Nukleoside

Um die internen Kalibriergeraden für die exakte Quantifizierung zu erhalten, wurde jeder Standard, d.h. $[D_3]$ -ms²i⁶A und $[D_2]$ -i⁶A, in verschiedenen Konzentrationen im Vergleich zum entsprechenden nicht markierten Nukleosid mit konstanter Konzentration analysiert. Daraus ergeben sich verschiedene Verhältnisse von der Stoffmenge des natürlichen Nukleosids zur Stoffmengen n^* des markierten Nukleosids. Diese Proben wurden als technische Triplikate gemessen und die lineare Regression wurde mit Origin® 6.0 (Microcal™) bestimmt. Dafür wurde das Verhältnis der Fläche unter der Kurve des unmarkierten Nukleosids (A) zum markierten Standard (A^*) gegen das Verhältnis der Stoffmenge des unmarkierten Nukleosids (n) zum markierten (n^*) aufgetragen. Für die UV-Kalibriergerade zur Quantifizierung der kanonischen Nukleoside, d.h. C, U, G und A wurden diese in einer Konzentrationsreihe gemessen. Für jedes Nukleosid wurden zwei verschiedene Lösungen mit einem Verdünnungsfaktor von zwei verdünnt, bis für jedes Nukleosid fünf verschiedene Konzentrationen erhalten wurden, die anschließend als technische Triplikate gemessen wurden. Die UV-Fläche unter der Kurve (A) wurde gegen die Stoffmenge des Nukleosids (n) aufgetragen. Akzeptable Genauigkeit (< 20 % relative Standardabweichung innerhalb des Triplikats) und Genauigkeit (80 – 120 %) wurde für die Kalibriergerade erreicht. Die Genauigkeit wird durch das Verhältnis der gemessenen zu den berechneten Verhältnissen der Flächen (A/A^*) unter den Kurven in Prozent berechnet. Die Verhältnisse der Flächen (A/A^*) können mit Hilfe der linearen Gleichungen für das entsprechende Stoffmengenverhältnis (n/n^*) berechnet werden. Die untere Nachweisgrenze wurde als die detektierte Menge definiert, die dreimal höher ist als der *Blank* (LOD). Die untere Nachweisgrenze zur Quantifizierung (LLOQ) und die obere Nachweisgrenze zur Quantifizierung (ULOQ) sind die niedrigste bzw. höchste Stoffmenge (n) und das Verhältnis der Flächen (A/A^*), die die Anforderungen der entsprechenden linearen Gleichung erfüllen.

4.1.4.6 Quantitative Echt-Zeit PCR Analyse

Bestimmung der prä-mRNA Anreicherung durch die Verwendung des Splicing Inhibitors Isoginkgetin

Die Zellen wurden unter Standardbedingungen, wie bereits oben beschrieben, in einer 175 cm² Zellkulturflasche bis zu 80 % Konfluenz kultiviert. Vor dem Experiment wurde das Kulturmedium gegen ein IMDM-Wachstumsmedium mit 50 µM Isoginkgetin (Merck Millipore) ausgetauscht und die Zellen bei 37 °C und 5 % CO₂ für 6 h inkubiert. Als Kontrolle wurde mit der gleichen Menge an DMSO im Medium kultiviert. Die Isolation der Gesamt-RNA wurde wie oben beschrieben durchgeführt. Mit dem TURBO DNA-free Kit (Ambion, Life Technologies) wurden übrige Rückstände genomischer DNA durch die Behandlung mit DNaseI entfernt. 1 µg RNA wurde mit dem iScript cDNA Synthesis Kit (Bio-Rad) in 1 µg cDNA transkribiert, was zu einer endgültigen cDNA-Konzentration von 50 ng/µL führte. Für die RT-PCR wurden 2 µL cDNA mit 10 µL iTaq Universal SYBR Green Supermix (Bio-Rad) gemischt. Für jedes Primerpaar wurden *forward*- und *reverse* Primer in gleichen Mengen zusammengegeben und mit nukleasefreiem ddH₂O auf eine Endkonzentration von 1,25 µM verdünnt. Die verwendeten Primer sind unten aufgelistet. Pro *Well* wurden 12 µL des cDNA-iTaq-Mastermix und 8 µL des Primer-Mix zugegeben. Die 96-well PCR-Platten (Axygen Scientific) wurden in Kombination mit Flat 8 Cap Strips verwendet und die Proben mit einer Eppendorf Zentrifuge 5810 R bei 1500 g für 3 min zentrifugiert. Jede Probe wurde als Triplikat mit einem Eppendorf Mastercycler Realplex4 gemessen. Die PCR-Bedingungen waren wie folgt: Schritt 1: 95 °C (3:00 min), Schritt 2: 95 °C (0:15 min), Schritt 3: 60 °C (0:15 min), Schritt 4: 72 °C (0:20 min); Schritt 2 - 4 wurde 40-mal wiederholt.

Das Verhältnis (R) von prä-mRNA zu mRNA wurde wie folgt berechnet:

$$R_{\text{Isoginkgetin}} = 2^{[C_t \text{ prä-mRNA} - C_t \text{ mRNA}]}$$

und im Anschluss auf die DMSO-behandelten Zellen normiert:

$$R_n = R_{\text{Isoginkgetin}}/R_{\text{DMSO}}$$

Die dafür verwendeten Primer sind unten abgebildet.

hTDG (exon2) *forward* Primer: 5'-TGAAGCTCCTAATATGGCAGTTG-3'

hTDG (exon3) *reverse* Primer: 5'-GGTTCTGTTGTTCTGGGTTTTCT-3'

hTDG (intron2) *reverse* Primer: 5'-AAGCTGGCTGATCCGATGTT-3'

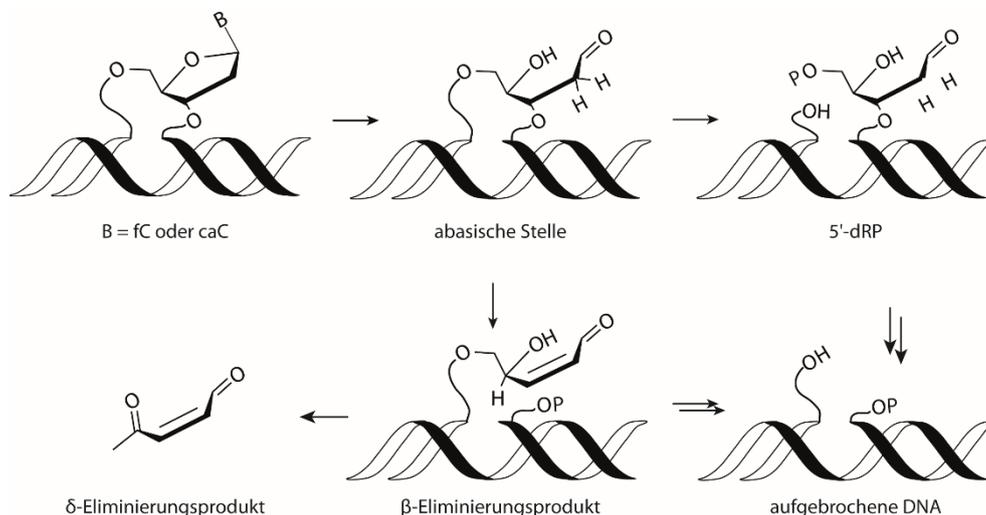
4.1.4.7 Visualisierung der Mitochondrien durch den Mito Tracker® Deep Red

Die MitoTracker® Deep Red FM (M22426, Invitrogen) Stammlösung (1 mM) wurde nach Herstellerangaben hergestellt. HAP1 und CDK5RAP1-Knockout-Zellen wurden in μ -Slides (8 Well, Ibidi) bis zu 60 % Konfluenz kultiviert. Das Kulturmedium wurde entfernt, vorgewärmtes Medium mit 100 nM MitoTracker® zugegeben und für 30 min bei 37 °C inkubiert. Nach der Färbung wurden die Zellen gewaschen und in DPBS (Sigma-Aldrich) abgebildet. Die erfolgreiche Färbung wurde mit einem konfokalen Mikroskop im Fluoreszenzmodus (Leica DMI4000B) visualisiert.

4.2 Studie zur aktiven Demethylierung mit verschiedenen dC- und C-Derivaten

4.2.1 Prolog

mdC hat eine wichtige Rolle bei der Genexpression, indem es verhindert, dass das jeweilige Gen transkribiert wird. Damit diese Modifikation aus der DNA wieder entfernt werden kann und die epigenetische Kontrolle dynamisch bleibt, sind Mechanismen zur Entfernung der Methylierung notwendig. Eine Möglichkeit hierzu ist das Ausschneiden der jeweiligen modifizierten Base aus dem Nukleotid. Dies führt zu einer abasischen Stelle, nach deren Entfernung durch die AP Endonuklease ein Nukleotid mit einer unmodifizierten Base eingefügt werden kann. Dieser Reparaturmechanismus wird als Basenexzisionsreparatur (BER) bezeichnet.^[33, 172-174] Die BER ist ein Mittel zur DNA Reparatur die man in allen Domänen des Lebens findet.^[175] Die zu entfernende Base wird an der betroffenen Stelle von Glykosylasen erkannt und ausgeschnitten. Während Pflanzen Glykosylasen besitzen, die mdC direkt entfernen können,^[176, 177] werden in Säugetieren wahrscheinlich nur die oxidierten Formen fdC und cadC erkannt. Dies wird von der Thymin-DNA-Glykosylase (Tdg) bewerkstelligt,^[30, 178] deren Funktion unter anderem die Entfernung von G:T-Fehlpaarungen ist.^[179] Der exakte Mechanismus der Entfernung von fdC und cadC ist jedoch Gegenstand aktueller Forschung.^[172]



Schema 2: Mechanismus der Basenexzisionsreparatur ausgehend von einem fdC oder cadC. Aus der entstandenen abasischen Stelle kann über zwei verschiedene Wege ein Einzelstrangbruch in der DNA gebildet werden. Entweder über die Entstehung eines 5'-2-Deoxyribose-5-phosphat (5'-dRP) oder über das β-Eliminierungsprodukt. Dieses kann wiederum zu einem δ-Eliminierungsprodukt weiterreagieren. (Schema nach *Rahimoff et. al.*)^[180]

Während der Reparatur des Genoms durch BER (siehe Schema 2) wird zunächst die modifizierte Base entfernt und schließlich ein unmodifiziertes Cytosin neu eingebaut. Es gibt zwei Arten von Reparatur-Glykosylasen: Zum einen die monofunktionalen Glykosylasen, die zuerst abasische Stellen produzieren, die im nächsten Schritt durch Endonukleasen wie der Ape1 entfernt werden.^[181] Dadurch entsteht ein 5'-2-Deoxyribose-5-phosphat (5'-dRP). Zum anderen gibt es bifunktionale Glykosylasen, diese prozessieren die abasischen Stellen intermediär zu β -Eliminierungsprodukten, welche durch eine anschließende δ -Eliminierung zu einem Einzelstrangbruch der DNA und somit zu einer Einzelnukleotid-Lücke führen.^[180]

Eine weitere Möglichkeit, die Methylierung von Cytidin aktiv zu entfernen, ist der direkte C-C Bindungsbruch an der C5-Position, der ohne das Ausschneiden der gesamten Base auskommen würde. Das wäre ein sehr attraktiver Mechanismus, da dieser einzelne Schritt weniger Energieaufwand erfordern würde als die BER und vermutlich weniger anfällig für Mutationen wäre, da es auch nicht zur Bildung von potentiell gefährlichen Strangbrüchen kommen kann.^[182] Außerdem könnte die Demethylierung über diesen Mechanismus deutlich schneller vonstattengehen als über die BER, was angesichts der rapiden globalen Demethylierung der DNA im väterlichen Pronukleus in Säugetierzygoten beachtenswert ist.^[183] Des Weiteren wurde bereits gezeigt, dass sowohl das mütterliche als auch das väterliche Genom der murinen Zygoten eine Tdg-unabhängige aktive Demethylierung durchlaufen.^[184] Erste Hinweise auf einen C-C Bindungsbruch konnten im Rahmen dieser Arbeit erhalten werden (Kapitel 3.4). Zur weiteren Untersuchung dieser Fragestellung wurden natürliche isotopologe Derivate gefüttert. Diese Derivate werden auch durch den alternativen Demethylierungsweg, die BER, aus dem Genom entfernt. Des Weiteren wurde kein fdC- sondern ein dC-Derivat eingesetzt, was sowohl Vorteile, als auch Nachteile mit sich bringt. Mit isotoopenmarkierten dC-Derivaten ist der komplette Metabolismus eines markierten Desoxycytidins nachverfolgbar. Man benötigt hierfür allerdings hoch-isotoopenmarkierte Derivate, die sich massenspektrometrisch stark von den natürlichen Nukleosiden unterscheiden müssen. Im Experiment wurden verschiedene murine embryonale Stammzelllinien, denen die Expression wichtiger Enzyme zur (De-)Methylierung von Cytidin nicht möglich ist, mit verschiedenen isotoopenmarkierten dC-Spezies gefüttert (siehe Abb. 20). All diese Metaboliten haben die Deuterium Markierung an der C5-Position der Base gemeinsam. Anschließend wurde das jeweilige Genom mithilfe der Massenspektrometrie quantitativ bezüglich der Mengen der verschiedenen Nukleoside analysiert. Durch den Vergleich der Wildtyp-Daten mit den Daten aus der jeweiligen Knockout-Zelllinie sollten Rückschlüsse auf den Abbaumechanismus von mdC gezogen werden. Der Verlust der D-Markierung an der C5-Position zeigt hierbei den potentiellen C-C Bindungsbruch an.

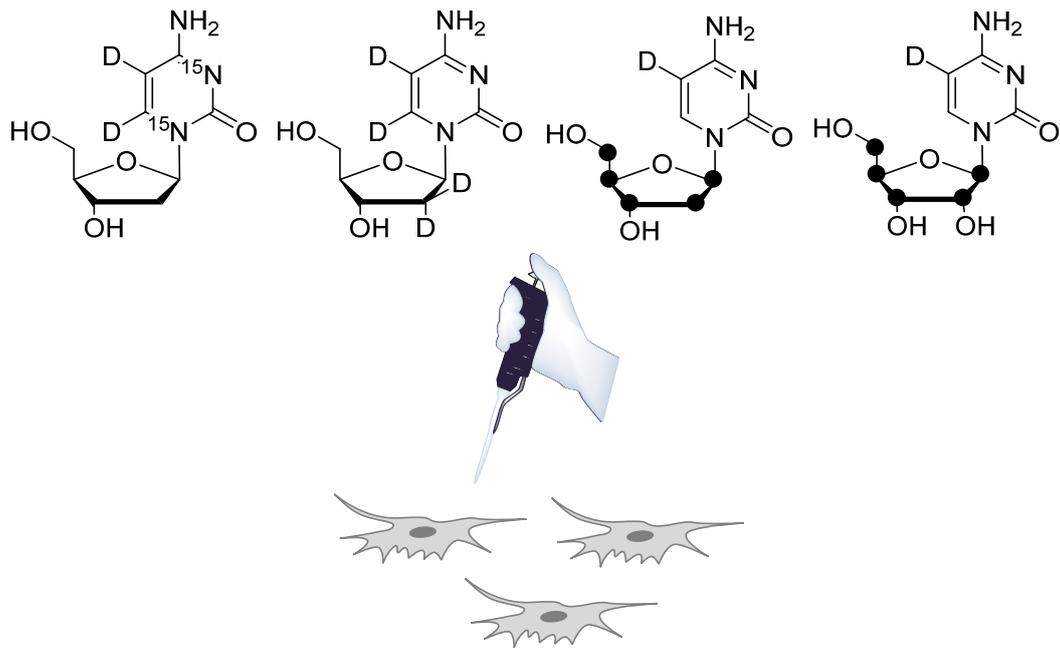
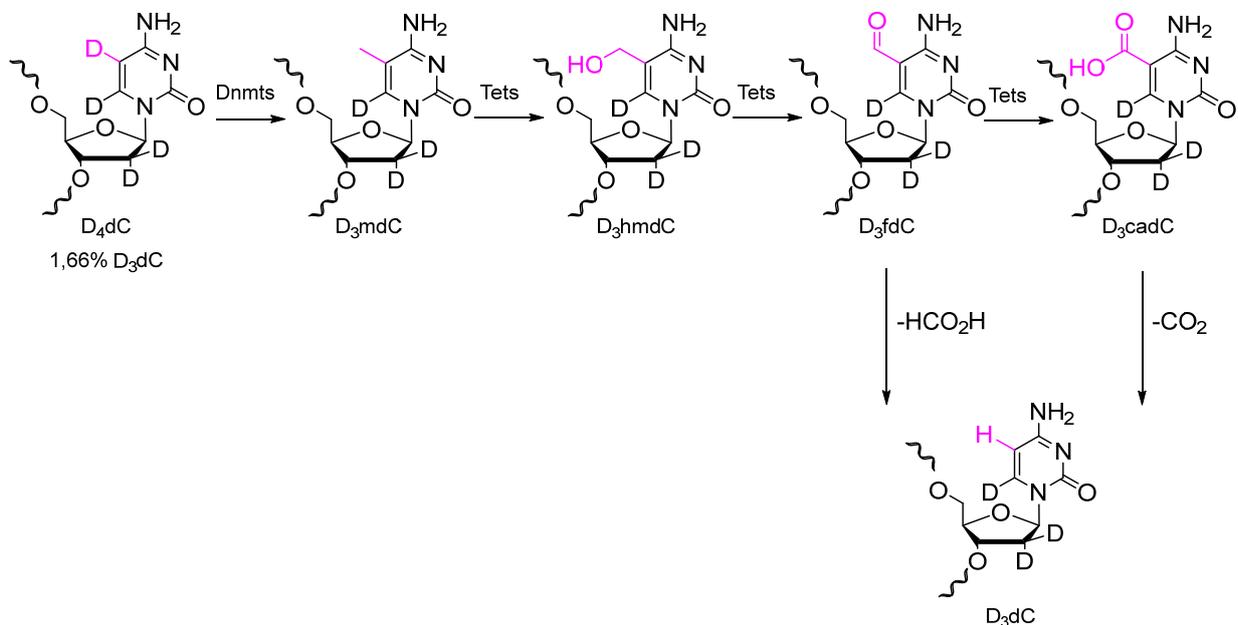


Abbildung 20: Die chemischen Strukturen der vier verschiedenen Metaboliten, die nacheinander entwickelt und an verschiedene Zellen gefüttert wurden. Schwarz ausgefüllte Kreise kennzeichnen die ^{13}C -Markierungen.

Diese Arbeit ist an dieser Stelle eine Fortsetzung der Experimente aus der Dissertation von *Jessica Steinbacher* und *Arne Schröder*. Dort wurde $D_2-^{15}N_2-dC$ an verschiedene Zelltypen gefüttert. Dieses Molekül hatte aber einige Nachteile. Zum einen konnte aufgrund der kommerziell erhältlichen Ausgangsmaterialien keine Isotopenreinheit von über 97 % erzielt werden. Zum anderen war die Inkorporationsrate dieser Verbindung in die verschiedenen Zelltypen oft zu gering, um den kompletten Metabolismus nachverfolgen zu können. Die folgende Arbeit zeigt die schrittweise Weiterentwicklung dieser Methode. Zuerst wurde D_4dC als markierte Substanz verwendet. Anschließend wurde $^{13}C_5D_1dC$ eingesetzt und schließlich wurde mit dem korrespondierenden Ribonukleotid $^{13}C_5D_1C$ experimentiert (Abb. 20).

4.2.2 Ergebnis und Diskussion

Die Weiterentwicklung des $D_2\text{-}^{15}\text{N}_2\text{-dCs}$ war notwendig, da eine isotopenreine $^{15}\text{N}_2$ -Quelle ($^{15}\text{N}_2$ -Harnstoff: $\sim 98\%$) nicht zugänglich war. Auch nach einer Nachdeuterierung durch *René Rahimoff* konnte die Isotopenreinheit nicht über 97% gesteigert werden. Ein neues isotopenmarkiertes dC-Derivat wurde daher entwickelt. Die Anforderungen waren, dass es einen Massenunterschied von mindestens 4 u zum natürlichen dC aufweisen sollte und sowohl eine Isotopenmarkierung an der C5-Position der Base sowie im Zucker haben sollte. Der Verlust der D-Markierung an der C5-Position sollte die aktive Demethylierung sichtbar machen. Durch die zusätzlichen Markierungen im Zucker sollte dieser Metabolit besser vom natürlichen dC in der DNA zu unterscheiden sein. Aufgrund der nicht ausreichenden Isotopenreinheit von $^{15}\text{N}_2$ -markiertem Stickstoff wurde ein Metabolit mit vier Deuterium Atomen, $D_4\text{dC}$, synthetisiert (Schema 3).

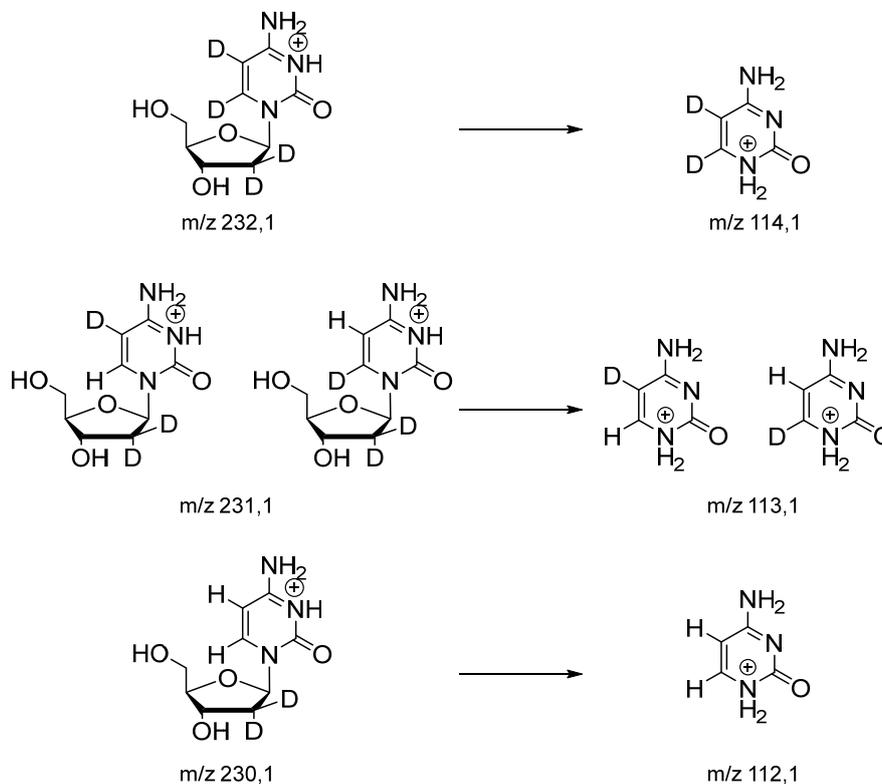


Schema 3: Der mögliche Weg der aktiven Demethylierung über einen C-C Bindungsbruch kann entweder über die Deformylierung ($-\text{HCO}_2\text{H}$) oder über die Decarboxylierung ($-\text{CO}_2$) verlaufen. Durch die D-Markierung können die Nucleoside nach dem Totalverdau von den natürlichen Nucleosiden unterschieden werden.

Für die massenspektrometrische Analyse dieser Verbindung und den potentiell daraus entstandenen Nucleosiden musste die bereits vorhandene UHPLC-MS/MS Messmethode weiterentwickelt werden. Zur Quantifizierung des Metaboliten und den daraus entstandenen Nucleosiden benötigt man zusätzlich noch stärker isotopenmarkierte Standards, um die Isotopenverdünnungsmethode anwenden zu können. Auch diese Standards dürfen nicht mit den Metaboliten interferieren. Die Isotopenreinheit dieses neu synthetisierten Metaboliten lag bei $\sim 98,34\%$. Damit ist die Substanz $1,34\%$ isotopener als die Vorläuferverbindung $D_2\text{-}^{15}\text{N}_2\text{-dC}$. Wie in Schema 3 gezeigt, kann $D_4\text{dC}$ in die DNA inkorporiert und durch die Dnmt Enzyme zu $D_3\text{mdC}$ methyliert werden. Die Tet Enzyme können ferner $D_3\text{mdC}$ zu $D_3\text{hmdC}$,

D₃fdC und wahrscheinlich auch zu D₃cadC oxidieren. Letzteres konnte allerdings nicht detektiert werden, da in allen Experimenten nur 1-2 µg DNA verdaut wurden, um so den Hintergrund möglichst gering zu halten. Dies sollte eine Verbesserung der Sensitivität mit sich bringen. Auch die D₃fdC Werte waren unter dem *Lower Limit of Quantification* (LLOQ). Würde eine C-C Bindungsbruchreaktion an der C5-Position stattfinden, entweder über eine Deformylierung von D₃fdC oder eine Decarboxylierung von D₃cadC, so würde als Produkt D₃dC entstehen. Das Schema 3 zeigt diesen möglichen aktiven Demethylierungsweg über die beiden genannten C-C Bindungsbruchreaktionen.

Da die Isotopenreinheit nur bei 98,34 % lag, ergeben sich Probleme aus den verbleibenden 1,66 % an D₃dC. Wie im Schema 4 gezeigt, fragmentieren alle dC-Derivate im positiven ESI-Modus und die einzige unter diesen Bedingungen mögliche Fragmentierung ist die Spaltung der glykosidischen Bindung.



Schema 4: Die spezifischen Fragmentierungen des D₄dC Metaboliten, der beiden möglichen D₃dC Spezies und der potentiell möglichen D₂dC Spezies. Alle dC-Derivate Fragmentieren im positiven ESI-Modus und es kommt zu einer Spaltung der glykosidischen Bindung.

Daraus wird ersichtlich, dass man nicht zwischen den beiden möglichen D₃dC Spezies unterscheiden kann, was für eine korrekte Analyse problematisch ist. Die 1,66 % Verunreinigung an D₃dC können darüber hinaus von zwei verschiedenen Molekülen herrühren: Fehlt das Deuterium an der C5-Position, so ist das genau das Molekül, welches nach einer Deformylierung bzw. Decarboxylierung entstehen würde. Fehlt das Deuterium

allerdings an der C6-Position, so würde nach der aktiven Demethylierung eine unmarkierte Base, d.h. ein D₂dC vorliegen. Auch dieses Molekül muss in der Messmethode enthalten sein. Theoretisch sollte auch die Messung dieses Metaboliten, D₂dC, kein Problem darstellen. Allerdings sollte die Wahrscheinlichkeit eines natürlich vorkommenden dC+2 Moleküls nicht außer Acht gelassen werden. Dabei handelt es sich nicht um ein natürlich vorkommendes zweifach deuteriertes dC, sondern um ein dC, das zweimal ¹³C statt ¹²C enthält. Die Wahrscheinlichkeit, dass ein dC, welches aus 9 C-Atomen besteht, ein ¹³C enthält, liegt bei 9,9 %. Die Wahrscheinlichkeit für ein ¹³C₂-dC liegt bei knapp 1 % und ist deshalb nicht zu vernachlässigen. Genauso wie das natürliche Vorkommen von ¹⁸O (0,20 %) neben ¹⁶O nicht außer Acht gelassen werden darf.

In einem ersten Experiment wurden 100 µM des D₄dCs für drei Tage an Tdg^{+/-} und Tdg^{-/-} murine embryonale Stammzellen (mESC) gefüttert. Diese befanden sich entweder im *naïve* Zustand oder waren bereits unter *priming* Bedingungen kultiviert. Bei der Linie Tdg^{+/-} handelt es sich um eine heterozygote Zelllinie, in deren doppeltem Chromosomensatz das für die Tdg kodierende Gen auf nur einem Allel vorhanden ist. Dies reicht aus, um das Enzym herzustellen, weshalb Tdg^{+/-} in der Lage ist, mdC über den BER-Mechanismus zu demethylieren. Die Linie Tdg^{-/-} exprimiert dagegen kein Tdg, weshalb diesen Zellen der Weg zur Demethylierung über BER nicht möglich ist. Damit bliebe den Tdg^{-/-} Zellen zum Abbau von mdC vermutlich nur die potentielle Demethylierung über einen C-C Bindungsbruch. Da eine Fütterungskonzentration von 100 µM nur zu einem Markierungsgrad von 2 – 6 % D₄dC pro dC führten, wurde das Experiment mit einer Konzentration von 200 µM wiederholt. Diesmal war der Markierungsgrad etwas höher (8 – 17 %) (Abb. 21 A, C), wobei die Tdg profizienten Zellen im *naïve* Zustand fast doppelt so viel D₄dC inkorporierten als die, die bereits *geprimed* waren. Interessant war hierbei, dass es in den Tdg defizienten Zellen zu einer leichten Anreicherung der deuterierten Spezies im Vergleich zu den heterozygoten Zellen kommt. Im prozentualen Vergleich aller deuterierten Spezies sieht es so aus, als ob in den Tdg^{-/-} Zellen jeweils mehr D₃dC vorhanden ist als in den Tdg^{+/-} Zellen. Betrachtet man allerdings die absolute Quantifizierung von D₃dC bezogen auf alle Nukleoside (dN), so stimmt dies nur für die *naïve* Bedingungen. Es ist also schwierig, hier von einer vermehrten Bildung an D₃dC in der Abwesenheit von Tdg zu sprechen. Auffällig ist jedoch, dass bei allen Messungen der prozentuale Anteil an D₃dC auf über 1,66 % gestiegen ist (Abb. 21 C, D), was der anfänglichen Verunreinigung durch D₃dC entsprach. Diese Erhöhung um ca. 1 % auf 2,39 – 2,60 % könnte auf eine aktive Demethylierung schließen lassen. Allerdings darf man hier nicht außer Acht lassen, dass diese Anreicherung von D₃dC auch durch einen von Enzymen verursachten D/H-Austausch, wie zum Beispiel durch Dnmts, an der C5-Position passieren könnte. Dnmts führen im Normalfall eine Methylgruppe an der C5-Position ein. Sind die Enzyme allerdings nicht mit

dem Cofaktor SAM beladen, kann keine Methylgruppe übertragen werden und stattdessen der Austausch mit Wasser stattfinden. Dies würde zu einem D/H-Austausch führen.

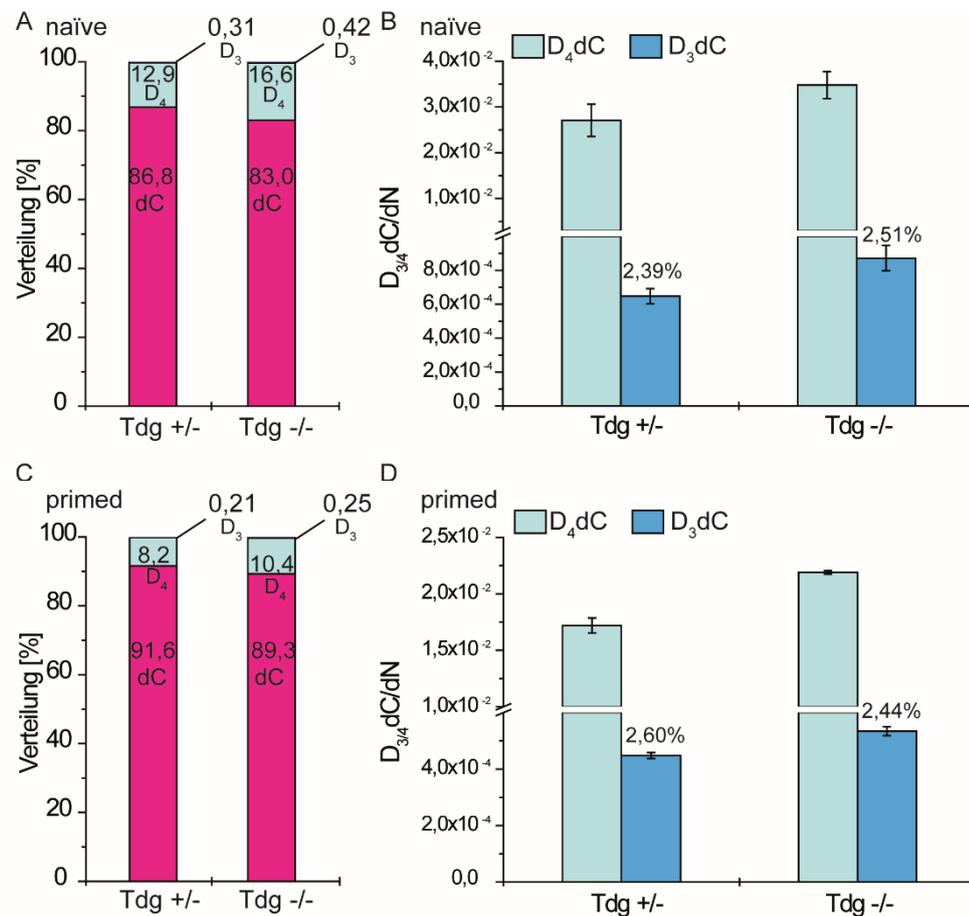


Abbildung 21: A) und C): Prozentuale Darstellung aller gemessenen dC-Spezies in Tdg^{+/-} und Tdg^{-/-} Zellen, welche entweder im *naïve* A) oder im *primed* B) Zustand kultiviert wurden. B) und D): Absolute Quantifizierung von D₄dC und D₃dC jeweils in Tdg^{+/-} und Tdg^{-/-} Zellen, welche entweder im *naïve* B) oder im *primed* D) Zustand kultiviert wurden. Zusätzlich wurde jeweils die prozentuale Menge an D₃dC zu D₄dC berechnet.

Wie es für die Tdg^{-/-} Zellen üblich ist, enthalten sie sowohl 5 - 8 Mal mehr fdC, wie auch 7 - 15 Mal mehr D₃fdC als die Tdg^{+/-} Zellen. Interessant ist hierbei, dass die Tdg^{-/-} Zellen mindestens 10 % mehr D₃fdC pro natürlichem fdC enthalten (Abb. 22). Das ist die gleiche Tendenz, die bereits von *Jessica Steinbacher* in ihrer Dissertation beschrieben wurde. Ob es sich dabei um einen Isotopeneffekt handelt, der dafür sorgt, dass das markierte fdC bevorzugt von den Tdg Enzymen ausgeschnitten wird, oder aber einen Hinweis darauf ist, dass sich die so generierte D₃fdC Spezies an spezifischen, für Tdg unzugänglichen Positionen im Genom befindet, ist noch unklar. Die Existenz von stabilem fdC enthaltenden Positionen im Genom wurde bereits beschrieben.^[185] Natürlich darf man an dieser Stelle nicht vergessen, dass es zur Detektion der D₃dC Spezies einer Deformylierung von D₃fdC bzw. einer Decarboxylierung von D₃cadC

bedarf. Die D₃fdC Spezies ist in den Tdg^{-/-} Zellen angereichert und sollte daher auch eine größere Menge an D₃dC liefern, was in guter Übereinstimmung mit den gemessenen Werten der *naïven* Stammzellen ist.

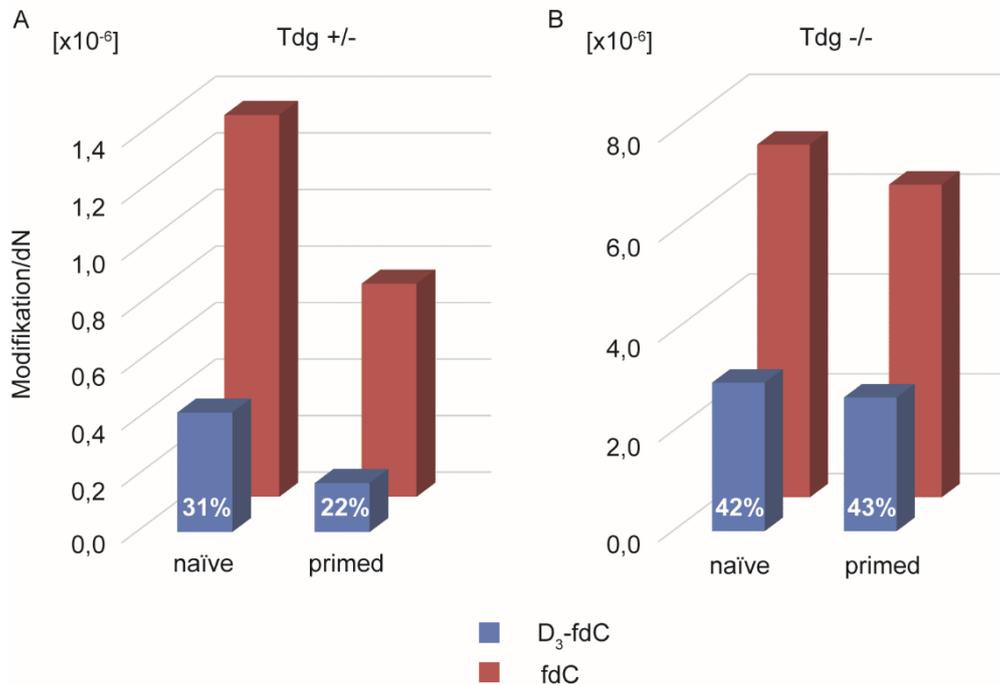


Abbildung 22: A): Absolute Quantifizierung der fdC und D₃fdC Mengen in den Tdg^{+/-} Zellen die im *naïve* oder *primed* Zustand kultiviert wurden. Die angegebenen Prozentzahlen sind die prozentualen Mengen der D₃fdC Spezies an allen fdCs Spezies. B): Absolute Quantifizierung der fdC und D₃fdC Mengen in den Tdg^{-/-} Zellen die im *naïve* oder *primed* Zustand kultiviert wurden. Die angegebenen Prozentzahlen sind die prozentualen Mengen der D₃fdC Spezies an allen fdCs Spezies.

Diese ersten potentiellen Hinweise auf eine aktive Demethylierung, die nur bedingt unabhängig von den Tdg Enzymen zu sein scheinen, sind noch kein Beweis für ein tatsächliches Auftreten eines C-C Bindungsbruchs. In einem nächsten Experiment wurde deshalb 100 µM D₄dC an Dnmt Triple Knockout (TKO) Zellen und den korrespondierenden Wildtyp (wt J1) verfüttert. Die *naïve* und *primed* mESCs wurden über drei Tage mit D₄dC enthaltendem Medium kultiviert. Die Wildtyp-Zellen sind im Gegensatz zu den Dnmt TKO Zellen in der Lage, die D₄dC Spezies zu methylieren. Die gebildete D₃mdC Spezies wiederum kann durch die Tet Enzyme oxidiert werden. In den Dnmt TKO Zellen ist schon der erste Schritt, die Methylierung des D₄dCs, nicht möglich, weswegen auch die oxidierte Spezies in dieser Zelllinie nicht auftreten. Der Vergleich der D₃dC-Werte pro inkorporiertem D₄dC beider Zelllinien sollte somit Rückschlüsse über eine aktive Demethylierung liefern, da dies nur im Wildtypen erfolgen kann.

In der folgenden Abbildung 23 A sind die relativen Mengen aller gemessenen dC-Spezies eines biologischen Duplikats abgebildet. Die Inkorporation des D₄dCs ist in den Dnmt TKO

Zellen ca. 2 – 4 % höher als in den korrespondierenden Wildtyp Zellen. Im Vergleich zu den Tdg^{+/-} und Tdg^{-/-} Zellen ist hier eine wesentlich höhere Inkorporation (25 – 30 %) bei nur 100 µM D₄dC Fütterungskonzentration möglich. Die Anteile des D₃dCs zeigen einen ähnlichen Verlauf wie die des D₄dCs, jedoch ist in den TKO Zellen ein kleiner Anstieg zu detektieren. Auf den ersten Blick erfüllt das nicht die Erwartungen, denn ohne die Dnmt Enzyme gibt es weder mdC noch D₃mdC und somit auch keine aktive Demethylierung. Erst die Betrachtung der D₂dC-Anteile zeigt den gegenläufigen Trend. Hierbei enthält der Wildtyp etwas höhere D₂dC-Werte im Vergleich zu den Dnmt TKO Zellen. Auch in der absoluten Quantifizierung der D₄dC, D₃dC und D₂dC Spezies in der Abbildung 23 B sind diese kleinen Unterschiede sichtbar. Diese sind nach zwei gemessenen biologischen Replikaten nicht signifikant verschieden und könnten sich auch gegenseitig bedingen.

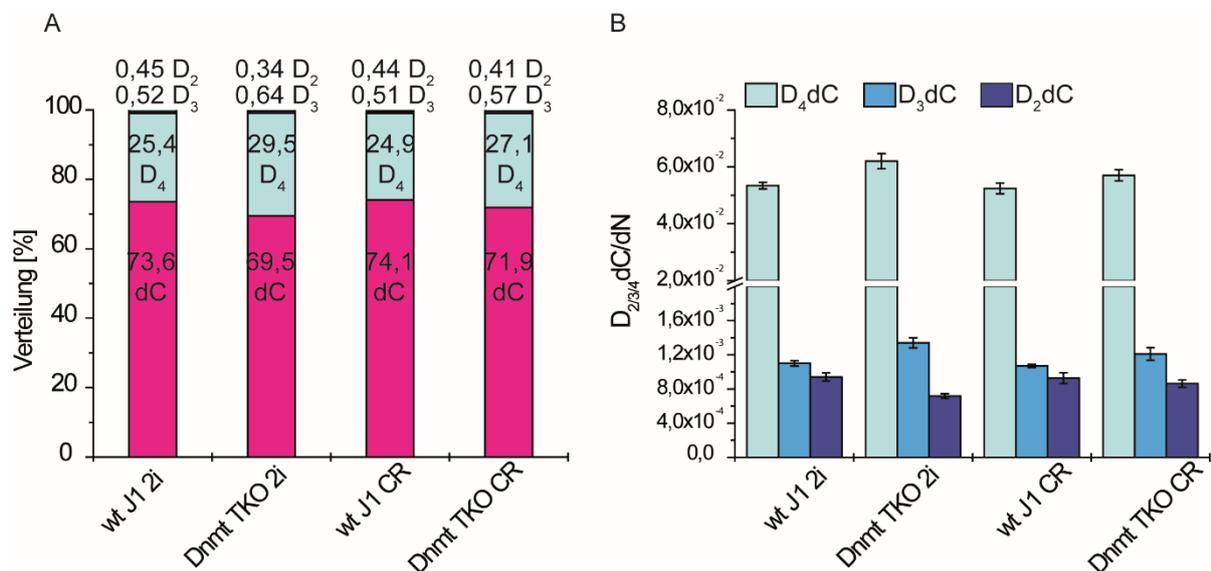


Abbildung 23: A): Relative Darstellung aller gemessenen dC-Spezies in Prozent in wt J1 und Dnmt TKO Zellen, welche entweder im *naïve* (2i) oder im *primed* (CR) Zustand kultiviert wurden. Die mESC werden *naïve* gehalten, indem dem Basismedium das Zytokin *Leukemia Inhibitory Factor* (LIF), der GSK3 Inhibitor CHIR99021 und der MEK Inhibitor PD0325901 zugesetzt wird.^[171] Den *naïve* mESCs wurden zum *priming* in ein CHIR99021 und IWR1-endo, einem Inhibitor des Wnt Signalwegs, enthaltendes Medium überführt, wobei die Zellen bei diesen Bedingungen (CR) für einige Tage pluripotent verbleiben.^[186] B): Absolute Quantifizierung von D₄dC, D₃dC und D₂dC jeweils in wt J1 und Dnmt TKO Zellen, welche entweder im *naïve* (2i) oder im *primed* (CR) Zustand kultiviert wurden. Technische Triplikate von jeweils zwei unterschiedlichen Zellkulturen wurden gemessen.

Eine etwas geringere Inkorporation von D₄dC führt somit auch zu einer niedrigeren Inkorporation der nicht isotoopenreinen Verbindung D₃dC. Dadurch ist der Anteil an natürlichem dC in der DNA höher und somit sind auch höhere Mengen an natürlich vorkommendem ¹³C₂-dC zu erwarten, welches mit dieser Messmethode nicht von D₂dC unterschieden werden kann. Wie bereits beschrieben könnte das D₂dC durch die Inkorporation des D₃dCs

(D-Markierung an der C6-Position der Base fehlt) und nachfolgender aktiver Demethylierung entstehen. Deshalb wird der Metabolit in der folgenden Diskussion dC+2 genannt.

Aufgrund dessen soll das Verhältnis der gemessenen dC-Spezies zueinander untersucht werden. Genauer gesagt, die dC-Spezies die potentiell aus ihrem Vorläufer entstanden sein könnte bzw. mit der es korrelieren könnte (Tabelle 5). Das Verhältnis von D₃dC zu D₄dC in Prozent ist in den Dnmt TKO Zellen um ca. 5 % größer als im Wildtyp. In den *naïve* Bedingungen ist der Unterschied zwischen dem Wildtyp und dem TKO sowohl bei D₂dC zu dC, als auch bei D₂dC zu D₃dC größer als unter *primed* Bedingungen.

Tabelle 5: Relatives Verhältnis der gemessenen dC-Spezies in Prozent aus Abbildung 23 A bezogen auf den jeweils korrespondierenden Wildtypen.

	D ₃ dC : D ₄ dC	dC+2 : dC	dC+2 : D ₃ dC
Dnmt TKO (<i>naïve</i>)	+5 %	-20 %	-37 %
Dnmt TKO (<i>primed</i>)	+4 %	-4 %	-14 %

In allen Fällen ist eine geringere Menge der D₂dC Spezies pro dC (-20 % bzw. -4 %) bzw. pro D₃dC (-37 % bzw. -14 %) im Dnmt TKO detektierbar als im Wildtyp. Dieser Unterschied ist unter *naïve* Bedingungen weitaus stärker ausgeprägt als unter *primed* Bedingungen. Das bedeutet, dass in den Dnmt TKO Zellen weniger der dC+2 Spezies pro dC gemessen bzw. gebildet wurde als im Wildtyp. Würde es sich bei der dC+2 Spezies nur um das natürliche ¹³C₂dC handeln müsste es direkt mit den dC Mengen korrelieren und den gleichen Wert im Wildtyp wie im Dnmt TKO liefern (d.h. 0 %). Wenn man nun annimmt, dass es sich nicht um eine Korrelation handelt, sondern um eine aktive Entstehung der dC+2 Spezies im Wildtyp und weiterhin annimmt, dass dieses dC+2 aus dem Metaboliten D₃dC entstanden sein muss, so muss man nun das Verhältnis von dC+2 zu D₃dC untersuchen. Da auch das Verhältnis von dC+2 zu D₃dC im Dnmt TKO kleiner ist als im Wildtyp (-37 % bzw. -14 %) könnte das bedeuten, dass im Wildtyp aktiv dC+2 entstanden ist. Natürlich ist ein Anstieg des dC+2 noch kein Beweis für eine aktive Demethylierung, aber es könnte ein weiterer Hinweis darauf sein. An dieser Stelle darf man allerdings nicht vergessen, dass auch ein D/H-Austausch an der Base zu einem Anstieg an dC+2 führen kann. Prädestiniert hierfür sind wie bereits erwähnt die Dnmt Enzyme. Da diese in einem Dnmt TKO nicht zugegen sind, könnte der Unterschied auch nur durch den D/H-Austausch mittels der Dnmt Enzyme zustande kommen.

Aus diesem Grund wurde das Experiment mit Tet TKO Zellen und den zugehörigen Wildtyp Zellen (E14) wiederholt. In beiden Zelltypen ist eine Methylierung durch die Dnmt Enzyme möglich. In den Tet TKO Zellen kann das gebildete mdC jedoch nicht weiter oxidiert werden. Eine aktive Demethylierung in Form einer Deformylierung bzw. Decarboxylierung ist hier also

nicht möglich. In der Abbildung 24 A sind erneut die prozentualen Verhältnisse aller gemessenen dC-Spezies abgebildet. Auch hier zeigt sich, dass die TKO Zellen eine höhere Inkorporation an D₄dC aufweisen. Die Inkorporation beträgt wieder knapp 30 %. Auch die D₃dC Verbindung folgt dem gleichen Muster: leicht höhere Werte in den Tet TKO Zellen, während sich die D₂dC bzw. dC+2 Verbindung genau gegenläufig verhält. Das sieht man besonders gut in Abbildung 24 B im Vergleich der absoluten Quantifizierung des D₃dCs und D₂dCs unter *primed* Bedingungen.

Wie zuvor beschrieben, müssen auch hier die erhaltenen relativen Ergebnisse aus Abbildung 24 A zueinander ins Verhältnis gesetzt werden (Tabelle 6). Der prozentuale Wert von D₃dC muss auf D₄dC normiert werden, denn daraus könnte diese Spezies durch aktive Demethylierung entstehen. Genauso wie der Anteil an D₂dC auf D₃dC und auf das natürliche dC normiert werden muss. Denn auch D₂dC könnte aus D₃dC (D-Markierung an der C6-Position der Base fehlt) durch aktive Demethylierung entstehen oder durch das natürliche Isotop ¹³C₂dC von dC. Interessant ist auch in diesem Experiment, dass das Verhältnis von D₃dC zu D₄dC im Tet TKO im Vergleich zum Wildtyp fast konstant bleibt.

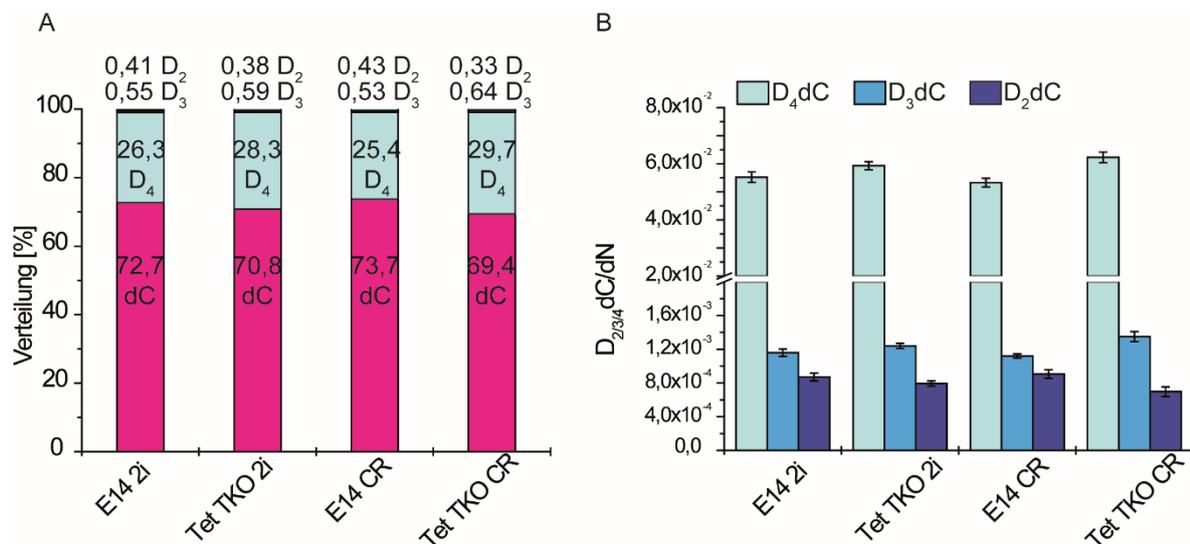


Abbildung 24: A): Relative Darstellung aller gemessenen dC-Spezies in Prozent in E14 und Tet TKO Zellen, welche entweder im *naïve* (2i) oder im *primed* (CR) Zustand kultiviert wurden. B): Absolute Quantifizierung von D₄dC, D₃dC und D₂dC jeweils in E14 und Tet TKO Zellen, welche entweder im *naïve* (2i) oder im *primed* (CR) Zustand kultiviert wurden. Technische Triplikate von jeweils zwei unterschiedlichen Zellkulturen wurden gemessen.

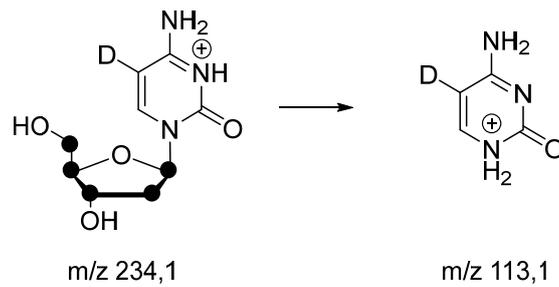
Das könnte darauf hindeuten, dass D₃dC im Wildtyp nicht vermehrt entsteht. Betrachtet man aber die Verhältnisse der dC+2 Spezies pro dC (-6 % bzw. -18 %) und D₃dC (-14 % bzw. -38 %), so sieht man hier deutlichere Unterschiede zwischen dem Wildtyp und der TKO Zelllinie. Wie in dem vorherigen Experiment diskutiert könnte das auf eine aktive Entstehung hindeuten, die weder mit den Mengen an natürlichem dC noch mit der inkorporierten Menge an D₄dC bzw. D₃dC zu korrelieren scheint.

Tabelle 6: Relatives Verhältnis der gemessenen dC-Spezies in Prozent aus Abbildung 24 A bezogen auf den jeweils korrespondierenden Wildtypen.

	D ₃ dC : D ₄ dC	dC+2 : dC	dC+2 : D ₃ dC
Tet TKO (<i>naïve</i>)	+/-0 %	-6 %	-14 %
Tet TKO (<i>primed</i>)	+3 %	-18 %	-38 %

In diesem Experiment ist des Weiteren sichergestellt, dass beide Zelllinien noch die Möglichkeit des D/H-Austauschs durch die Dnmt Enzyme aufweisen. Eine andere Möglichkeit, um diesen D/H-Austausch sichtbar und messbar machen zu können, ist die Mengen von D₃dC/dN in den Dnmt TKO Zellen mit denen der Tet TKO Zellen zu vergleichen. Denn die Dnmt TKO Zellen haben theoretisch keine Möglichkeit D₃dC zu bilden, während den Tet TKO Zellen nur die Möglichkeit eines D/H-Austauschs zur Bildung von D₃dC bleibt. Der Vergleich der quantifizierten Mengen von D₃dC/dN in den beiden Zelllinien zeigte allerdings, dass in den *naïve* Bedingungen mehr D₃dC/dN in den Dnmt TKO Zellen vorkommt. Erst der Vergleich in den *primed* Bedingungen zeigt, dass $1,4 \cdot 10^{-4}$ mehr der D₃dC Spezies pro dN in den Tet TKO Zellen vorkommt. Theoretisch könnte das die Menge an D₃dC sein, die nur durch einen D/H-Austausch zustanden gekommen ist. Da die beiden Bedingungen gegenläufige Ergebnisse zeigen, ist das kein Beweis für oder gegen einen D/H-Austausch.

Wie auch das vorherige Experiment ist dieses ein biologisches Duplikat, welches nur kleine und bisher nicht signifikante Unterschiede liefert. Die große Ungewissheit in diesem Experiment resultiert aber vor allem aus der Tatsache, dass die gemessenen Werte eine Momentaufnahme der jeweiligen DNA und der darin enthaltenen deuterierten Nukleotide einer Zelllinie sind. Die Zellen werden jeweils drei Tage mit D₄dC im Medium kultiviert und können während dieser Zeit D₄dC bzw. D₃dC (1,66 % als Verunreinigung) inkorporieren und metabolisieren. Es gibt also in diesem Versuch weder einen bestimmten Anfang noch ein eindeutiges Ende. Die detektierten deuterierten Nukleoside des Wildtyps könnten bereits mehrfache Methylierungszyklen mit anschließender aktiver Demethylierung durchlaufen haben. Dies macht eine definitive Aussage äußerst schwierig. Eine weitere Schwierigkeit ist, dass zwischen den beiden möglichen D₃dC-Spezies messtechnisch nicht unterschieden werden kann. Aus der Deuterierung an der C6-Position resultiert eine größere Ungewissheit, als dass diese Markierung von Nutzen wäre. Aus diesem Grund wurde erneut ein verbessertes Molekül synthetisiert: ¹³C₅D₁dC (Schema 5).

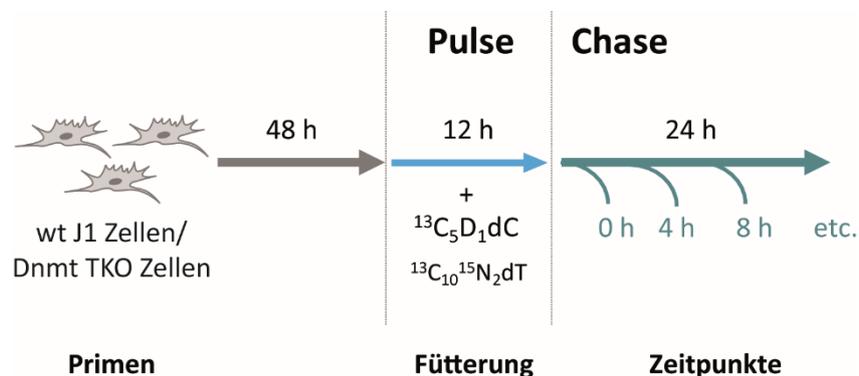


Schema 5: Die spezifische Fragmentierung des $^{13}\text{C}_5\text{D}_1\text{dCs}$. Die Fragmentierung erfolgt im positiven ESI-Modus und führt zu einer Spaltung der glykosidischen Bindung. Die ausgefüllten Kreise stellen die ^{13}C -Markierung dar.

Die fünf C-Atome des Zuckers sind ^{13}C -markiert und die C5-Position der Base ist deuteriert. Dadurch ist das Molekül 6 u schwerer als das natürliche dC und daher aufgrund der Massendifferenz leicht vom biologischen Hintergrund zu unterscheiden, der hier gegen Null geht. Auch die Synthese liefert eine höhere Isotopenreinheit; die Verunreinigung durch $^{13}\text{C}_5\text{dC}$ beträgt nur 0,8 %. Die Reinheit wurde also nochmals um 0,86 % verbessert.

Da die vorherigen Fütterungsexperimente keinen Rückschluss auf die Kinetik des möglicherweise entstandenen Produkts einer aktiven Demethylierungsreaktion zugelassen haben, wurde ein neues Experiment gestaltet. In einem Pulse-Chase-Experiment werden die Zellen über einen bestimmten Zeitraum mit markierten Nucleosiden im Medium kultiviert. Nach einer bestimmten Dauer wird das Medium durch frisches, Nucleosid-freies Medium ersetzt, die Zellen zu definierten Zeitpunkten geerntet und deren genomische DNA analysiert. Dieses Experiment ist im Schema 6 graphisch dargestellt.

In diesem Pulse-Chase-Experiment sollte der zeitliche Verlauf der Metabolisierung von $^{13}\text{C}_5\text{D}_1\text{dC}$ verfolgt werden, ohne dass die Zellen Zugriff auf das Nucleosid aus dem Medium haben. Dies sollte den fortlaufenden Neueinbau von $^{13}\text{C}_5\text{D}_1\text{dC}$ als unbekanntem Faktor vermeiden. Hierzu werden die mESCs zunächst mit $^{13}\text{C}_5\text{D}_1\text{dC}$ gefüttert.



Schema 6: Der Ablauf des Pulse-Chase Experiments. Die beiden Zelllinien werden ausgesät, über 48 h *geprimed* und für 12 h mit $^{13}\text{C}_5\text{D}_1\text{dC}$ und $^{13}\text{C}_{10}^{15}\text{N}_2\text{dT}$ im Medium kultiviert. Danach wird das Medium gewechselt und die Zellen werden zu definierten Zeitpunkten (0 h, 4 h, 8 h, 12 h und 24 h) geerntet und auf ihren Gehalt an den markierten Nucleosiden analysiert.

Mit Beginn der Probennahme werden die Zellen in Nukleosid-freiem Medium überführt, woraufhin im Abstand von einigen Stunden die Kulturen geerntet werden und deren Genom analysiert wurde. Theoretisch wäre es sinnvoll das Medium nicht gegen ein Nukleosid-freies Medium zu ersetzen, sondern gegen ein Medium, das die gleiche Konzentration an unmarkierten Nukleosiden enthält. Denn es ist denkbar, dass Nukleoside einen Einfluss auf den Metabolismus haben.

Bei den beiden Zelltypen handelte es sich um den Dnmt TKO und den korrespondierenden Wildtyp J1. Nur der Wildtyp kann alle Möglichkeiten des (De-)Methylierungsmetabolismus nutzen, während der Dnmt TKO in diesem Experiment als Negativkontrolle genutzt wurde. Die zusätzliche Verwendung von dT+12 (Abb. 25) ist notwendig, um über dessen Kinetik auf die Teilungsrate normieren zu können.

Der Grund, warum gerade dT+12 für dieses Experiment ausgewählt wurde, ist dass in der DNA vorhandenes Thymidin im Normalfall unter Zellkulturbedingungen weder auf chemische Weise beschädigt noch in andere Nukleotide umgewandelt werden kann. Zudem sind Markierungsraten von mehr als 90 % erreichbar. Eine Veränderung des Gehalts an dT+12 im Chase ist somit maßgeblich durch die passive Verdünnung während der Replikation bedingt. Die folgende Normierung der Cytidinemengen auf die Menge an dT+12 im Genom zum jeweiligen Zeitpunkt berücksichtigt also die möglicherweise unterschiedlichen Teilungsraten der verschiedenen Zelllinien und die damit verbundenen Eliminierungskinetiken. Vorteilhaft ist, dass der Einbau von dT+12 aufgrund der höheren Konzentration von Thymidin in der Zelle^[187] in größerem Maße stattfindet als der von $^{13}\text{C}_5\text{D}_1\text{dC}$, sodass sich nahezu alle Thymidin-Nukleotide im Genom mit einer Isotopenmarkierung versehen lassen.^[180] Ein weiterer Grund ist, dass im Vergleich zu den dC Metaboliten dT keiner Desaminierung unterlaufen kann und somit effizienter eingebaut wird.

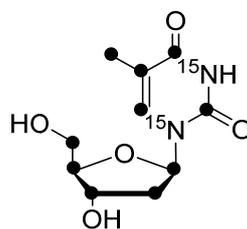


Abbildung 25: Darstellung des $^{13}\text{C}_{10}^{15}\text{N}_2\text{-dT}$ (dT+12) Standards. Die ausgefüllten Kreise stellen die ^{13}C -Markierung dar.

In einem Vorexperiment sollte zunächst die Konzentration an dT+12 im Medium bestimmt werden, bei der ausreichende Mengen ins Genom integriert werden. Dies ist unter anderem deshalb sinnvoll, da größere Mengen dT im Medium zu einem Arrest des Zellzyklus führen können. Hierzu wurden die beiden Zelltypen neben $^{13}\text{C}_5\text{D}_1\text{dC}$ entweder mit 5 μM oder mit

10 μM dT+12 gefüttert. Nach 2 h, 4 h und 8 h wurden Proben genommen und anschließend massenspektrometrisch auf den Gehalt an $^{13}\text{C}_5$ -haltigen Nukleosiden sowie an dT+12 untersucht. Der Vergleich der dT+12-Mengen bei unterschiedlichen Konzentrationen zeigt, dass bei 10 μM zwar mehr eingebaut wird als bei 5 μM , allerdings nicht annähernd die doppelte Menge, wie man erwarten würde.

Da die Zellen den Nukleosiden im Medium im Pulse-Chase-Experiment über 12 h ausgesetzt sein sollen, bevor die Probenentnahme beginnt, sollte auch bei einer Konzentration von 5 μM ausreichend dT+12 ins Genom inkorporiert werden, um auswertbare Daten zu erzeugen. Außerdem inkorporiert der TKO die isotoopenmarkierten Nukleoside sehr viel schneller ins Genom als der entsprechende Wildtyp. Bei einer Konzentration von 5 μM dT+12 im Medium hat der TKO nach 2 h bereits 35 % mehr dT+12 integriert als der Wildtyp. Nach 8 h war der Unterschied sogar auf 46 % angewachsen. Der Einbau läuft, vorbehaltlich einer linearen Zunahme, im Dnmt TKO also um 60 % schneller ab als in wt J1.

Im Pulse-Chase-Experiment wurden die wt J1 und die Dnmt TKO Zellen zunächst mit 100 μM $^{13}\text{C}_5\text{D}_1\text{dC}$ und gleichzeitig 5 μM dT+12 gefüttert. Nach 12 h wurde das Medium durch Nukleosid-freies Medium ersetzt und mit dem Ernten der ersten Kulturen begonnen. Dieser Erntezeitpunkt entspricht dem 0 h Wert. Ausgehend davon wurden nach 4 h, 8 h, 12 h und 24 h weitere Kulturen der Zellen geerntet, um den zeitlichen Verlauf des Abbaus der isotoopenmarkierten Nukleotide im Genom zu untersuchen. In der folgenden Abbildung (Abb. 26) sind die Ergebnisse der verschiedenen isotoopenmarkierten Spezies auf alle kanonischen Nukleoside des Genoms normiert gezeigt. Nach 12 h Fütterung, d.h. zum Zeitpunkt 0 h, besteht das Genom des Dnmt TKO zu 28,9 % aus dT+12, was einem vollständigen Ersatz der natürlichen Thymidinnukleotide durch die isotoopenmarkierte Spezies in der DNA entspricht. Im Wildtyp sind 24,3 % des Genoms, d.h. ca. 82 % der Thymidine durch dT+12 ersetzt (Abb. 26 E und F). Nach Entzug der Nukleoside aus dem Medium nimmt die Menge an dT+12 im TKO um 78 % schneller ab als im Wildtyp, sodass der Anteil im TKO nach 24 h nur noch 3,3 % beträgt. Im Wildtyp hingegen sind noch ca. 1/3 der Thymidine zu diesem Zeitpunkt markiert. Das Experiment deckt damit den Zeitraum, in dem sinnvolle Daten erhalten werden können, komplett ab. Man könnte den 24 h Zeitpunkt durch einen 18 h Zeitpunkt ersetzen, um auch in den TKO Zellen noch etwas höhere Werte an dT+12 zu messen. Denn die quantifizierte Menge an dT+12 in den TKO Zellen waren nach 24 h bereits recht klein und eine Normierung auf diesen Wert erschien nicht sinnvoll. Die Mengen an $^{13}\text{C}_5\text{D}_1\text{dC}$ sind zu Beginn der Messungen in beiden Zelllinien etwa gleich groß. Die Werte (Abb. 26 A) entsprechen einem Einbau von ca. 20 %. Auch die Abnahme findet in beiden Fällen gleich schnell statt (Abb. 26 B), sodass nach 24 h noch 38 % der ursprünglichen Menge an $^{13}\text{C}_5\text{D}_1\text{dC}$ vorhanden sind. Zu Beginn beinhaltet das Genom des wt J1 0,052 % $^{13}\text{C}_5\text{dC}$, das des TKOs 0,054 % (Abb. 26 D). Bezogen auf die Menge an $^{13}\text{C}_5\text{D}_1\text{dC}$ entspricht dies 1,1 % bzw. 1,2 %.

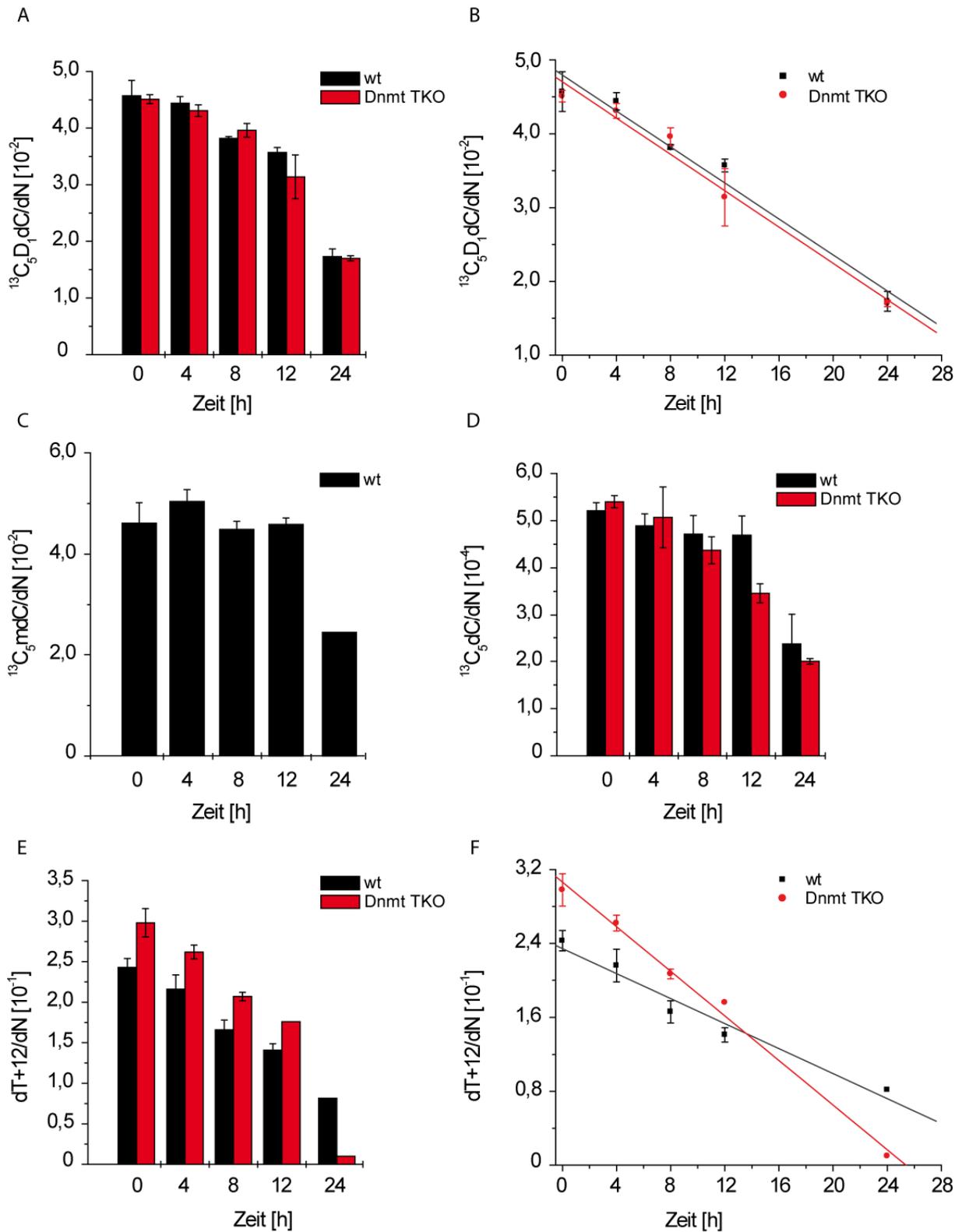


Abbildung 26: A): Absolute Quantifizierung der $^{13}\text{C}_5\text{D}_1\text{dC}$ Mengen pro dN nach 0 h, 4 h, 8 h, 12 h und 24 h in den Wildtyp Zellen und in den Dnmt TKO Zellen. B): Die Abnahme der Mengen aus A) werden als Geraden pro Zeit dargestellt. C): Absolute Quantifizierung der $^{13}\text{C}_5\text{mdC}$ Mengen pro dN nach 0 h, 4 h, 8 h, 12 h und 24 h in den Wildtyp-Zellen, die Dnmt TKO Zellen liefern kein $^{13}\text{C}_5\text{mdC}$. D): Absolute Quantifizierung der $^{13}\text{C}_5\text{dC}$ Mengen pro dN nach 0 h, 4 h, 8 h, 12 h und 24 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen. E): Absolute Quantifizierung

der dT+12 Mengen pro dN nach 0 h, 4 h, 8 h, 12 h und 24 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen. F): Die Abnahme der Mengen aus E) werden als Geraden pro Zeit dargestellt.

Der Wildtyp kann $^{13}\text{C}_5\text{D}_1\text{dC}$ zu $^{13}\text{C}_5\text{dC}$ durch Methylierung und möglicherweise anschließender aktiver Demethylierung und D/H-Austausch verstoffwechseln, während dem Dnmt TKO nur die passive Verdünnung zur Verfügung steht. Über 12 h verringert sich die Menge an $^{13}\text{C}_5\text{dC}$ auf 0,047 % im Wildtyp und auf 0,035 % im TKO. Das Verhältnis von $^{13}\text{C}_5\text{dC}$ zu $^{13}\text{C}_5\text{D}_1\text{dC}$ steigt um 15 % im Wildtyp auf 1,3 % an, im TKO sinkt es um 8 % auf 1,1 %. Dieser relative Anstieg im Wildtyp kommt entweder durch ein verzögertes Einsetzen der metabolischen Prozesse, wie z.B. der Methylierung zustande, oder ist ein weiterer Hinweis für eine aktive Demethylierung. In der Abbildung 26 C sind die Mengen an $^{13}\text{C}_5\text{mdC}$ im Wildtyp gezeigt. Der Dnmt TKO enthält erwartungsgemäß kein $^{13}\text{C}_5\text{mdC}$ und ist deshalb nicht abgebildet. Die Mengen sind die ersten 12 h annähernd konstant und in der gleichen Größenordnung wie das $^{13}\text{C}_5\text{D}_1\text{dC}$. Das ist kein valides Ergebnis und zeigt die Grenzen des Massenspektrometers in diesem Fall deutlich auf. Die zu großen Mengen an $^{13}\text{C}_5\text{mdC}$ kommen durch den sogenannten *Crosstalk* in der Kollisionszelle des Triplequadrupols zustande.^[188] Nach der Fragmentierung des $^{13}\text{C}_5\text{mdC}$ im positiven ESI Modus durch die Spaltung der glykosidischen Bindung wird eine unmarkierte Base detektiert, die sich massenspektrometrisch nicht von dem natürlichen Methylcytosin unterscheiden lässt. Auch die UHPLC kann das $^{13}\text{C}_5\text{mdC}$ nicht vom unmarkierten mdC trennen und die beiden Moleküle sind somit mehr oder weniger gleichzeitig in der Kollisionszelle und können nicht mehr voneinander unterschieden werden. Die Werte für $^{13}\text{C}_5\text{mdC}$ entsprechen somit nicht der tatsächlichen zu erwartenden Menge; $^{13}\text{C}_5\text{mdC}$ ist mit dieser Methode nicht auswertbar.

In der Abbildung 27 ist die Normierung der $^{13}\text{C}_5$ -haltigen Nukleoside auf die Menge an dT+12 gezeigt. Dieser weitere Schritt sorgt dafür, dass die unterschiedlichen Abbaukinetiken der beiden Zelltypen berücksichtigt werden. Dadurch werden die zeitlichen Veränderungen der Nukleosid-Mengen, die nicht durch die Replikation zustande kommen, sichtbar. Da der 24 h Messpunkt der Dnmt TKO Zellen kaum mehr dT+12 enthält, wurde dieser in der folgenden Auswertung nicht berücksichtigt. Auf den ersten Blick sieht man, dass die Mengen aller isotoopenmarkierten Cytidine ansteigen. Im Fall von $^{13}\text{C}_5\text{dC}$ und auch $^{13}\text{C}_5\text{mdC}$ könnte man das erwarten, da z.B. D/H-Austausch, Methylierung und auch aktive Demethylierung durch Deformylierung/Decarboxylierung zu einem Anstieg dieser Spezies führen könnte. Für $^{13}\text{C}_5\text{D}_1\text{dC}$ ist das kontraintuitiv, denn die Menge sollte mindestens im gleichen Maße wie die des dT+12 abnehmen. Ein möglicher Grund könnte sein, dass aufgrund der höheren Fütterungskonzentration von $^{13}\text{C}_5\text{D}_1\text{dC}$ (100 μM im Vergleich zu 5 μM des dT+12) sich dieses noch immer im Zytosol (*soluble Pool*) der Zellen als Nukleosid und Nukleotid befindet und somit für einen zeitverzögerten Einbau zur Verfügung steht. Diese Beobachtung erschwert die Auswertung eines Pulse-Chase-Experiments und sollte auch in Zukunft nicht außer Acht

gelassen werden. Man könnte in einem weiteren Experiment versuchen dem gegen zu wirken, indem man wie bereits erwähnt das Medium mit unmarkierten Nukleosiden versieht. Eventuell würde man so die überschüssigen markierten Nukleoside in einer Art Konkurrenzreaktion rausverdünnen können.

In der Abbildung 27 A sind die auf dT+12 normierten Mengen an $^{13}\text{C}_5\text{D}_1\text{dC}$ gezeigt. Am 0 h Messpunkt enthält der TKO 19 % weniger $^{13}\text{C}_5\text{D}_1\text{dC}$ als der Wildtyp. Wie man in der Abbildung 27 B sehen kann, ist auch die Zunahme im TKO etwas geringer, d.h. die Gerade weist eine flachere Steigung auf. Auch das ist im ersten Moment unerwartet, denn der Wildtyp ist in der Lage, $^{13}\text{C}_5\text{D}_1\text{dC}$ zu metabolisieren.

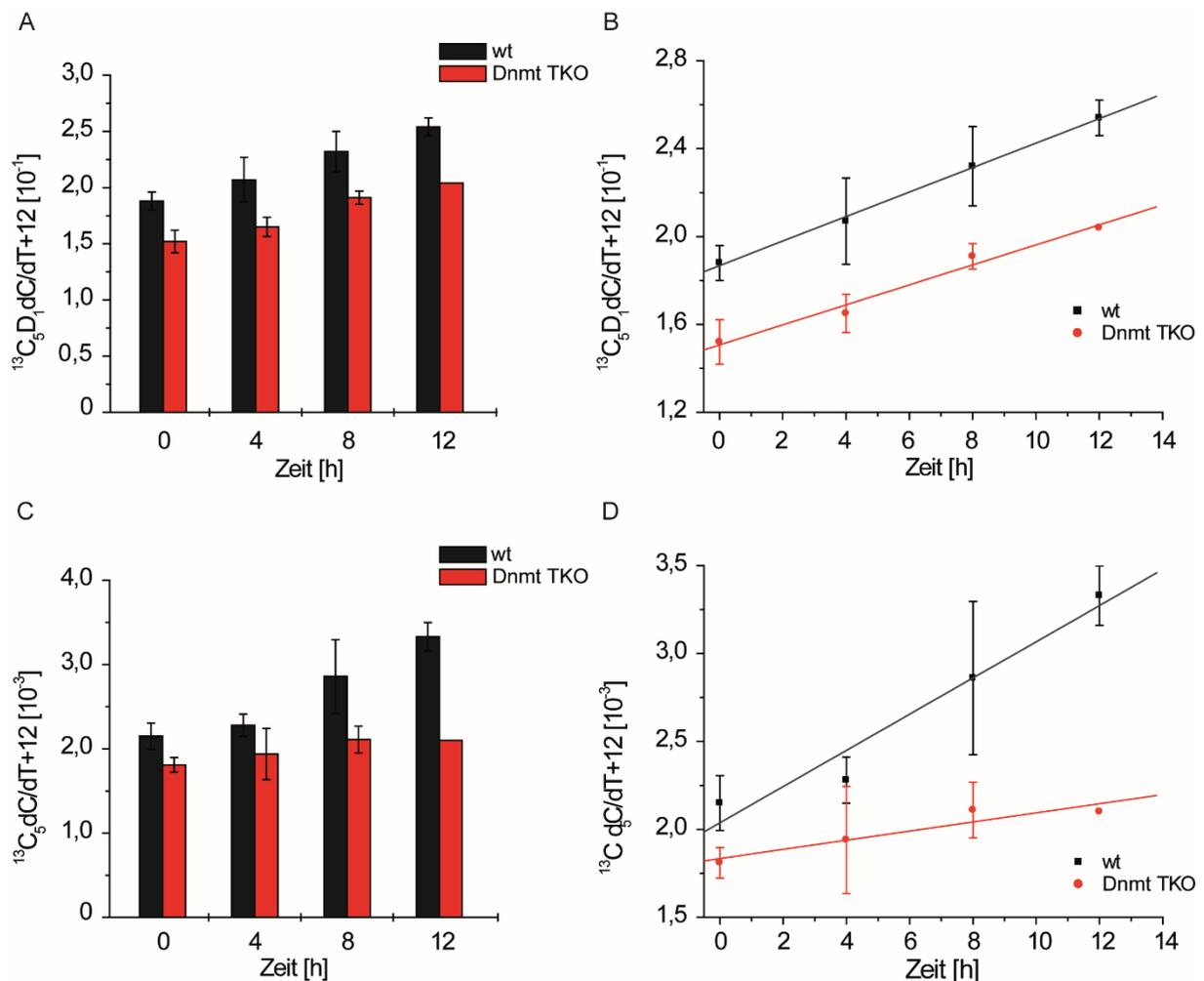


Abbildung 27: A): Absolute Quantifizierung der $^{13}\text{C}_5\text{D}_1\text{dC}$ Mengen pro dT+12 nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen. B): Die Mengen aus A) werden als Geraden pro Zeit dargestellt. C): Absolute Quantifizierung der $^{13}\text{C}_5\text{dC}$ Mengen pro dN nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen. D): Die Mengen aus C) werden als Geraden pro Zeit dargestellt.

Eine Erklärung hierfür könnte sein, dass der TKO seinen Vorrat an $^{13}\text{C}_5\text{D}_1\text{dC}$ im Zytosol durch den insgesamt schnelleren Einbau früher aufbraucht. In der Abbildung 27 C sind die Mengen an $^{13}\text{C}_5\text{dC}$ pro dT+12 abgebildet: der Wildtyp enthält auch hier zu Beginn ca. 19 % mehr $^{13}\text{C}_5\text{dC}$

als der Dnmt TKO. Im Laufe der 12 h wächst der Anteil von $^{13}\text{C}_5\text{dC}$ im Wildtyp um 55 % an, der im TKO hingegen nur um 16 %. Die Geraden in der Abbildung 27 D zeigen die ca. dreifach stärkere Steigung im Wildtyp recht deutlich. Die Prozesse, die $^{13}\text{C}_5\text{dC}$ bilden, sind hier deutlicher ausgeprägt als im Dnmt TKO. Eine stattfindende aktive Demethylierung über eine Deformylierung und Decarboxylierung würde zu einer vermehrten Bildung von $^{13}\text{C}_5\text{dC}$ führen. Aber auch ein D/H-Austausch durch die DnmTs wäre denkbar.

Dieses Pulse-Chase-Experiment zeigt, dass sich die Auswirkungen der Prozesse der Metabolisierung von $^{13}\text{C}_5\text{D}_1\text{dC}$ im Wildtyp J1 und der schnelleren Replikation im Dnmt TKO bezogen auf die Mengen der Nukleoside (dN) gegenseitig aufheben. Die Normierung der Daten auf das dT+12 ist notwendig, um die beiden Zelllinien wirklich miteinander vergleichen zu können.

Dieses Experiment sollte nun mit dem Tet TKO und dem korrespondierenden Wildtyp E14 wiederholt werden. Beiden Zelllinien ist es möglich, das inkorporierte $^{13}\text{C}_5\text{D}_1\text{dC}$ zu methylieren. Dadurch unterscheiden sich die Zelllinien nur noch in den höher oxidierten Spezies, die aber vergleichbar selten vorkommen. Das führt insgesamt zu einer besseren Vergleichbarkeit der gemessenen Werte. Ein anderer wichtiger Grund für dieses Experiment ist die potentielle Teilnahme der Dnmt Enzyme bei einem D/H-Austausch. Angenommen dieser D/H-Austausch wird durch die Dnmt Enzyme katalysiert, so ist diese Möglichkeit in beiden Zelllinien gegeben. Entsteht das potentielle Produkt $^{13}\text{C}_5\text{dC}$ durch eine Deformylierungs- bzw. Decarboxylierungsreaktion, so ist das nur im Wildtyp, nicht aber im Tet TKO möglich. Wie in Abb. 27 gezeigt, werden auch in Abb. 28 die erhaltenen $^{13}\text{C}_5\text{D}_1\text{dC}$ bzw. $^{13}\text{C}_5\text{dC}$ Mengen auf die Menge an dT+12 normiert. Die Abb. 28 A und C zeigen als Säulendiagramm die Verhältnisse von $^{13}\text{C}_5\text{D}_1\text{dC}$ bzw. $^{13}\text{C}_5\text{dC}$ pro dT+12 in Abhängigkeit von der Zeit. Durch diese Darstellung ist es schwierig, einen möglichen Trend zu erkennen. Erst die zweite Darstellung der Daten (Abb. 28 B und D) in Form einer Geraden lässt einen Vergleich zwischen Tet TKO (rot) und Wildtyp (schwarz) zu. Da die Steigungen der Geraden sowohl für das Edukt $^{13}\text{C}_5\text{D}_1\text{dC}$ pro dT+12 als auch für das potentielle Produkt $^{13}\text{C}_5\text{dC}$ pro dT+12 in beiden Zelllinien annähernd gleich sind, kann man aufgrund von diesen Daten eigentlich nicht von einer aktiven Demethylierung durch Deformylierung bzw. Decarboxylierung sprechen. Die erhaltenen Daten sehen insgesamt nach einer Korrelation aus, genauer gesagt wird im Wildtyp mehr $^{13}\text{C}_5\text{D}_1\text{dC}$ pro dT+12 eingebaut und genau das spiegelt sich in den Werten von $^{13}\text{C}_5\text{dC}$ pro dT+12 wider. Das könnte nun bedeuten, dass tatsächlich das meiste an $^{13}\text{C}_5\text{dC}$ als Verunreinigung ins Genom eingebracht wurde, oder dass die Dnmt Enzyme zu einem kleinen Teil für den D/H-Austausch zuständig sind. Auch möglich ist, dass die Zellkulturbedingungen für diese Experimente mit dem Testmolekül, $^{13}\text{C}_5\text{D}_1\text{dC}$, nicht zum gewünschten Ziel führen. Denn auch wenn die gewonnenen Daten auf den ersten Blick recht vielversprechend aussehen, darf man nicht außer Acht lassen, dass auch dieses Molekül noch Nachteile mit sich bringt wie zum

Beispiel die verzögerte Inkorporation in die DNA und die relativ geringe Inkorporationsrate generell. Ein Grund könnte hierfür der Metabolismus des dC auf Nukleosidebene sein (Schema 7).

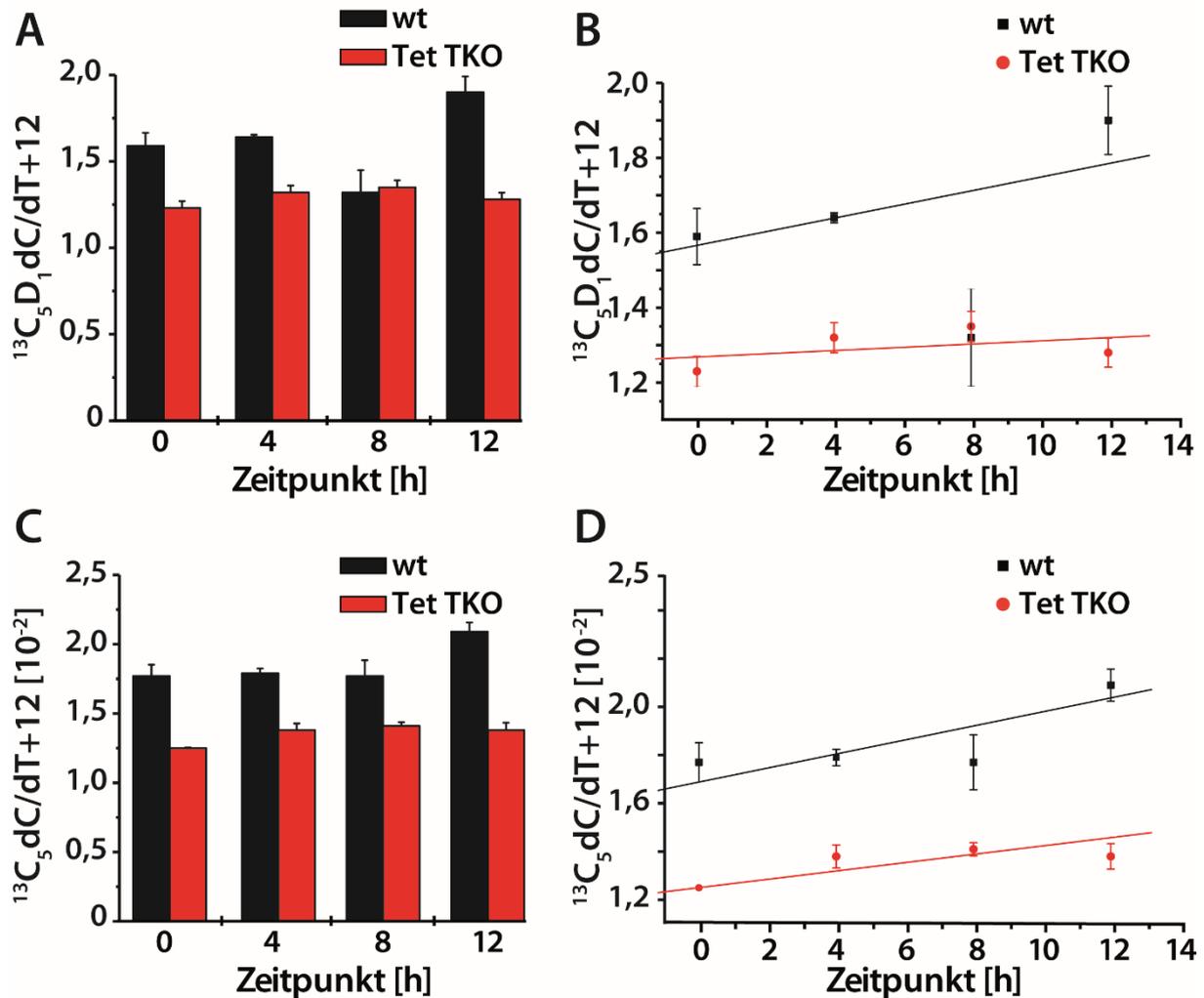
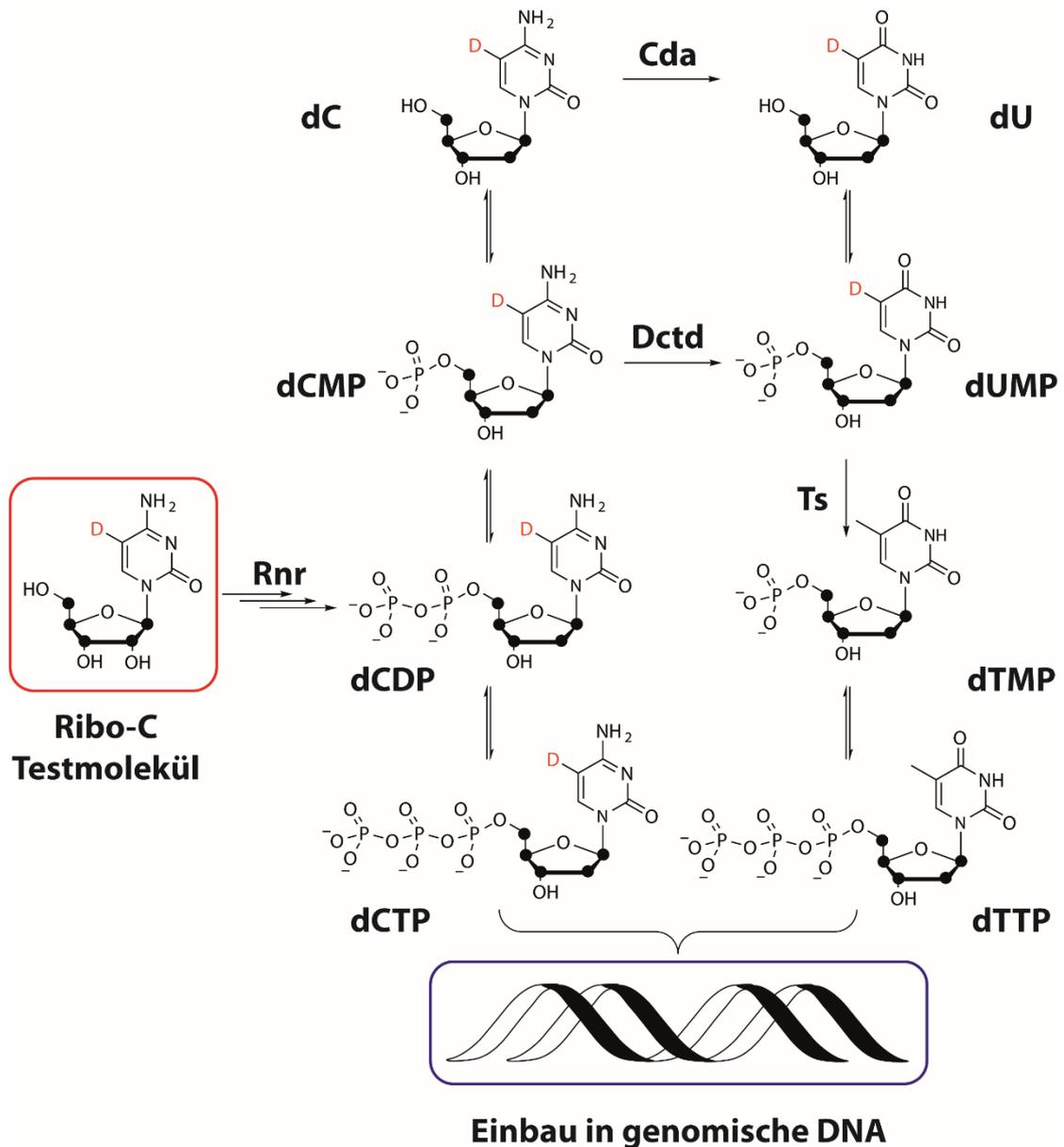


Abbildung 28: A): Absolute Quantifizierung der $^{13}\text{C}_5\text{D}_1\text{dC}$ Mengen pro dT+12 nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Tet TKO Zellen. B): Die Mengen aus A) werden als Geraden pro Zeit dargestellt. C): Absolute Quantifizierung der $^{13}\text{C}_5\text{dC}$ Mengen pro dN nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Tet TKO Zellen. D): Die Mengen aus C) werden als Geraden pro Zeit dargestellt. (Abb. von René Rahimoff)

Die Bildung des dC- bzw. $^{13}\text{C}_5\text{D}_1\text{dC}$ -Triphosphats ist eine Gleichgewichtsreaktion und auf der Stufe des Nukleosids (dC) sowie des Monophosphats können zusätzliche enzymatische Reaktionen stattfinden und so das Gleichgewicht beeinflussen. Das Nukleosid kann entweder zum Monophosphat (dCMP) phosphoryliert werden oder durch die Cytidin Desaminase (Cda) zu dU desaminiert werden. Das Gleiche gilt für dCMP, es kann entweder zum Diphosphat (dCDP) phosphoryliert werden oder durch die Desoxycytidylat Desaminase (Dctd) zu Uridinmonophosphat (dUMP) desaminiert werden. Das dU wird zum dUMP phosphoryliert und kann durch die Thymidylat Synthase (Ts) zum Thymidinmonophosphat (dTMP) methyliert

werden. Nach der anschließenden Phosphorylierung zum Thymidintriphosphat (dTTP) kann dieses, genauso wie das dCTP in die DNA inkorporiert werden. Durch diesen verzweigten Metabolismus wird deutlich, warum die Inkorporationsrate von $^{13}\text{C}_5\text{D}_1\text{dC}$ relativ gering ist und diese Inkorporation einige Zeit dauert. Ein anderes Problem im Zusammenhang mit dem Testmolekül $^{13}\text{C}_5\text{D}_1\text{dC}$ ist der sogenannte *Crosstalk* in der Kollisionszelle des Massenspektrometers, der zu einem unbrauchbaren Ergebnis bei der $^{13}\text{C}_5\text{mDc}$ Spezies führt.



Schema 7: Metabolismus des dCs auf Nukleosidebene. dC wird schrittweise durch eine Gleichgewichtsreaktion zum Triphosphat (dCTP) phosphoryliert und schließlich in die DNA eingebaut. Auf der Stufe des dCs bzw. des Monophosphats können enzymatische Deaminierungen zu dU bzw. dUMP stattfinden. Gefolgt von einer Methylierung zum dTMP und einer Phosphorylierung zum korrespondierten dT-Triphosphats (dTTP) kann auch diese in die DNA inkorporiert werden. Die schwarz ausgefüllten Kreise stellen die ^{13}C -Markierung dar. (Abb. von René Rahimoff)

Auch die Isotopenreinheit ist nicht 100 % und bringt deshalb eine leichte Ungewissheit mit. Diese Schwierigkeiten sollte durch die Entwicklung eines neuen Metaboliten behoben werden. Dieser neue Metabolit ist im Schema 7 abgebildet. Es handelt sich dabei um das RNA-Derivat $^{13}\text{C}_5\text{D}_1\text{C}$ (roter Kasten). Durch die Verwendung des RNA-Bausteins und damit Ausnutzung des *RNA Salvage Pathway* könnte man einen höheren Einbau in die DNA erreichen, denn das $^{13}\text{C}_5\text{D}_1\text{C}$, bzw. das $^{13}\text{C}_5\text{D}_1\text{C}$ kann nach der Phosphorylierung als das korrespondierende Diphosphat direkt durch die Ribonukleotidreduktase (Rnr) an der 2'-Position desoxygeniert werden. Das dadurch entstandene dCDP kann nicht mehr desaminiert werden und wird nach einer nachfolgenden Phosphorylierung direkt in die DNA eingebaut.

Das hat zwei Vorteile gegenüber dem vorherigen Metaboliten $^{13}\text{C}_5\text{D}_1\text{dC}$. Zum einen erhoffte man sich durch die Verwendung von $^{13}\text{C}_5\text{D}_1\text{C}$ einen höheren Einbau und zum anderen keine komplizierte Verteilung der ^{13}C -markierten Spezies durch den weitläufigen Metabolismus und dadurch eine vereinfachte Auswertung der Daten. Auch die Synthese des neuen Testmoleküls führte zu einer nochmals verbesserten Isotopenmarkierung von 99,7 %. Die Isotopenreinheit konnte also nochmals um 0,5 % verbessert werden. Mit diesem neuen Molekül $^{13}\text{C}_5\text{D}_1\text{C}$ wurde dann das gleiche Pulse-Chase Experiment durchgeführt werden, wie es in Schema 6 beschrieben wurde. Das Experiment wurde mit den Dnmt TKO Zellen sowie mit den Tet TKO Zellen und den jeweils korrespondierenden Wildtypen durchgeführt. In der folgenden Abbildung sind die quantifizierten Mengen des zusätzlich gefütterten dT+12 pro dN sowie das inkorporierte $^{13}\text{C}_5\text{D}_1\text{dC}$ als auch das potentielle C-C Bindungsbruchprodukt $^{13}\text{C}_5\text{dC}$ pro dN abgebildet.

Als erstes sollte überprüft werden, wie gut die Inkorporation des RNA Testmoleküls $^{13}\text{C}_5\text{D}_1\text{C}$ über den *RNA Salvage Pathway* funktioniert. Im Vergleich zur Fütterung mit dem korrespondierenden DNA Nukleosid erreicht man über diese Methode eine doppelt so hohe Inkorporation von $^{13}\text{C}_5\text{D}_1\text{dC}$ von ca. 43 % am 0 h Zeitpunkt (Abb. 29 C). Es wurden 200 μM $^{13}\text{C}_5\text{D}_1\text{C}$ gefüttert. Die anfängliche Inkorporation von $^{13}\text{C}_5\text{D}_1\text{dC}$ ist im Dnmt TKO um 3 % höher als im Wildtyp. Im Chase zeigen beide Geraden eine sehr langsame Reduktion der Mengen und weisen eine vergleichbare negative Steigung auf (Abb. 29 D). Allerdings fällt deutlich auf, dass der Einbau von dT+12 (Abb. 29 A) um mehr als eine Zehnerpotenz geringer ist, als bei den bisherigen Pulse-Chase Experimenten. Es könnte sein, dass der Einbau des RNA Testmoleküls über den *RNA Salvage Pathway* die gleichzeitige Inkorporation des dT+12 beeinträchtigt. Die Darstellung der Menge von dT+12 pro dN in Abhängigkeit von der Zeit als Geraden zeigt deutlich die unterschiedliche Geschwindigkeit der beiden Zelllinien in der Abnahme über die Zeit (Abb. 29 B). Die Gerade des Dnmt TKOs startet bei einer minimal höheren Inkorporationsmenge von dT+12 und hat eine etwas flachere negative Steigung. Der Vergleich der Mengen an dT+12 des 0 h Zeitpunkts mit dem 12 h Zeitpunkt zeigt, dass der Wert im Dnmt TKO nur um 22 % absinkt, während er im Wildtyp um 32 % abnimmt.

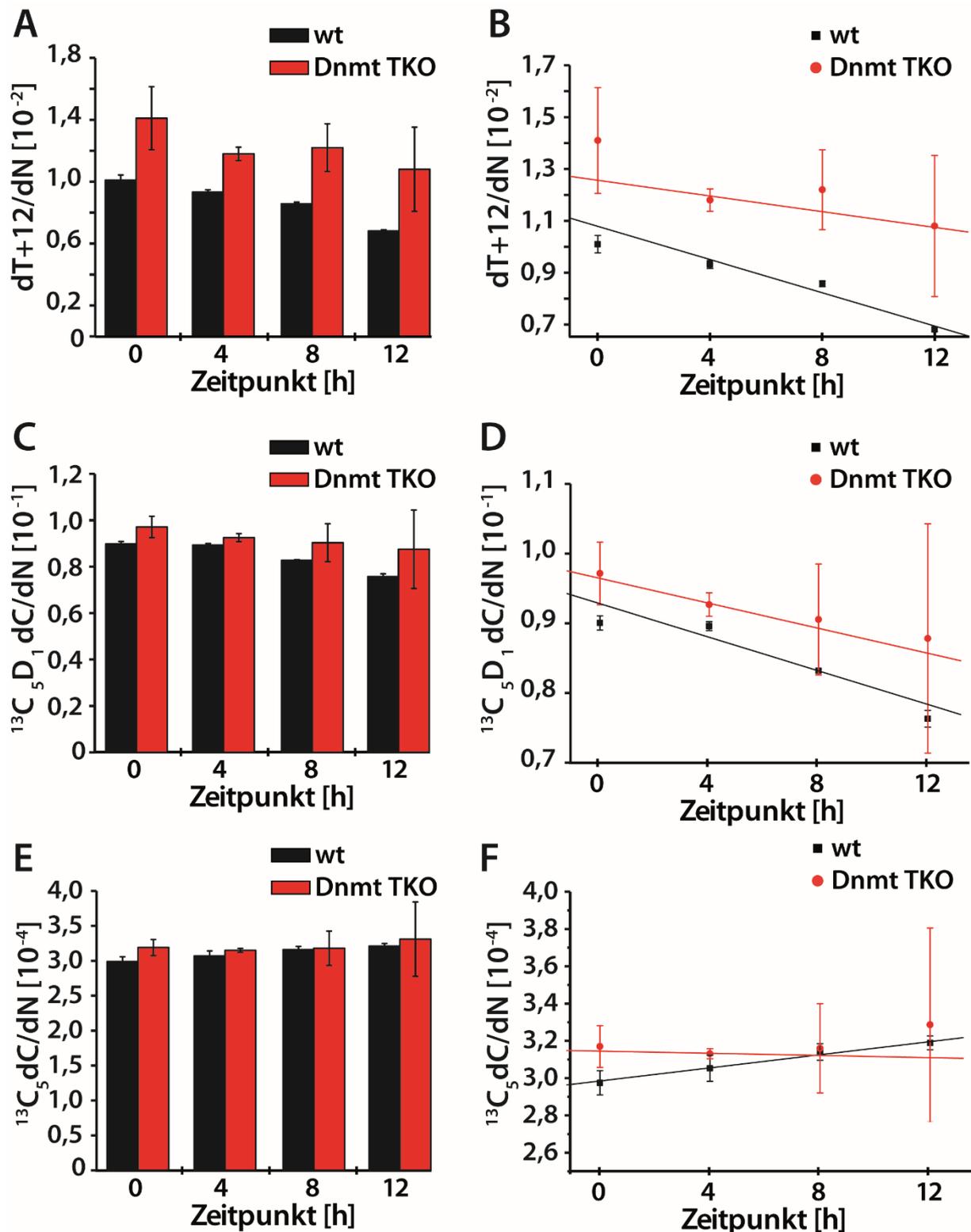


Abbildung 29: A): Absolute Quantifizierung der dT+12 Mengen pro dN nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen. B): Die Mengen aus A) werden als Geraden pro Zeit dargestellt. C): Absolute Quantifizierung der ¹³C₅D₁dC Mengen pro dN nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen. D): Die Mengen aus C) werden als Geraden pro Zeit dargestellt. E): Absolute Quantifizierung der ¹³C₅dC Mengen pro dN nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen. F): Die Mengen aus E) werden als Geraden pro Zeit dargestellt. (Abb. von René Rahimoff)

Die Daten des Tet TKOs und des korrespondierenden Wildtyps sind nicht gezeigt, vermitteln aber den gleichen Trend. Wie in den vorausgegangenen Experimenten wird an dieser Stelle wieder deutlich, wie wichtig die Normierung auf das zusätzlich gefütterte dT+12 ist. Die absoluten Mengen des potentiellen C-C Bindungsbruchprodukts $^{13}\text{C}_5\text{dC}$ sind pro dN in den Abbildungen 29 E und F gezeigt. Die Mengen scheinen über den kompletten Zeitraum von 12 h fast gleich zu bleiben und auch nicht mit der Menge des eingebauten $^{13}\text{C}_5\text{D}_1\text{dC}$ zu korrelieren. Geringe Unterschiede der gebildeten Mengen von $^{13}\text{C}_5\text{dC}$ pro dN werden in der Abbildung 29 F etwas deutlicher. Durch die Darstellungsweise über Geraden sieht man, dass die Menge im Wildtyp leicht ansteigt, während sie im Dnmt TKO leicht abnimmt. Doch auch diese Abbildung ist schwer zu interpretieren; erst durch die Normierung auf die dT+12 Werte in der Abbildung 30 werden die tatsächlichen Unterschiede sichtbar.

In der Abbildung 30 A wird die inkorporierte Menge an $^{13}\text{C}_5\text{D}_1\text{dC}$ pro dT+12 in den Dnmt TKO Zellen und dem korrespondierendem Wildtyp abgebildet und auf den 0 h Zeitpunkt normiert. Insgesamt steigt die Menge an $^{13}\text{C}_5\text{D}_1\text{dC}$ pro dT+12 in beiden Zelllinien leicht über die Zeit an. Jedoch ist der Anstieg nicht linear, denn der 4 h bzw. 8 h Zeitpunkt sind entweder fast gleich (wt J1) oder sogar invertiert (Dnmt TKO). Das könnte entweder an messtechnischen Ungenauigkeiten liegen oder daran, dass die Zellteilung Auswirkungen auf die tatsächlich quantifizierte Menge hat. Die dazugehörigen $^{13}\text{C}_5\text{dC}/\text{dT}+12$ Werte aus diesem Experiment sind in der Abbildung 30 B gezeigt. Auch diese steigen über die 12 h mehr oder weniger stetig an. Interessant ist hierbei, dass die Werte im Wildtyp aussehen, als würden sie fast exponentiell ansteigen, während die Werte im Dnmt TKO nur leicht und unstetig ansteigen. Das ist ein starker Hinweis darauf, dass dieses $^{13}\text{C}_5\text{dC}$ aktiv im Wildtyp gebildet wurde. Jedoch ist an dieser Stelle noch immer fraglich, ob es durch eine C-C Bindungsbruchreaktion oder einen D/H-Austausch eingeführt wurde. Deswegen wurde das Experiment auch mit dem Tet TKO und dem korrespondierenden Wildtyp wiederholt. In der Abbildung 30 C sind wieder die $^{13}\text{C}_5\text{D}_1\text{dC}$ pro dT+12 Werte abgebildet und in der Abbildung 30 D die korrespondierenden $^{13}\text{C}_5\text{dC}$ pro dT+12 Werte. Dadurch wird sichtbar, dass die resultierenden Kurven des Wildtyps und des Tet TKOs für die beiden Derivate jeweils mehr oder weniger deckungsgleich sind. Die Werte für $^{13}\text{C}_5\text{dC}/\text{dT}+12$ steigen sowohl im Wildtyp als auch im Tet TKO stetig an, unterscheiden sich aber fast nicht. Dieses Ergebnis weist darauf hin, dass es in beiden Zelllinien zu einer aktiven Bildung von $^{13}\text{C}_5\text{dC}$ kommt, dieses aber aus einem D/H-Austausch, verursacht durch die Dnmt Enzyme resultiert. Man sieht man, dass ab dem 8 h Zeitpunkt eine Art Plateau erreicht ist und sich die Menge an $^{13}\text{C}_5\text{D}_1\text{dC}$ genauso verhält wie die Menge an dT+12.

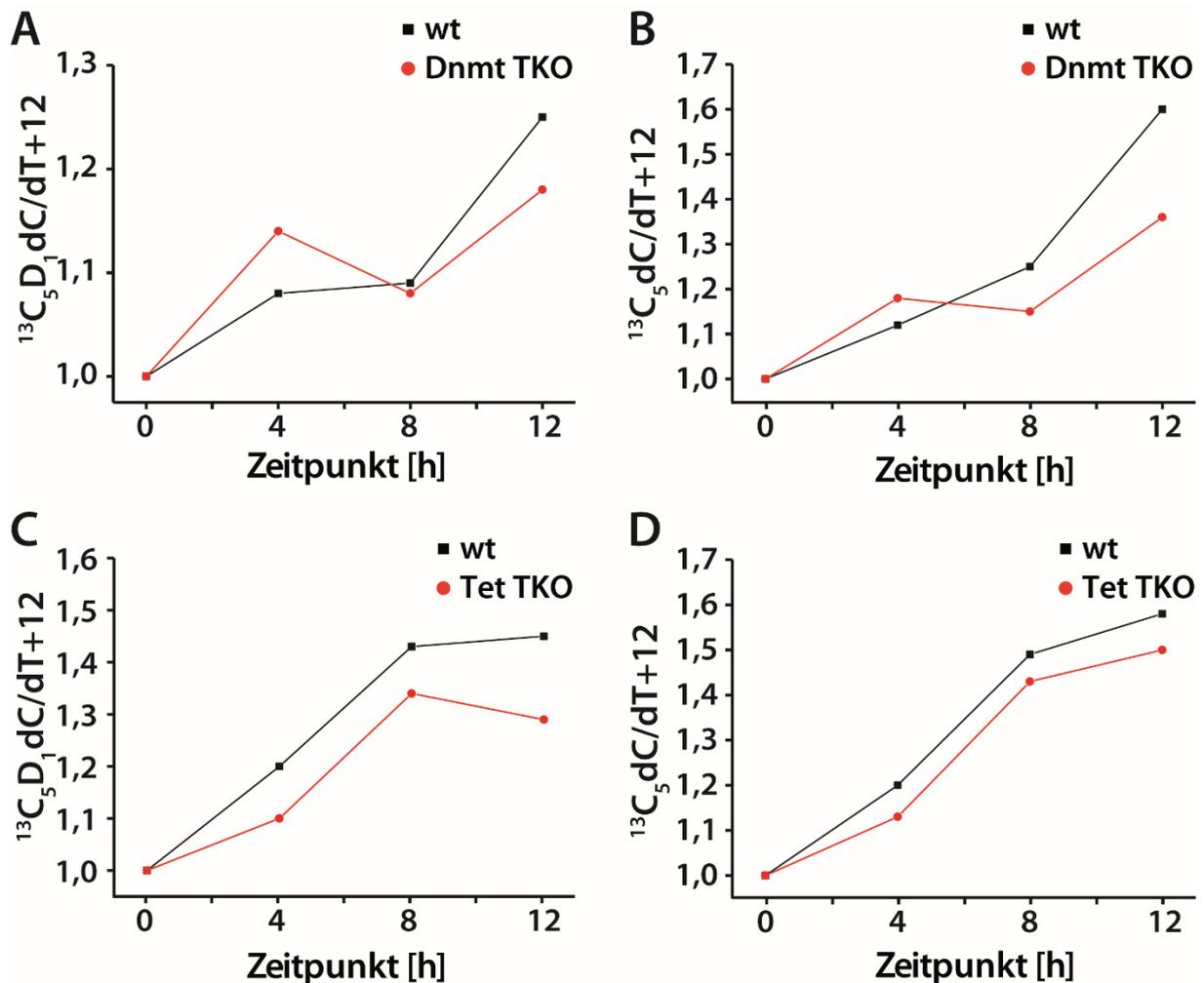


Abbildung 30: A): Absolute Quantifizierung der $^{13}\text{C}_5\text{D}_1\text{dC}$ Mengen pro dT+12 nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen, dargestellt als Gerade der Mengen pro Zeit. B): Absolute Quantifizierung der $^{13}\text{C}_5\text{dC}$ Mengen pro dT+12 nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen, dargestellt als Gerade der Mengen pro Zeit C): Absolute Quantifizierung der $^{13}\text{C}_5\text{D}_1\text{dC}$ Mengen pro dT+12 nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Tet TKO Zellen, dargestellt als Gerade der Mengen pro Zeit. D): Absolute Quantifizierung der $^{13}\text{C}_5\text{dC}$ Mengen pro dT+12 nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen und in den Dnmt TKO Zellen, dargestellt als Gerade der Mengen pro Zeit.

In einem zukünftigen Experiment wäre es daher sinnvoll, den ersten Erntezeitpunkt erst nach 8 h zu nehmen, denn davor scheint es noch immer zum Einbau von $^{13}\text{C}_5\text{D}_1\text{dC}$ zu kommen. Der Grund hierfür könnte zum Beispiel die verzögerte Inkorporation über den *RNA Salvage Pathway* sein. Als nächstes sollte überprüft werden, ob durch das neue Testmolekül auch das entstandene $^{13}\text{C}_5\text{mdC}$ quantifizierbar und sinnvoll auszuwerten ist oder ob es wieder zu einem *Crosstalk* in der Kollisionszelle kommt. Hierfür wurden die quantifizierten Mengen an natürlichem mdC (rot) im Vergleich zum markierten $^{13}\text{C}_5\text{mdC}$ (blau) in der folgenden Abbildung sowohl für den Wildtypen (Abb. 31 A) als auch für den Tet TKO (Abb. 31 B) dargestellt. Durch die Berechnung der Summe der beiden Derivate und dem anschließenden Vergleich mit der ungenährten Kontrolle (0 h Zeitpunkt) wird deutlich, dass die Werte zu den verschiedenen

Zeitpunkten mit nur kleinen Schwankungen die gleichen Ergebnisse liefern. Die angegebenen Prozentzahlen zeigen den Verlust der Isotopenmarkierung des $^{13}\text{C}_5\text{mdCs}$ über 12 Stunden an. Der Vergleich des Tet TKOs mit dem korrespondierenden Wildtyp zeigt, dass die prozentuale Abnahme der Markierung 8-10 % beträgt.

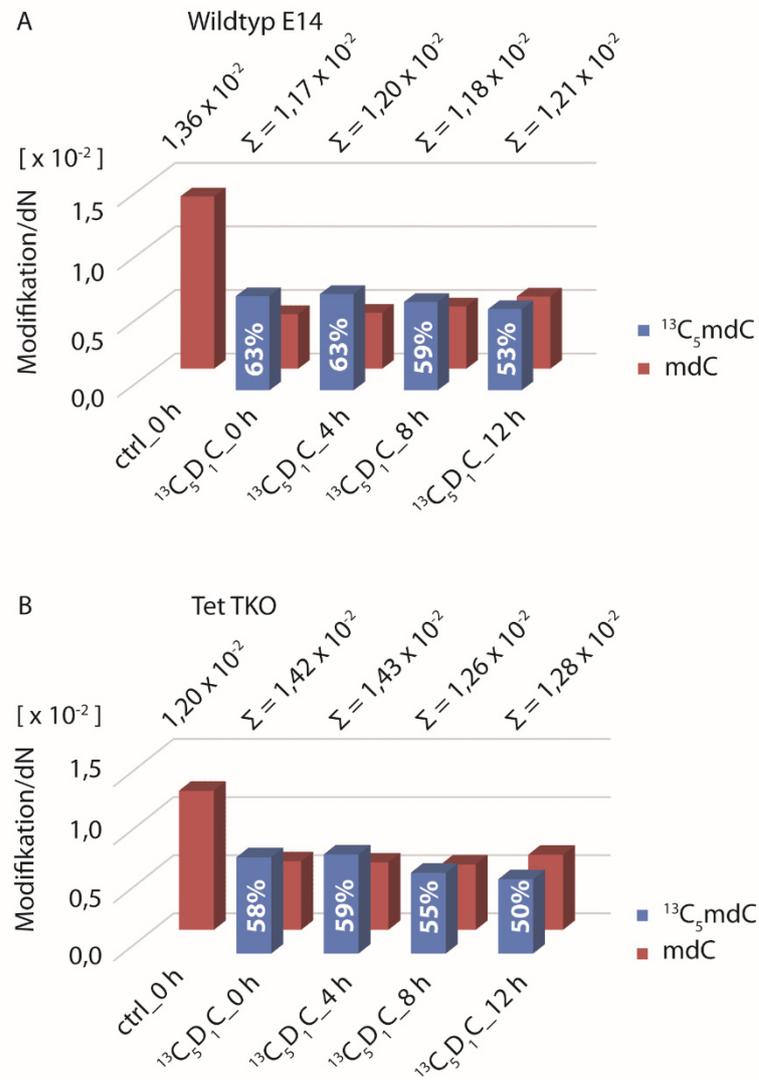


Abbildung 31: A): Absolute Quantifizierung des mdCs (rot) bzw. des $^{13}\text{C}_5\text{mdCs}$ (blau) pro dN nach 0 h, 4 h, 8 h und 12 h in den Wildtyp-Zellen. Für den 0 h Zeitpunkt ist sowohl eine Negativkontrolle, d.h. ohne $^{13}\text{C}_5\text{D}_1\text{C}$ Fütterung als auch der erste Erntezeitpunkt des *Pulse-Chase* Experiments abgebildet. B): Absolute Quantifizierung des mdCs (rot) bzw. des $^{13}\text{C}_5\text{mdCs}$ (blau) pro dN nach 0 h, 4 h, 8 h und 12 h in den Tet TKO Zellen. Für den 0 h Zeitpunkt ist sowohl eine Negativkontrolle, d.h. ohne $^{13}\text{C}_5\text{D}_1\text{C}$ Fütterung als auch der erste Erntezeitpunkt des *Pulse-Chase* Experiments abgebildet.

Die ermittelten Ergebnisse lassen darauf schließen, dass es durch das neue Testmolekül zu weniger *Crosstalk* in der Kollisionszelle kommt und die Daten für $^{13}\text{C}_5\text{mdC}$ und mdC quantifizierbar sind.

Die hohen Anforderungen an die Reinheit der zu verfütternden Metaboliten wurden in diesem Projekt sehr deutlich. Nicht nur eine hohe Inkorporationsrate in die DNA ist wichtig, sondern auch die massenspektrometrischen Eigenschaften. In diesem Fall war eine ständige Weiterentwicklung der Testmoleküle erforderlich. Die Verbesserung von D_4dC zu $^{13}\text{C}_5\text{D}_1\text{dC}$ und schließlich zu dem RNA-Derivat $^{13}\text{C}_5\text{D}_1\text{C}$ zeigen, wie aufwändig und komplex massenspektrometrische Analysen einer biochemischen Verfütterungsstudie sind. Auch abschließend ist keine definitive Aussage über das Vorkommen einer C-C Bindungsbruchreaktion möglich. Die Hinweise deuten entweder auf eine aktive Entfernung der D-Markierung an der C5-Position durch Dnmt Enzyme und somit einem direkten D/H-Austausch oder auf eine aktive Demethylierung in Form einer Deformylierung oder Decarboxylierung hin. Durch weitere Pulse-Chase Experimente mit dem RNA-Derivat $^{13}\text{C}_5\text{D}_1\text{C}$ könnte es in Zukunft aber dennoch möglich sein, Beweise für den einen oder den anderen Weg zu sammeln. Wichtig wäre es, ein Zellsystem zu finden, indem man auch tatsächliche Veränderungen der mdC Werte über die Zeit beobachten kann. Wie in der Abbildung 32 gezeigt, bleiben die Gesamtmengen an mdC (natürliches mdC und $^{13}\text{C}_5\text{mdC}$) über diesen Zeitraum von 12 h annähernd konstant. Das könnte darauf hindeuten, dass es über diesen Zeitraum keine aktive Demethylierung gibt oder aber, dass es sich um eine Art Fließgleichgewicht handelt und zu jeder Zeit aktive Demethylierung und gleichzeitige Remethylierung stattfinden. Im Folgenden sollte man sich auf ein Zellsystem konzentrieren in welchem man eine tatsächliche Abnahme der Mengen an mdC oder den höher oxidierten Spezies, hmdC , fdC und cadC zu einem bestimmten Zeitpunkt detektieren kann.

4.2.3 Zusammenfassung der Projektbeiträge

Die erste Synthese des D₄dCs wurde von *Alexander Schön* während seiner Masterarbeit unter der Anleitung von *René Rahimoff* durchgeführt. Alle weiteren Metaboliten zur Verfütterung (D₄dC, ¹³C₅D₁dC und ¹³C₅D₁C) wurden von *René Rahimoff* synthetisiert. Die Zellkulturversuche wurden von *Angie Kirchner* durchgeführt und entwickelt. Auch die Isolation aller DNA Proben wurde von ihr vorgenommen. Die Entwicklung der Quantifizierungsmethoden für alle gefütterten deuterierten dC-Spezies und deren Derivaten wurde von mir entwickelt. Diese basierten auf der Vorarbeit von *Toni Pfaffeneder* und *Jessica Steinbacher* für das ¹⁵N₂-D₂-dC Projekt. Der enzymatische Verdau, die Messung und Auswertung der Daten wurden von mir vorgenommen. Die Daten wurden gemeinsam von *Angie Kirchner*, *René Rahimoff* und mir interpretiert.

4.2.4 Material und Methoden

4.2.4.1 Zelllinien

Die verwendeten J1-mESCs (embryonalen Stammzellen der Maus, *murine Embryonic Stem Cells*) entstammen der Zelllinie 129/Sv/Jae. Die Dnmt-TKO-J1-mESCs wurden von Maskai Okano bereitgestellt.^[189] Die Zelllinien Tdg^{+/+} und Tdg^{-/-} wurden von Primo Schär zur Verfügung gestellt. Sowohl der Wildtyp E14Tg2a^[190] als auch die Tet-TKO-Zellen^[191] wurden von Yi Zhang bezogen.

4.2.4.2 Enzymatischer Verdau

Der enzymatische Verdau der genomischen DNA aus den Fütterungsexperimenten mit D₄dC erfolgte wie in dem in Abschnitt 3.4 beschriebenen Protokoll. Der enzymatische Verdau der genomischen DNA aus den Fütterungsexperimenten mit ¹³C₅D₁dC und ¹³C₅D₁C erfolgte wie in dem in Abschnitt 3.1 beschriebenen Methodenteil.

4.2.4.3 UHPLC-MS/MS Analyse

Die quantitative UHPLC-MS/MS-Analyse der hydrolysierten DNA-Proben erfolgte wie unter 3.4 beschrieben mit einem 1290 UHPLC-System von Agilent, welches über einen UV Detektor verfügt und mit einem 6490 Triple-Quadrupol Massenspektrometer von Agilent verbunden war. Die Chromatographie wurde mit einer Poroshell 120 SB-C8Säule (Agilent, 2.7 µm, 2.1 mm × 150 mm) bei 35 °C mit einem Gradienten aus Wasser und Acetonitril mit jeweils 0,0075 % (v/v) Ameisensäure durchgeführt. Der Fluss betrug 0,35 mL/min: 0 →4 min; 0 →3,5% (v/v) MeCN; 4 →6,9 min; 3,5 →5 % MeCN; 6,9 →7,2 min; 5 →80 % MeCN; 7.2 →10,5 min; 80 % MeCN; 10,5 →11,3 min; 80 →0 % MeCN; 11,3 →13 min; 0 % MeCN. Die Quellparameter waren dieselben wie unter 3.4 beschrieben. Die für die jeweiligen Analyten spezifischen Parameter sind in Tabelle 7 und 8 zusammengefasst.

Tabelle 7: Substanzabhängige LC/MS-MS Parameter zur Quantifizierung der markierten Nukleoside in genomischer DNA aus dem Fütterungsexperiment mit D₄dC.

Verbindung	Precursor Ion (m/z)	MS1 Auflösung	Produkt Ion (m/z)	MS2 Auflösung	Dwell time [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 1,5-3,1 min								
dC	228.1	Unit	112.1	Unit	5	30	5	Positiv
D ₂ dC	230.11	Unit	112.05	Unit	40	30	5	Positiv
D ₃ dC	231.12	Unit	113.06	Unit	40	30	5	Positiv
D ₄ dC	232.12	Unit	114.06	Unit	40	30	5	Positiv
[¹⁵ N ₃ , ¹³ C ₉]-dC	240.1	Unit	119.1	Unit	20	30	5	Positiv
mdC	242.11	Unit	126.07	Unit	40	40	5	Positiv
[D ₃]-mdC	245.13	Unit	129.09	Unit	40	40	5	Positiv
D ₃ mdC	245.13	Unit	127.07	Unit	40	40	5	Positiv
hmdC	258.11	Unit	142.06	Unit	40	27	5	Positiv
[¹⁵ N ₂ ,D ₂]-hmdC	262.12	Unit	146.07	Unit	40	27	5	Positiv
D ₃ hmdC	261.13	Unit	143.07	Unit	40	27	5	Positiv
cadC	272.09	Unit	156.04	Unit	20	5	5	Positiv
[¹⁵ N ₂]-cadC	274.08	Unit	158.03	Unit	20	5	5	Positiv
D ₃ cadC	275.11	Unit	157.05	Unit	20	5	5	Positiv
Zeitsegment 3,1-4,4 min								
hmdU	257.08	Wide	214.07	Wide	55	7	5	Negativ
[D ₂]-hmdU	259.09	Wide	216.08	Wide	60	7	5	Negativ
D ₃ hmdU	260.09	Wide	217.09	Wide	60	7	5	Negativ
fdU	255.06	Wide	212.06	Wide	55	6	5	Negativ
D ₃ fdU	258.1	Wide	214.1	Wide	60	6	5	Negativ
dU	227.07	Wide	184.06	Wide	60	5	5	Negativ
[¹⁵ N ₂]-dU	229.06	Wide	185.06	Wide	60	5	5	Negativ
D ₃ dU	230.08	Wide	187.08	Wide	60	5	5	Negativ
Zeitsegment 4,4-9,0 min								
8oxodG	284.1	Wide	168.05	Wide	40	9	7	Positiv
[¹⁵ N ₅]-8oxodG	289.08	Wide	173.04	Wide	40	9	7	Positiv
fdC	256.09	Unit	140.05	Unit	75	5	5	Positiv
[¹⁵ N ₂]-fdC	258.09	Unit	142.05	Unit	30	5	5	Positiv
D ₃ fdC	259.11	Unit	141.05	Unit	75	5	5	Positiv
[¹³ C ₅ , ¹⁵ N ₂]-fdC	263.1	Unit	142.05	Unit	75	5	5	Positiv
dT	243.1	Unit	127.05	Unit	30	3	5	Positiv
D ₃ dT	246.12	Unit	128.06	Unit	30	3	5	Positiv

Tabelle 8: Substanzabhängige LC/MS-MS Parameter zur Quantifizierung der markierten Nucleoside in genomischer DNA aus dem Fütterungsexperiment mit $^{13}\text{C}_5\text{D}_1\text{dC}$ und $^{13}\text{C}_5\text{D}_1\text{C}$.

Verbindung	Precursor Ion (m/z)	MS1 Auflösung	Produkt Ion (m/z)	MS2 Auflösung	Dwell time [ms]	CE (V)	CAV (V)	Polarität
Zeitsegment 1,5-3,8 min								
$^{13}\text{C}_5\text{D}_1\text{dC}$	234.12	Unit	113.06	Unit	50	30	5	Positiv
$^{13}\text{C}_5\text{dC}$	233.11	Unit	112.05	Unit	50	30	5	Positiv
$^{15}\text{N}_3,^{13}\text{C}_9\text{-dC}$	240.1	Unit	119.1	Unit	50	30	5	Positiv
mdC	242.11	Unit	126.07	Unit	40	40	5	Positiv
$[\text{D}_3]\text{-mdC}$	245.13	Unit	129.09	Unit	40	40	5	Positiv
$^{13}\text{C}_5\text{mdC}$	247.13	Unit	126.07	Unit	40	40	5	Positiv
hmdC	258.11	Unit	142.06	Unit	50	27	5	Positiv
$^{15}\text{N}_2, \text{D}_2\text{-hmdC}$	262.12	Unit	146.07	Unit	50	27	5	Positiv
$^{13}\text{C}_5\text{hmdC}$	263.13	Unit	142.06	Unit	50	27	5	Positiv
cadC	272.09	Unit	156.04	Unit	20	5	5	Positiv
$^{15}\text{N}_2\text{-cadC}$	274.08	Unit	158.03	Unit	20	5	5	Positiv
Zeitsegment 3,8-9,0 min								
8oxodG	284.1	Unit	168.05	Unit	25	9	7	Positiv
$^{15}\text{N}_5\text{-8oxodG}$	289.08	Unit	173.04	Unit	25	9	7	Positiv
fdC	256.09	Unit	140.05	Unit	100	5	5	Positiv
$^{13}\text{C}_5,^{15}\text{N}_2\text{-fdC}$	263.1	Unit	142.05	Unit	100	5	5	Positiv
$^{13}\text{C}_5\text{fdC}$	261.11	Unit	140.05	Unit	100	5	5	Positiv
$^{15}\text{N}_2,^{13}\text{C}_1\text{-dT}$	246.12	Unit	130.05	Unit	40	3	5	Positiv
$^{15}\text{N}_2,^{13}\text{C}_{10}\text{-dT}$	255.12	Unit	134.06	Unit	40	3	5	Positiv
$^{13}\text{C}_5\text{dT}$	248.11	Unit	127.05	Unit	40	3	5	Positiv

4.2.4.4 Auswertung

Die Auswertung erfolgte wie im Abschnitt 3.1 beschrieben. Die Kalibriergeraden der verschiedenen Modifikationen wurden aus bereits veröffentlichten Publikationen der Arbeitsgruppe verwendet.^[32, 192] Die Darstellung der Daten erfolgte mit Origin und Adobe Illustrator.

4.3 Studien zur Identität des Enzyms mit Deformylierungsaktivität

4.3.1 Prolog

Die Studien aus 3.4 und 4.2 verwenden verschiedene Isotopenmarkierungen, um den aktiven Demethylierungsprozess in genomischer DNA zu erforschen. Vieles deutet daraufhin, dass es zusätzlich zu der BER einen alternativen Weg zur Entfernung von 5-Formylcytidin durch einen C-C Bindungsbruch, d.h. eine Deformylierungsreaktion, gibt. Seither stellt sich vor allem die Frage, welches Enzym eine solche Reaktion katalysieren könnte. Wie bereits in der Veröffentlichung aus dem Abschnitt 3.4 beschrieben, gibt es in der Natur eine Vielzahl an sogenannten Deformylasen^[193] wie beispielsweise das Enzym Lanosterol 14 α -Demethylase (CYP51A1). Die Reaktion erfolgt in drei Schritten, wobei jeder davon ein Molekül Sauerstoff und zwei Moleküle NADPH als Reduktionsäquivalent verbraucht.^[194] Die ersten zwei Schritte wandeln die 14 α -Methylgruppe zuerst in den 14 α -Carboxylalkohol und schließlich in den 14 α -Carboxyaldehyd um. Diese Reaktionen sind typisch für die P450-Monooxygenasen. Diese Enzyme enthalten Häm als Cofaktor.^[195] Das Eisen des Häm Cofaktors ist über einen Thiolat-Liganden an einen konservierten Cystein-Rest gebunden. Über eine sechste Koordinationstelle kann O₂ gebunden werden, welches dann wahrscheinlich auf das Substrat übertragen wird.^[195, 196] Im letzten Schritt wird der Aldehyd unter Abspaltung von Ameisensäure deformyliert und zurück bleibt eine Doppelbindung. Die Katalyse dieser C-C Bindungsspaltung ist allerdings komplizierter und zwei verschiedene Reaktionsmechanismen kommen in Frage. Entweder verläuft diese Reaktion radikalisch oder ionisch.^[197]

Ein anderes Enzym, das auch eine Deformylierungsreaktion katalysiert, ist die *Cyanobacterial Aldehyde Decarbonylase* (AD)^[198], die später als *Aldehyde-Deformylating Oxygenase* (ADO)^[199] bezeichnet wurde. Auch für diese Reaktion wird molekularer Sauerstoff benötigt. In der Abbildung 32 ist der postulierte Mechanismus gezeigt. Zuerst bindet der Sauerstoff (O₂) an den reduzierten Zustand des Sauerstoff-überbrückten Eisenclusters und generiert so ein Metall-gebundenes Peroxid-Nukleophil. Dieses greift das Substrat an der Carbonylgruppe an und initiiert die Spaltung der C-C Bindung. Es entsteht das korrespondierende Alkan und Formiat. Des Weiteren ist gezeigt, dass Aldehyde mit einer um ein C-Atom verkürzten Kettenlänge und Alkohole entstehen können. Anschließend wird der benötigte Eisencluster wieder reduziert und somit in die aktive Form überführt. Durch dieses Enzym werden die Aldehyde von Fettsäuren deformyliert und es entstehen die korrespondierenden Alkane und Formiat.^[199] Der Unterschied zu der möglichen Deformylierungsreaktion des 5-Formylcytidins ist der beteiligte Kohlenstoff, denn bei den beiden erwähnten Beispielen ist die Formylgruppe benachbart zu einem gesättigten Kohlenstoff. Im Fall des fdCs ist der Aldehyd an einen aromatischen Heterozyklus gebunden.^[193]

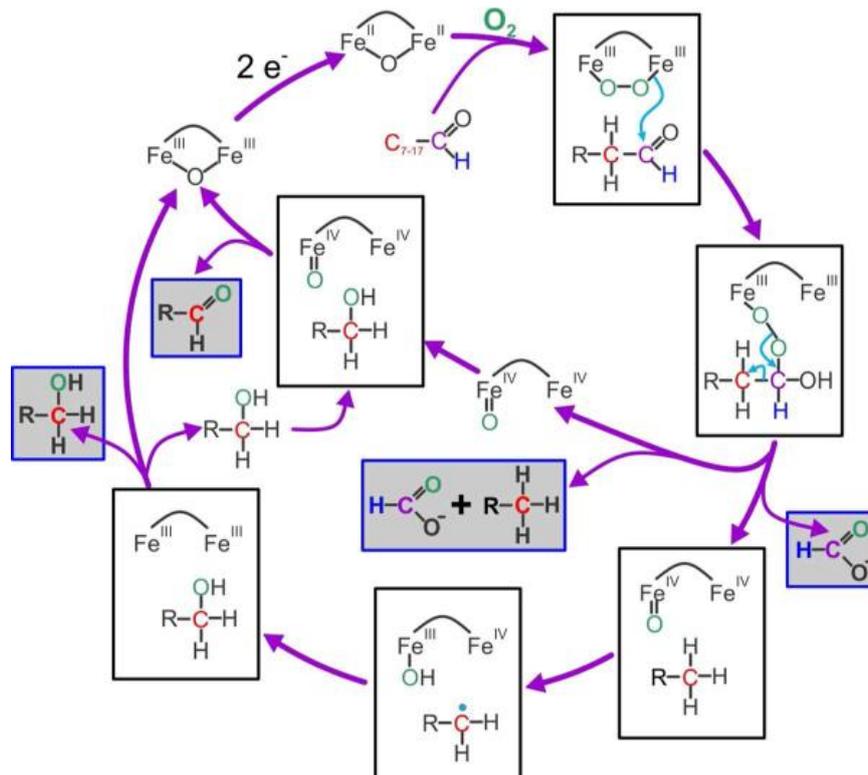


Abbildung 32: Postulierter Deformylierungsmechanismus des Enzyms *Aldehyde-Deformylating Oxygenase* (ADO). Die grau hinterlegten Strukturen sind bereits beobachtet worden. (Abbildung von *Aukema et al.*)^[199]

Substrate mit aromatischen Heterozyklen sind bisher aus Decarboxylierungsreaktionen bekannt. Diese Decarboxylierungen werden von den Enzymen Isoorotat Decarboxylase (IDCase)^[200, 201] und Orotidin 5'-Monophosphat Decarboxylase (ODCase)^[202] katalysiert. Die ODCase katalysiert zum Beispiel den letzten Schritt der *de novo* Pyrimidinbiosynthese, d.h. die Umwandlung von Orotidin 5'-Monophosphat (OMP) zu Uridin 5'-Monophosphat. Vor allem die Stabilisierung des Vinylianion Intermediats durch einen konservierten Lysin-Rest im katalytischen Zentrum ist ein wichtiger Faktor für die Katalyse.^[202] Das Enzym IDCase hingegen katalysiert in Pilzen die Decarboxylierung von 5-Carboxyluracil zu Uracil. Auch die Decarboxylierung von 5-Carboxylcytosin zu Cytosin konnte *in vitro* durch die IDCase katalysiert werden, allerdings mit einer schwächeren Aktivität.^[201] Nach dieser Entdeckung stellte sich die Frage, ob dieses Enzym eine mögliche Vorlage für eine potentielle 5-Carboxylcytidin Decarboxylase in DNA ist. Da die Ergebnisse der Veröffentlichung aus Abschnitt 3.4 einen C-C Bindungsbruch in Form einer Deformylierung unabhängig von einer Tet-vermittelten Weiteroxidation zeigen, ist nun die Frage, unter welchen katalytischen Bedingungen eine solche Reaktion *in vivo* abläuft.^[193]

4.3.2 Ergebnis und Diskussion

Auf der Suche nach einem Enzym, das die Formylgruppe der DNA-Modifikation fdC entfernen kann, wurden zuerst die Tet-Enzyme mittels eines Tet TKOs und die Dnmt-Enzyme mittels eines Dnmt TKOs untersucht. Diese Enzyme modifizieren die besagte C5-Position des Cytidins. Die Daten hierzu sind in der Veröffentlichung im Abschnitt 3.4 zu finden. Durch die Verfütterung des F-fdCs an Tet TKO Zellen fand man heraus, dass es sich bei der Entfernung der Formylgruppe um eine Tet-unabhängige Reaktion handelt. Der Einfluss anderer Faktoren ist allerdings nicht auszuschließen. F-fdC könnte zunächst zu F-cadC aufoxidiert werden und anschließend eine Decarboxylierungsreaktion erfolgen. Diese Ergebnisse zeigten jedoch, dass es sich bei dem gesuchten Enzym nicht um die Tet-Enzyme handelt. Auch die Verfütterung des F-fdCs an den Dnmt TKO zeigte die Entstehung des Deformylierungsprodukts F-dC; hier war allerdings erwartungsgemäß im Dnmt Triple Knockout kein F-mdC detektierbar. Es zeigten sich überraschenderweise jedoch Effekte in den Einzelknockouts, d.h. Dnmt1^{-/-}, Dnmt3a^{-/-} und 3b^{-/-} im Vergleich zum Dnmt TKO. Hierfür wurden diese vier Zelllinien und der dazu korrespondierende Wildtyp J1 mit F-fdC gefüttert. Nach der absoluten Quantifizierung der fluorierten Derivate in der genomischen DNA wurden zwei zusätzliche Normierungen eingeführt, um eine bessere Vergleichbarkeit der sehr unterschiedlichen Zelllinien zu erreichen. Zuerst wurden die erhaltenen F-dC Werte auf die F-fdC Werte normiert und im Anschluss wurden diese Ergebnisse auf den korrespondierenden Wildtyp J1 bezogen.

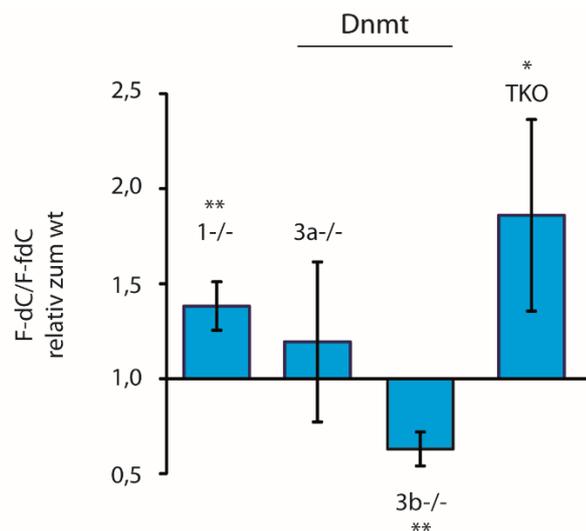


Abbildung 33: Normierte Quantifizierungsergebnisse (absolute Menge an F-dC pro F-fdC und relativ zum korrespondierenden Wildtyp J1) der Fütterung von F-fdC (350 μ M) an Dnmt1^{-/-}, Dnmt3a^{-/-}, Dnmt3b^{-/-} und Dnmt TKO und dem korrespondierenden Wildtyp J1. Technische Triplikate von jeweils drei unterschiedlichen Zellkulturen wurden gemessen.

Die Abbildung 33 zeigt die dadurch gewonnenen Daten. Der Dnmt1 KO und der Dnmt TKO liefern bezogen auf den Wildtypen ein größeres Verhältnis von F-dC zu F-fdC. Das bedeutet, dass in diesen KO Zellen mehr Deformylierung stattfindet als in den Wildtyp-Zellen. Dnmt1 ist die *maintenance* Methyltransferase und bindet als solche an die hemi-methylierte DNA in CpG-reichen Sequenzen. Unter hemi-methylierter DNA versteht man DNA, die nur an einem der beiden Stränge methyliert ist, wie kurz nach der Replikation. Der elterliche Strang liegt methyliert vor, während der neu gebildete noch unmethyliert ist. Die *maintenance* Methyltransferase Dnmt1 bindet an diese hemi-methylierten CpG-Stellen, methyliert diese und erhält so das Methylierungsmuster.^[203] Die *de novo* Methyltransferasen Dnmt3a und Dnmt3b binden auch an nicht methylierte DNA und sind essentiell für die frühe Entwicklung, da es nach der Nidation des Embryos zur genomweiten Methylierung kommt.^[204] Die Untersuchung der *de novo* Methyltransferase Dnmt3b hinsichtlich ihrer Deformylierungsaktivität lieferte ein interessantes Ergebnis. Im Gegensatz zu Dnmt1 führt der Dnmt3b KO zu einer Erniedrigung des Verhältnisses von F-dC zu F-fdC, d.h. die Deformylierung ist reduziert. Das deutet darauf deuten, dass die DNA Methyltransferase Dnmt3b einen Einfluss auf die Deformylierung hat. Wäre das Enzym Dnmt3b tatsächlich die Deformylase, so dürfte das Produkt der Deformylierungsreaktion, d.h. F-dC, nicht mehr detektierbar sein. Da es sich jedoch um eine Verringerung im Vergleich zum Wildtyp handelt, ist es eher wahrscheinlich, dass es sich hierbei um einen sekundären Effekt handelt. Darunter könnte man sich zum Beispiel eine Veränderung der DNA-Struktur durch ein anderes Methylierungsmuster oder Ähnliches vorstellen. Theoretisch ist auch denkbar, dass sich der Dnmt3b-/- bezüglich anderer Faktoren von den anderen Zelllinien unterscheidet und diese der Grund für eine verminderte Deformylierung sind. Die Daten des Dnmt3a-/- führen zu keinem signifikanten Ergebnis.

Im Folgenden sollten weitere Enzyme wie z.B. die Poly (ADP-Ribose) Polymerasen (PARPs) auf ihre Beteiligung an der Deformylierung, untersucht werden. Diese Enzyme sind dafür bekannt, den Transfer einer ADP-Ribose auf ein Ziel-Protein zu katalysieren.^[205-207] Zusätzlich dazu wurde herausgefunden, dass die PARP Enzyme eine wichtige Rolle in der DNA Reparatur spielen. Sie sind Teil des Basenexzisionskomplexes und reagieren auf Einzelstrangbrüche.^[208] Des Weiteren wurde gezeigt, dass PARPs auch eine Rolle in der Nukleotidexzisionsreparatur (NER) spielen.^[209] Da BER und NER die beiden Hauptwege der DNA Reparatur sind, geht mit der Inhibierung der PARP Enzyme eine genomische Instabilität einher. Die Inhibierung der PARP Enzyme führt außerdem zum Zellzyklusarrest der beschädigten Zellen.^[210] Ein weiterer Grund, warum die PARP Enzyme auf eine Aktivität als Deformylase untersucht werden sollen, ist dass die ADP-Ribosylierung der DNA durch die PARPs zu einer entspannten Chromatinstruktur führt.^[211] Durch diese entspannte Chromatinstruktur könnten die DNA Replikation, die Transkription und die Reparatur leichter vonstattengehen, da die beschädigte DNA leichter von den Reparaturenzymen erreicht

werden kann.^[211] Abgesehen von den Reparaturenzymen könnten auch Enzyme mit einer Deformylierungsaktivität besser binden. Im Folgenden sollte daher durch die Verwendung von zwei PARP Inhibitoren (Abb. 34) getestet werden, ob diese Einfluss auf die Deformylierungsreaktion haben. Hierfür wurde erneut F-fdC an mESCs gefüttert, welche zusätzlich mit jeweils einem der beiden PARP Inhibitoren (c (Olaparib) = 20 nM; c (Veliparib) = 1 μ M) versetzt wurden. Nach der massenspektrometrischen Analyse aller fluorierten Nukleosid-Analoga des Genoms stellte sich heraus, dass die Deformylierungsreaktion weiterhin und auch im gleichen Maße stattfinden konnte.

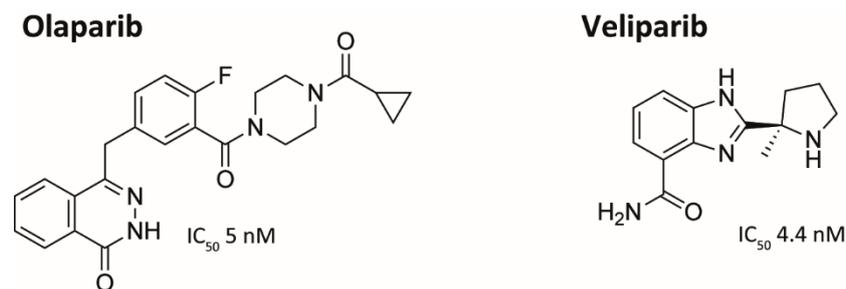


Abbildung 34: Schematische Darstellung der chemischen Strukturen der beiden PARP Inhibitoren Olaparib und Veliparib, sowie deren mittlere inhibitorische Konzentration.

Da die Inhibierung der PARP Enzyme zu keiner Reduktion der Deformylierungsreaktion geführt haben, ist nicht davon auszugehen, dass die PARP Enzyme die gesuchte Deformylase darstellen.

Eine andere Enzymklasse die als Deformylase vorstellbar ist, ist die der Histon Deacetylasen (HDAC). HDACs sind Enzyme, welche die Entfernung von Acetylgruppen von Lysin-Resten der Histone katalysieren. Diese Acetylierungen an Histonen sind sogenannte posttranslationale Modifikationen (PTMs) und führen zu einer funktionellen Veränderung der Genexpression oder der Chromatinstruktur. Acetylierungen an der ϵ -Aminogruppe der Lysin-Reste von Histonen werden durch Histon/Lysin Acetyltransferasen (HATs/KATs) eingeführt. Sie sind jedoch reversibel und können durch die HDACs wieder entfernt werden. In Menschen sind 18 HDAC Enzyme bekannt und in vier verschiedene Klassen unterteilt. Zu der Klasse I gehören HDAC1, HDAC2, HDAC3 und HDAC8. Die Klasse II besteht aus HDCA4, HDAC5, HDAC6, HDAC7, HDAC9 und HDAC10. Die Klasse III hingegen enthält Sirtuine (SIRT1, SIRT2, SIRT3, SIRT4, SIRT5, SIRT6 und SIRT7). HDAC11 bildet die Klasse IV. Jede Histon Deacetylase gehört also entweder zur Klasse der Histon Deacetylase Familie oder zu den Sir2 Regulatoren.^[212] Die Klasse III unterscheidet sich stark von den anderen Klassen, da diese NAD^+ als Reaktanden benutzt.^[213] Bei der Deacetylierung des Lysin-Rests entsteht neben dem deacetylierten Produkt somit Nicotinamid und 2'-O-Acetyl-ADP-Ribose.^[212] Die biologische Relevanz der HDACs ist vor allem dadurch gegeben, dass sie die Funktion, die durch die

Acetylierung der Histone entsteht, aufhebt und dadurch z.B. die Konformation des Chromatins verändert. Da HDACs auch eine große Rolle in der Entstehung von Krankheiten spielen, sind viele HDAC Inhibitoren kommerziell erhältlich. Zwei dieser Inhibitoren sind in der Abbildung 35 gezeigt. Trichostatin A (TSA) gehört zu der Gruppe der Inhibitoren, deren Struktur auf der Hydroxamsäure basiert, und Valproinsäure gehört zu den kurzkettigen Fettsäuren. Bisher ist bekannt, dass TSA zu den pan-HDAC Inhibitoren gehört und die Klassen I, II und IV inhibieren kann.^[212]

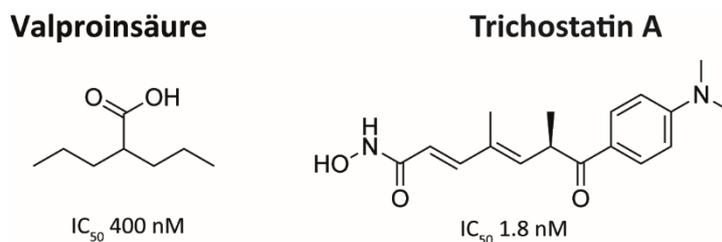


Abbildung 35: Schematische Darstellung der chemischen Strukturen der beiden HDAC Inhibitoren Valproinsäure und Trichostatin A, sowie die mittlere inhibitorische Konzentration.

Durch die Verwendung dieser beiden HDAC Inhibitoren sollten die HDACs als potentielle Deformylase getestet werden. Auch hierfür wurden die Zellen sowohl mit Valproinsäure (c (Valproinsäure) = 0,5 mM) bzw. TSA (c (TSA) = 10 nM) behandelt und gleichzeitig mit 350 µM F-fdC versetzt. Nach 3 d wurden die Zellen geerntet, die DNA isoliert und anschließend massenspektrometrisch auf die fluorierten Nukleoside untersucht. Wie auch schon bei der Analyse der PARP Enzyme ergaben sich keine Unterschiede im Grad der Deformylierung, d.h. die Menge des deformylierten Produkts F-dC, blieb gleich. Die Klassen I, II und IV der HDAC Enzyme sind somit wahrscheinlich nicht die Enzyme, die eine Deformylierungsreaktion an F-fdC katalysieren können.

Im Folgenden sollte deshalb die Klasse III der HDACs mit Sirtuin Inhibitoren untersucht werden. Die Abbildung 36 zeigt drei verschiedene Sirtuin Inhibitoren, wobei EX-527 (c (EX-527) = 50 µM) ein kommerziell erhältlicher Sirt1 Inhibitor, SirReal2 (c (SirReal2) = 10 µM) ein Sirt2 Inhibitor und Nicotinamid (NAM) (c (NAM) = 10 mM) ein Sirtuin-Breitband Inhibitor ist. Um zu testen, ob die Sirtuine im Allgemeinen einen Effekt auf die Deformylierungsreaktion haben, wurde erneut F-fdC an mESCs gefüttert, die zusätzlich noch mit einem der drei Sirtuin Inhibitoren versetzt wurden.

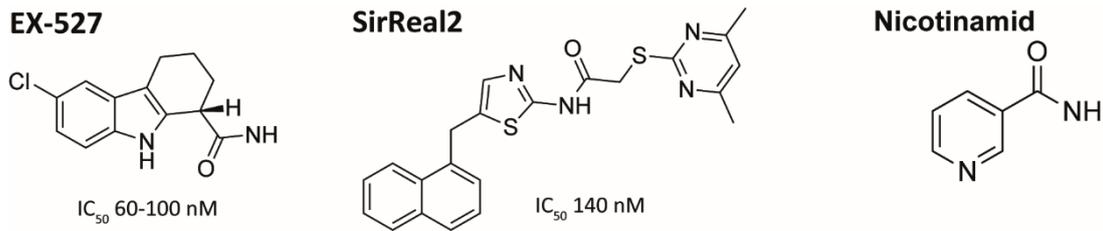


Abbildung 36: Schematische Darstellung der chemischen Strukturen verschiedener Sirtuin (Sirt) Inhibitoren: EX-527 ist ein Sirt1 Inhibitor, SirReal2 ein Sirt2 Inhibitor und Nicotinamid (NAM) ein Sirtuin-Breitband Inhibitor, sowie deren mittlere inhibitorische Konzentration.

Nach 3 d Inkubation mit F-fdC und dem jeweiligen Inhibitor wurden die Zellen geerntet, DNA isoliert und massenspektrometrisch auf die enthaltenen Fluoro-Nukleoside analysiert. Wie in der Abbildung 37 ersichtlich, führt die Behandlung mit dem Breitband-Inhibitor Nicotinamid zu einer signifikanten Verminderung der Deformylierung um 24 %. Um die Daten aus den einzelnen biologischen Replikaten, bestehend aus technischen Triplikaten, besser vergleichbar zu machen, wurden zwei Normierungen durchgeführt. Zum einen wurde die quantifizierte Mengen des entstandenen F-dCs auf die quantifizierte Menge des Vorläufermoleküls F-fdC bezogen und zum anderen wurde anschließend das Verhältnis von F-dC zu F-fdC der Inhibitor-behandelten Probe auf die der Negativkontrolle normiert. Diese Normierungen zeigen direkt den Unterschied in der Deformylierung.

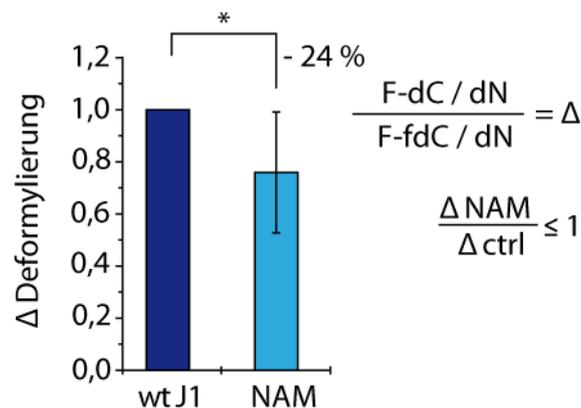


Abbildung 37: Normierte Quantifizierungsergebnisse (absolute Menge an F-dC pro F-fdC und relativ zur korrespondierenden Negativkontrolle) der Fütterung von F-fdC (350 μM) an Wildtyp J1 Zellen die mit Nicotinamid behandelt wurden und die korrespondierende Negativkontrolle. Technische Triplikate von jeweils vier unterschiedlichen Zellkulturen wurden gemessen.

Die Verwendung von Inhibitoren ist nie ganz einfach, denn ohne weitere Analyse lässt sich der Grad der Inhibition nicht bestimmen. Deshalb könnte man bei einer Wiederholung des Experiments auch ein zusätzliches Kit zur Bestimmung der NAD⁺ Mengen in der Zelle bestimmen. So könnte man in jedem einzelnen Experiment den Grad der Inhibition bestimmen und auf diesen normieren, das würde zu verlässlicheren Werten mit kleinerem Fehler führen.

Trotzdem ist eine Reduktion der Deformylierungsreaktion durch die Behandlung mit Nicotinamid deutlich sichtbar (Abb. 37). Die Sirtuine könnten also einen deformylierenden Effekt aufweisen. Dieses Ergebnis ist nicht mehr als ein Hinweis, denn Nicotinamid könnte auch andere Enzyme und Abläufe in der Zelle beeinflussen.

Nach diesem ersten Hinweis auf eine mögliche Beteiligung der Sirtuine an der Deformylierungsreaktion sollte dieses Experiment im Folgenden mit den beiden anderen Sirtuin Inhibitoren Ex-527 und SirReal2 wiederholt werden. Die massenspektrometrische Analyse der Zellen, die mit dem Sirtuin 1 Inhibitor versetzt wurden, ergab keine signifikanten Ergebnisse, da es sehr große Unterschiede zwischen den einzelnen biologischen Replikaten gab. Hier wäre es unbedingt notwendig gewesen, die tatsächliche Inhibition des Sirt1 in jeder Probe zu bestimmen und darauf zu normieren. Ein ähnliches Ergebnis zeigt sich bei der Inhibition des Sirt2. Die Messung von zwei biologischen Replikaten (Rep1 und Rep2) lieferte sehr unterschiedliche Mengen an F-dC pro F-fdC. Auch für dieses Experiment wurden die zwei genannten Normierungen durchgeführt. Wie Abb. 38 zeigt, kommt es durch die Verwendung des Sirt2 Inhibitors (SR2) im ersten biologischen Replikat zu einer Verminderung der Deformylierung um 20 % und im Zweiten um 75 %. Das könnte darauf hindeuten, dass der Effekt des Inhibitors nicht immer gleich stark ausgeprägt ist, aber die generelle Tendenz eine Reduktion der Deformylierung durch den Sirt2 Inhibitor zeigt.

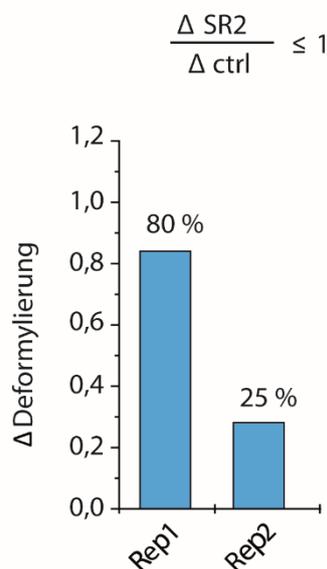


Abbildung 38: Normierte Quantifizierungsergebnisse (absolute Menge an F-dC pro F-fdC und relativ zur korrespondierenden Negativkontrolle) der Fütterung von F-fdC (350 μ M) an Wildtyp J1 Zellen die mit SirReal2 (SR2) behandelt wurden. Technische Triplikate von jeweils einer Zellkultur wurden gemessen.

Da diese Experimente gezeigt haben, dass die Verwendung von Inhibitoren in diesem Fall nicht zu gleichbleibenden Ergebnissen führt bzw. es unbedingt einer zweiten Analyse zur Kontrolle der tatsächlichen Inhibition in jeder Probe bedarf, wurde im Folgenden eine Knockout-Zelllinie verwendet. Obwohl die Inhibition durch den Sirt1 Inhibitor Ex-527 nicht zu einer signifikanten Verminderung der Deformylierung geführt hat, wurde das folgende Experiment mit einer Sirt1 Knockout-Zelllinie durchgeführt. Wie zuvor wurde F-fdC an diese Knockout-Zelllinie und ihren korrespondierenden Wildtyp R1 gefüttert, für 3 d kultiviert und die Zellen geerntet. Nach der Isolation der DNA wurden die Proben verdaut und massenspektrometrisch auf die Zusammensetzung der genomischen DNA analysiert, wobei das Hauptaugenmerk auf der Quantifizierung der Fluoro-Nukleoside lag. Erneut wurden die Daten zweifach normiert, wodurch deutlich wird, dass im KO 60 % weniger Deformylierung stattfindet als im korrespondierenden Wildtyp (Abb. 39). Das ist ein erster Hinweis darauf, dass Sirtuin 1 einen Einfluss auf die Deformylierung hat.

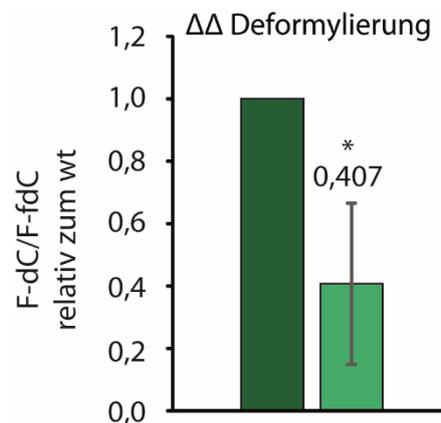


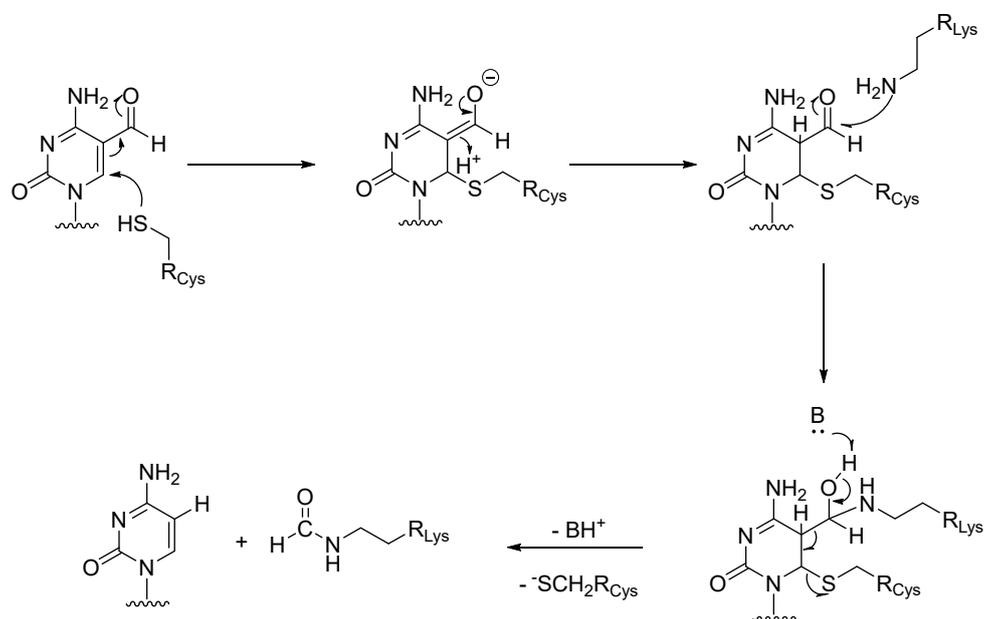
Abbildung 39: Normierte Quantifizierungsergebnisse (absolute Menge an F-dC pro F-fdC und relativ zum korrespondierenden Wildtyp R1) der Fütterung von F-fdC (350 μ M) an Sirt1 Knockout Zellen und den korrespondierenden Wildtyp R1. Technische Triplikate von jeweils vier unterschiedlichen Zellkulturen wurden gemessen.

Wäre Sirtuin 1 tatsächlich die einzige Deformylase, so dürfte das Produkt (F-dC) gar nicht mehr detektierbar sein und nicht nur in reduzierter Menge vorliegen. Es wäre denkbar, dass es sich nur um einen sekundären Effekt handelt. Andererseits wäre es auch möglich, dass durch die Deletion des Sirtuin 1 andere Sirtuin Enzyme dessen Aufgabe in der Zelle übernehmen und F-fdC deformylieren können. Deswegen sollte in Zukunft der Sirt1 KO verwendet werden und eine zusätzliche Inhibition von Sirt2 durch den Inhibitor SirReal2 getestet werden. Es wäre interessant zu sehen, ob die Deformylierungsaktivität im Vergleich zum Sirt1 KO alleine noch weiter absinkt.

In einem zukünftigen Experiment sollten die Acetylierungslevel der Histone in den Sirt1 Knockout Zellen und dem korrespondierenden Wildtyp bestimmt werden. Diese Ergebnisse

könnte man verwenden, um die erhaltenen Quantifizierungsdaten der DNA darauf zu beziehen. Die tatsächlichen Acetylierungslevel könnten theoretisch auch erklären warum die Reduktion der Deformylierung um 60 %, aber eben nicht um 100 %, stattfindet. Damit ist gemeint, dass in dem Sirt1 KO auch nicht zwangsläufig alle in Frage kommenden Histone acetyliert vorliegen sondern auch nur zu einem bestimmten Prozentsatz.

Wenn der Knockout des Sirtuin 1 bzw. die Inhibierung möglicherweise alle Sirtuine durch den Breitband-Inhibitor Nicotinamid zu einer Verminderung der Deformylierung führen, dann könnte das auch daran liegen, dass spezielle Lysine nicht mehr deacetyliert werden und kein freies, reaktives Amin tragen. Hier könnte man die Hypothese aufstellen, dass nur das freie Amin des Lysins von Histonen nukleophil an der Carbonylgruppe des Aldehyds angreifen kann. Ein möglicher Mechanismus ist im folgenden Schema gezeigt.



Schema 8: Möglicher Mechanismus der Deformylierungsreaktion von fdC durch die freien Amine der Lysin Reste von Histonen.

Nicht nur das freie Amin des Lysin-Rests, sondern auch eine Methylierung zum Methyllysin wäre als angreifendes Nucleophil denkbar.

4.3.3 Zusammenfassung der Projektbeiträge

Das F-fdC wurde von *René Rahimoff* und *Eva Korytiakovà* synthetisiert. Die Zellkulturversuche wurden von *Angie Kirchner* und *Fabio Spada* durchgeführt und entwickelt. Auch die Isolation aller DNA Proben wurde von *Angie Kirchner* vorgenommen. Der enzymatische Verdau wurde von mir mit Hilfe von *Kerstin Kurz* durchgeführt. Die anschließende Messung und Auswertung der Daten wurden von mir vorgenommen. Die Daten wurden gemeinsam von *Angie Kirchner* und mir interpretiert.

4.3.4 Material und Methoden

4.3.4.1 Zelllinien

Die verwendeten R1-mESCs^[214] (embryonalen Stammzellen der Maus, *murine Embryonic Stem Cells*) als auch die Sirt1-/- mESCs^[215] wurden von Michael W. McBurney (University of Ottawa) bereitgestellt. Die Dnmt1-/- Zelllinie^[216], sowie die Dnmt3a-/- und Dnmt3b-/- mESCs^[204] wurden von En Li (*Novartis Institutes for Biomedical Research, Shanghai*) zur Verfügung gestellt. Die verwendeten J1-mESCs (embryonalen Stammzellen der Maus, *murine Embryonic Stem Cells*) entstammen der Zelllinie 129/Sv/Jae.

4.3.4.2 Enzymatischer Verdau

Der enzymatische Verdau der genomischen DNA aus den Fütterungsexperimenten mit F-fdC erfolgte wie in dem in Abschnitt 3.4 beschriebenen Methodenteil.

4.3.4.3 UHPLC-MS/MS Analyse

Die UHPLC-MS/MS Analyse erfolgte wie in dem in Abschnitt 3.4 beschriebenen Protokoll.

4.3.4.4 Auswertung

Die Auswertung der Daten erfolgte wie in dem in Abschnitt 3.1 beschriebenen Protokoll. Die Darstellung der Daten erfolgte mit Excel und Adobe Illustrator.

5 Abkürzungsverzeichnis

(R)	<i>ribo</i>
(S)	<i>arabino</i>
-/-	Knockout beider Allele eines Enzyms
+/-	Knockout eines der beiden Allele eines Enzyms
°	Grad
8oxodG	8-Oxo-2'-desoxyguanosin
A	Adenin/Adenosin
A	<i>area under the curve</i>
ADAR	<i>Adenosine deaminases acting on RNA</i>
ALKBH5	<i>RNA demethylase ALKBH5</i>
amu	<i>atomic mass unit</i>
ATP	Adenosintriphosphat
BER	Basenexzisionsreparatur
bp	Basenpaar
C	Cytosin/Cytidin
C	Celsius
cadC	5-Carboxyl-2'-desoxycytidin
Cda	Cytosin-Desaminase
CDK5	<i>cyclin-dependent protein kinase 5</i>
CDK5RAP1	<i>CDK5 regulatory subunit associated protein 1</i>
cDNA	<i>complementary DNA</i>
CE	<i>collision energy</i>
CHCl ₃	Chloroform
Cm	2'-O-Methylcytidin
CpG	dC-dG-Dinukleotid
dA	2'-Desoxyadenosin
dC	2'-Desoxycytidin
dCDP	2'-Desoxycytidin-5'-diphosphat
dCMP	2'-Desoxycytidin-5'-monophosphat
Dctd	Desoxycytidilat-Desaminase
dG	2'-Desoxyguanosin
DMAPP	Dimethylallyl-diphosphat
dN	2'-Desoxynukleosid
DNA	<i>deoxyribonucleic acid</i> ; Desoxyribonukleinsäure

Dnmt	DNA-Methyltransferase
dT	2'-Desoxythymidin
dTTP	2'-Desoxycytidin-5'-triphosphat
dU	2'-Desoxyuridin
dUMP	2'-Desoxyuridin-5'-monophosphat
<i>E.coli</i>	<i>Escherichia Coli</i>
EDTA	Ethylendiamintetraessigsäure
eIF3	<i>eukaryotic initiation factor 3</i>
eIF4E	<i>eukaryotic translation initiation factor 4E</i>
EMSA	<i>Electrophoretic Mobility Shift Assay</i>
ESI	Elektrospray-Ionisierung
FAM120C	<i>constitutive coactivator pf PPAR-gamma-like protein 2</i>
FBS	<i>Fetal bovine serum</i>
fdC	5-Formyl-2'-desoxycytidin
fdU	5-Formyl-2'-desoxyuridin
FTO	<i>Fat Mass and Obesity-associated Protein</i>
G	Guanin/Guanosin
GFP	Grün fluoreszierendes Protein
GTC	Guanidinisothiocyanat
HAP1	<i>near haploid cell line</i>
HAT	Histon-Acetyltransferase
HDAC	Histon-Deacetylase
HEK	<i>human embryoid kidney</i>
HeLa	humane Zervix Adenokarzinom Zelllinie
hmdC	5-Hydroxymethyl-2'-desoxycytidin
hmdU	5-Hydroxymethyl-2'-desoxyuridin
HNRNPA2B1	<i>heterogeneous nuclear ribonucleoprotein A2/B1</i>
HPLC	<i>high performance liquid chromatography</i>
HRMS	<i>high resolution Massenspektrometrie</i>
I	Inosin
i ⁶ A	<i>N⁶-Isopentenyladenosin</i>
IDCase	Isoorotat Decarboxylase
io ⁶ A	<i>N⁶-(4-Hydroxyisopentenyl)adenosin</i>
KAT	Lysin-Acetyltransferase
KBM-7	<i>chronic myelogenous leukemia cell line</i>
KIAA1429	<i>vir-like m⁶A methyltransferase associated</i>
KO	<i>Knockout</i>

LC	<i>liquid chromatography</i>
LIF	<i>Leukemia Inhibitory Factor</i>
LLOQ	<i>lower limit of quantification</i>
lncRNA	<i>long non-coding RNA</i>
LOD	<i>limit of detection</i>
M.SssI	CpG Methyltransferase
m ¹ A	<i>N¹-Methyladenosin</i>
m ² ₂ G	<i>N²,N²-Dimethylguanosin</i>
m ⁵ C	<i>5-Methylcytidin</i>
m ⁶ ₂ A	<i>N⁶,N⁶-Dimethyladenosin</i>
m ⁶ A	<i>N⁶-Methyladenosin</i>
m ⁶ Am	<i>N⁶,2'-O-diemthyladenosin</i>
m ⁷ G	<i>N⁷-Methylguanosin</i>
mdC	<i>5-Methyl-2'-desoxycytidin</i>
MeCN	Acetonitril
MEF	<i>mouse embryonic fibroblast</i>
MELAS	<i>mitochondrial encephalomyopathy, lactic adicosis, and stroke-like episodes</i>
MeOH	Methanol
MERRF	<i>myoclonic epilepsy with ragged red fibers</i>
mESC	<i>murine embryonic stem cells</i>
METTL14	<i>N⁶-Adenosine methyltransferase non-catalytic subunit 14</i>
METTL3	<i>N⁶-Adenosine methyltransferase non-catalytic subunit 3</i>
MiaA	<i>Methylthio-isopentenyl-adenosine synthesis enzyme A</i>
MiaB	<i>Methylthio-isopentenyl-adenosine synthesis enzyme B</i>
miCLIP	<i>m⁶A individual-nucleotide-resolution crosslinking and immunoprecipitation</i>
miRNA	<i>micro RNA</i>
mRNA	<i>messenger RNA</i>
MS	Massenspektrometrie/Massenspektrometer
MS/MS	Tandemmassenspektrometrie
ms ² _i ⁶ A	<i>N⁶-Isopentenyl-2-thiomethyladenosin</i>
mt-RNA	<i>mitochondrial RNA</i>
NAM	Nicotinamid
Nck	<i>neuronal Cdc2-like Kinase</i>
NMR	<i>nuclear magnetic resonance</i>
nt	<i>nucleotides</i>

Oct4	Octamer-bindender Transkriptionsfaktor 4
ODCase	Orotidin 5'-Monophosphat Decarboxylase
OMP	Orotidin-5'-monophosphat
PARP	Poly (ADP) Polymerasen
P-bodies	<i>processing bodies</i>
PBS	<i>Phosphate-buffered saline</i>
PCR	<i>Polymerase-Kettenreaktion</i>
piRNA	<i>piwi-interacting RNA</i>
PTC	<i>premature termination codon</i>
PTM	posttranslationale Modifikationen
q	<i>quantitative</i>
R	<i>ratio</i> , Verhältnis
RNA	<i>Ribonucleic Acid</i> , Ribonukleinsäure
ROS	<i>reactive oxygen species</i>
rRNA	ribosomale RNA
S	<i>Svedberg units</i>
SAM	S-Adenosylmethionin
SILAC	<i>stable isotope labeling by amino acids in cell culture</i>
siRNA	<i>small-interfering RNA</i>
Sirt	Sirtuin
snoRNA	<i>small nucleolar RNA</i>
SNP	<i>single nucleotide polymorphism</i>
snRNA	<i>small nuclear RNA</i>
ssRNA	<i>single stranded RNA</i>
T	Thymin/Thymidin
t ⁶ A	N ⁶ -Threonylcarbamoyladenosin
TAE	TRIS-Acetat-EDTA
Tdg/TDG	Thymin-DNA-Glykosylase
Tet	<i>Ten-eleven translocation</i> Enzym
TKO	<i>triple knockout</i>
TRIT1	tRNA isopentenyl transferase 1
Ts	Thymidilat Synthase
TSA	Trichostatin A
U	Uracil
U2AF2	<i>U2 small nuclear RNA auxiliary factor 2</i>
UHPLC	<i>ultrahigh performance liquid chromatography</i>
ULOQ	<i>upper limit of quantification</i>

UTR	<i>untranslated region</i>
UV	Ultraviolettstrahlung
v/v	Volumenanteil
WTAP	<i>Pre-mRNA-splicing regulator WTAP</i>
XIST	<i>X-inactive specific transcript</i>
YBX1	<i>Y-box binding protein 1</i>
yW	Wybutosin
Ψ	Pseudouridin

6 Literaturverzeichnis

- [1] P. A. Levene, The Structure of Yeast Nucleic Acid. *J. Biol. Chem.* **1919**, *40*, 415–424.
- [2] E. Chargaff, Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **1950**, *6*, 201-209.
- [3] J. D. Watson, F. H. C. Crick, Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737-738.
- [4] M. H. F. Wilkins, et al., Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **1953**, *171*, 738-740.
- [5] O. T. Avery, et al., Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III *J. Exp. Med.* **1944**, *79*, 137-158.
- [6] A. D. Hershey, M. Chase, Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage. *J. Gen. Physiol.* **1952**, *36*, 39-56.
- [7] J. D. Watson, F. H. C. Crick, Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **1953**, *171*, 964-967.
- [8] M. Meselson, F. W. Stahl, The Replication of DNA in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 671-682.
- [9] F. H. C. Crick, On protein synthesis. *Symp. Soc. Exp. Biol.* **1958**, *12*, 138-163.
- [10] F. Crick, Central Dogma of Molecular Biology. *Nature* **1970**, *227*, 561-563.
- [11] R. Holliday, Epigenetics: An overview. *Dev. Genet.* **1994**, *15*, 453-457.
- [12] R. Holliday, Epigenetics: A Historical Overview. *Epigenetics* **2006**, *1*, 76-80.
- [13] A. Bird, Perceptions of epigenetics. *Nature* **2007**, *447*, 396-398.
- [14] G. Egger, et al., Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **2004**, *429*, 457-463.
- [15] O. J. Rando, K. J. Verstrepen, Timescales of Genetic and Epigenetic Inheritance. *Cell*, *128*, 655-668.
- [16] R. Wing, et al., Crystal structure analysis of a complete turn of B-DNA. *Nature* **1980**, *287*, 755-758.
- [17] G. R. Wyatt, Occurrence of 5-Methyl-Cytosine in Nucleic Acids. *Nature* **1950**, *166*, 237-238.
- [18] R. Bonasio, et al., Molecular Signals of Epigenetic States. *Science* **2010**, *330*, 612-616.
- [19] H. Cedar, Y. Bergman, Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **2009**, *10*, 295-304.
- [20] Z. D. Smith, A. Meissner, DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **2013**, *14*, 204-220.
- [21] T. B. Johnson, R. D. Coghill, Researches on Pyrimidines. C111. The Discovery of 5-Methyl-Cytosine in Tuberculinic Acid, the Nucleic Acid of the Tubercle Bacillus *J. Am. Chem. Soc.* **1925**, *47*, 2838-2844.
- [22] P. A. Jones, D. Takai, The Role of DNA Methylation in Mammalian Epigenetics. *Science* **2001**, *293*, 1068-1070.
- [23] J. A. Law, S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **2010**, *11*, 204-220.
- [24] E. Baxter, et al., Epigenetic regulation in cancer progression. *Cell Biosci.* **2014**, *4*, 45.
- [25] R. J. Klose, A. P. Bird, Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **2006**, *31*, 89-97.
- [26] R. Jaenisch, A. Bird, Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **2003**, *33*, 245-254.
- [27] G. R. Wyatt, S. S. Cohen, A New Pyrimidine Base from Bacteriophage Nucleic Acids. *Nature* **1952**, *170*, 1072-1073.
- [28] G. R. Wyatt, S. S. Cohen, The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochem. J.* **1953**, *55*, 774-782.

- [29] T. Pfaffeneder, et al., The Discovery of 5-Formylcytosine in Embryonic Stem Cell DNA. *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012.
- [30] Y.-F. He, et al., Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* **2011**, *333*, 1303-1307.
- [31] S. Ito, et al., Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011**, *333*, 1300-1303.
- [32] T. Pfaffeneder, et al., Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **2014**, *10*, 574-581.
- [33] T. Carell, et al., Structure and Function of Noncanonical Nucleobases. *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131.
- [34] Cornelia G. Spruijt, et al., Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives. *Cell*, *152*, 1146-1159.
- [35] J. Du, et al., DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 519-532.
- [36] R. K. McGinty, S. Tan, Nucleosome Structure and Function. *Chem. Rev.* **2015**, *115*, 2255-2273.
- [37] J.-S. Lee, et al., The Language of Histone Crosstalk. *Cell* **2010**, *142*, 682-685.
- [38] T. Jenuwein, C. D. Allis, Translating the Histone Code. *Science* **2001**, *293*, 1074-1080.
- [39] A. J. Bannister, T. Kouzarides, Regulation of chromatin by histone modifications. *Cell Res.* **2011**, *21*, 381-395.
- [40] K. E. Gardner, et al., OPERating ON Chromatin, a Colorful Language where Context Matters. *J. Mol. Biol.* **2011**, *409*, 36-46.
- [41] H. Wu, Y. Zhang, Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell* **2014**, *156*, 45-68.
- [42] C. Jiang, B. F. Pugh, Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **2009**, *10*, 161-172.
- [43] H. Huang, et al., Quantitative Proteomic Analysis of Histone Modifications. *Chem. Rev.* **2015**, *115*, 2376-2418.
- [44] J. Wu, M. Grunstein, 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* **2000**, *25*, 619-623.
- [45] A. F. Kebede, et al., Novel types and sites of histone modifications emerge as players in the transcriptional regulation contest. *FEBS J.* **2015**, *282*, 1658-1674.
- [46] C. He, Grand Challenge Commentary: RNA epigenetics? *Nat. Chem. Biol.* **2010**, *6*, 863-865.
- [47] I. A. Roundtree, C. He, RNA epigenetics—chemical messages for posttranscriptional gene regulation. *Curr. Opin. Chem. Biol.* **2016**, *30*, 46-51.
- [48] T. Sibbritt, et al., Mapping and significance of the mRNA methylome. *WIREs RNA* **2013**, *4*, 397-422.
- [49] X. Wang, C. He, Dynamic RNA Modifications in Posttranscriptional Regulation. *Mol. Cell* **2014**, *56*, 5-12.
- [50] W. A. Cantara, et al., The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.* **2011**, *39*, D195-D201.
- [51] M. Guttman, J. L. Rinn, Modular regulatory principles of large non-coding RNAs. *Nature* **2012**, *482*, 339.
- [52] L. He, G. J. Hannon, MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **2004**, *5*, 522.
- [53] T. R. Cech, The efficiency and versatility of catalytic RNA: implications for an RNA world. *Gene* **1993**, *135*, 33-36.
- [54] S. R. Eddy, Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2001**, *2*, 919.
- [55] M. E. Dinger, et al., Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **2008**, *18*, 1433-1445.
- [56] P. Boccaletto, et al., MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **2018**, *46*, D303-D307.

- [57] P. A. Limbach, et al., Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* **1994**, *22*, 2183-2196.
- [58] M. Sprinzl, K. S. Vassilenko, Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **2005**, *33*, D139-D140.
- [59] E. M. Phizicky, J. D. Alfonzo, Do all modifications benefit all tRNAs? *FEBS Lett.* **2010**, *584*, 265-271.
- [60] H. Laten, et al., Isopentenyladenosine deficient tRNA from an antisuppressor mutant of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **1978**, *5*, 4329-4342.
- [61] M. E. Dihanich, et al., Isolation and characterization of MOD5, a gene required for isopentenylation of cytoplasmic and mitochondrial tRNAs of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **1987**, *7*, 177-184.
- [62] P. F. Agris, et al., tRNA's Wobble Decoding of the Genome: 40 Years of Modification. *J. Mol. Biol.* **2007**, *366*, 1-13.
- [63] H. Grosjean, et al., Posttranscriptionally modified nucleosides in transfer RNA: Their locations and frequencies. *Biochimie* **1995**, *77*, 139-141.
- [64] M. Charette, M. W. Gray, Pseudouridine in RNA: What, Where, How, and Why. *IUBMB Life* **2000**, *49*, 341-351.
- [65] D. Globisch, et al., Systems-Based Analysis of Modified tRNA Bases. *Angew. Chem. Int. Ed.* **2011**, *50*, 9739-9742.
- [66] V. Reiter, Synthese von eukaryotischen RNA-Modifikationen und Quantifizierung nicht kanonischer Nucleoside sowie Untersuchungen zu deren Biosynthese *Dissertation Ludwig-Maximilian-Universität München* **2013**.
- [67] T. Brückl, Synthese natürlicher und isotoopenmarkierten tRNA Nucleoside und deren massenspektrometrische Quantifizierung *Dissertation Ludwig-Maximilian-Universität München* **2010**.
- [68] A. Czerwoniec, et al., MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res.* **2009**, *37*, D118-D121.
- [69] C. T. Y. Chan, et al., A Quantitative Systems Approach Reveals Dynamic Control of tRNA Modifications during Cellular Stress. *PLOS Genet.* **2010**, *6*, e1001247.
- [70] K. H. Kaminska, et al., Structural bioinformatics analysis of enzymes involved in the biosynthesis pathway of the hypermodified nucleoside ms2io6A37 in tRNA. *Proteins* **2008**, *70*, 1-18.
- [71] J. Caillet, L. Droogmans, Molecular cloning of the *Escherichia coli* miaA gene involved in the formation of delta 2-isopentenyl adenosine in tRNA. *J. Bacteriol.* **1988**, *170*, 4147-4152.
- [72] J. A. Moore, C. D. Poulter, *Escherichia coli* Dimethylallyl Diphosphate:tRNA Dimethylallyltransferase: A Binding Mechanism for Recombinant Enzyme. *Biochemistry* **1997**, *36*, 604-614.
- [73] B. Esberg, et al., Identification of the miaB Gene, Involved in Methylthiolation of Isopentenylated A37 Derivatives in the tRNA of *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.* **1999**, *181*, 7256-7265.
- [74] H. L. Hernández, et al., MiaB, a Bifunctional Radical-S-Adenosylmethionine Enzyme Involved in the Thiolation and Methylation of tRNA, Contains Two Essential [4Fe-4S] Clusters. *Biochemistry* **2007**, *46*, 5140-5147.
- [75] F. Pierrel, et al., MiaB Protein from *Thermotoga maritima*: Characterization of an extremely thermophilic tRNA-Methylthiotransferase *J. Biol. Chem.* **2003**, *278*, 29515-29524.
- [76] F. Pierrel, et al., Enzymatic Modification of tRNAs: MiaB is an Iron-Sulfur Protein. *J. Biol. Chem.* **2002**, *277*, 13367-13370.
- [77] F. Forouhar, et al., Two Fe-S clusters catalyze sulfur insertion by radical-SAM methylthiotransferases. *Nat. Chem. Biol.* **2013**, *9*, 333.
- [78] F. Pierrel, et al., MiaB Protein Is a Bifunctional Radical-S-Adenosylmethionine Enzyme Involved in Thiolation and Methylation of tRNA. *J. Biol. Chem.* **2004**, *279*, 47555-47563.

- [79] S. Kellner, et al., Profiling of RNA modifications by multiplexed stable isotope labelling. *Chem. Commun.* **2014**, *50*, 3516-3518.
- [80] S. Kellner, et al., Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **2014**, *42*, e142-e142.
- [81] A. Golovko, et al., Cloning of a human tRNA isopentenyl transferase. *Gene* **2000**, *258*, 85-93.
- [82] V. Reiter, et al., The CDK5 repressor CDK5RAP1 is a methylthiotransferase acting on nuclear and mitochondrial RNA. *Nucleic Acids Res.* **2012**, *40*, 6235-6240.
- [83] F.-Y. Wei, et al., Cdk5-dependent regulation of glucose-stimulated insulin secretion. *Nat. Med.* **2005**, *11*, 1104.
- [84] Y.-P. Ching, et al., Identification of a Neuronal Cdk5 Activator-binding Protein as Cdk5 Inhibitor. *J. Biol. Chem.* **2002**, *277*, 15237-15240.
- [85] G. R. Björk, et al., A primordial tRNA modification required for the evolution of life? *EMBO J.* **2001**, *20*, 231-239.
- [86] J. Rozenski, et al., The RNA Modification Database: 1999 update. *Nucleic Acids Res.* **1999**, *27*, 196-197.
- [87] S. Dunin-Horkawicz, et al., MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res.* **2006**, *34*, D145-D149.
- [88] I. Díaz, M. Ehrenberg, ms2i6A deficiency enhances proofreading in translation. *J. Mol. Biol.* **1991**, *222*, 1161-1171.
- [89] M. L. Gefter, R. L. Russell, Role of modifications in tyrosine transfer RNA: A modified base affecting ribosome binding. *J. Mol. Biol.* **1969**, *39*, 145-157.
- [90] S. S. Ashraf, et al., The uridine in "U-turn": contributions to tRNA-ribosomal binding. *RNA* **1999**, *5*, 503-511.
- [91] J. Cabello-Villegas, et al., Solution Conformations of Unmodified and A37N6-dimethylallyl Modified Anticodon Stem-loops of *Escherichia coli* tRNAPhe. *J. Mol. Biol.* **2002**, *319*, 1015-1034.
- [92] F. V. Murphy IV, et al., The role of modifications in codon discrimination by tRNA^{Lys}UUU. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186.
- [93] R. K. Wilson, B. A. Roe, Presence of the hypermodified nucleotide N⁶-(Δ²-isopentenyl)-2-methylthioadenosine prevents codon misreading by *Escherichia coli* phenylalanyl-transfer RNA. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 409-413.
- [94] L. B. Jenner, et al., Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.* **2010**, *17*, 555.
- [95] F.-Y. Wei, et al., Cdk5rap1-Mediated 2-Methylthio Modification of Mitochondrial tRNAs Governs Protein Translation and Contributes to Myopathy in Mice and Humans. *Cell Metab.* **2015**, *21*, 428-442.
- [96] T. Suzuki, et al., Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases. *EMBO J.* **2002**, *21*, 6581-6589.
- [97] Y. Kirino, T. Suzuki, Human Mitochondrial Diseases Associated with tRNA Wobble Modification Deficiency. *RNA Biol.* **2005**, *2*, 41-44.
- [98] Y. Kirino, et al., Specific correlation between the wobble modification deficiency in mutant tRNAs and the clinical features of a human mitochondrial disease. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7127-7132.
- [99] M.-K. Shin, et al., Association between CDK5RAP1 polymorphisms and susceptibility to vitiligo in the Korean population *Eur. J. Dermatol.* **2012**, *22*, 495-499.
- [100] V. H. Cowling, Regulation of mRNA cap methylation. *Biochem. J.* **2010**, *425*, 295-302.
- [101] K. Nishikura, A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 83-96.
- [102] S. Farajollahi, S. Maas, Molecular Diversity through RNA Editing: A Balancing Act. *Trends Genet.* **2010**, *26*, 221-230.
- [103] L. Han, et al., The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers. *Cancer cell* **2015**, *28*, 515-528.

- [104] L. Chen, et al., Recoding RNA editing of antizyme inhibitor 1 predisposes to hepatocellular carcinoma. *Nat. Med.* **2013**, *19*, 209-216.
- [105] N. Paz-Yaacov, et al., Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors. *Cell Reports* **2015**, *13*, 267-276.
- [106] D. Weissmann, et al., Region-specific alterations of A-to-I RNA editing of serotonin 2c receptor in the cortex of suicides with major depression. *Transl. Psychiat.* **2016**, *6*, e878.
- [107] K. A. Berg, et al., RNA-editing of the 5-HT(2C) receptor alters agonist-receptor-effector coupling specificity. *Brit. J. Pharmacol.* **2001**, *134*, 386-392.
- [108] C. M. Niswender, et al., RNA Editing of the Human Serotonin 5-Hydroxytryptamine 2C Receptor Silences Constitutive Activity. *J. Biol. Chem.* **1999**, *274*, 9472-9478.
- [109] R. D. Price, E. Sanders-Bush, RNA Editing of the Human Serotonin 5-HT(2C) Receptor Delays Agonist-Stimulated Calcium Release. *Mol. Pharmacol.* **2000**, *58*, 859-862.
- [110] R. D. Price, et al., RNA Editing of the Human Serotonin 5-HT2C Receptor Alters Receptor-mediated Activation of G13 Protein. *J. Biol. Chem.* **2001**, *276*, 44663-44668.
- [111] M. S. Paul, B. L. Bass, Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *The EMBO Journal* **1998**, *17*, 1120-1127.
- [112] M. Sakurai, et al., Inosine cyanoethylation identifies A-to-I RNA editing sites in the human transcriptome. *Nat. Chem. Biol.* **2010**, *6*, 733.
- [113] M. Sakurai, et al., A biochemical landscape of A-to-I RNA editing in the human brain transcriptome. *Genome Res.* **2014**, *24*, 522-534.
- [114] B.-E. Wulff, et al., Elucidating the inosinome: global approaches to adenosine-to-inosine RNA editing. *Nat. Rev. Genet.* **2011**, *12*, 81-85.
- [115] G. Nigita, et al., A-to-I RNA Editing: Current Knowledge Sources and Computational Approaches with Special Emphasis on Non-Coding RNA Molecules. *Front. Bioeng. Biotechnol.* **2015**, *3*, 37.
- [116] R. C. Desrosiers, et al., Characterization of Novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5' terminus. *Biochemistry* **1975**, *14*, 4367-4374.
- [117] R. Desrosiers, et al., Identification of Methylated Nucleosides in Messenger RNA from Novikoff Hepatoma Cells. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3971-3975.
- [118] C. M. Wei, et al., 5'-Terminal and internal methylated nucleotide sequences in HeLa cell mRNA. *Biochemistry* **1976**, *15*, 397-401.
- [119] C. M. Wei, B. Moss, Nucleotide sequences at the N6-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **1977**, *16*, 1672-1676.
- [120] F. Rottman, et al., Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. *Cell* **1974**, *3*, 197-199.
- [121] P. Narayan, F. Rottman, An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. *Science* **1988**, *242*, 1159-1162.
- [122] P. Narayan, et al., Context effects on N6-adenosine methylation sites in prolactin mRNA. *Nucleic Acids Res.* **1994**, *22*, 419-426.
- [123] T. Csepány, et al., Sequence specificity of mRNA N6-adenosine methyltransferase. *J. Biol. Chem.* **1990**, *265*, 20117-20122.
- [124] D. Dominissini, et al., Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **2012**, *485*, 201.
- [125] K. D. Meyer, et al., Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and Near Stop Codons. *Cell* **2012**, *149*, 1635-1646.
- [126] B. Linder, et al., Single-nucleotide resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* **2015**, *12*, 767-772.
- [127] T. Ni, et al., A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods* **2010**, *7*, 521.
- [128] B. Moss, et al., Histone mRNAs contain blocked and methylated 5' terminal sequences but lack methylated nucleosides at internal positions. *Cell* **1977**, *10*, 113-120.
- [129] Y. Fu, et al., Gene expression regulation mediated through reversible m6A RNA methylation. *Nat. Rev. Genet.* **2014**, *15*, 293.

- [130] J. Mauer, et al., Reversible methylation of m(6)A(m) in the 5' cap controls mRNA stability. *Nature* **2017**, *541*, 371-375.
- [131] K. D. Meyer, S. R. Jaffrey, Rethinking m6A Readers, Writers, and Erasers. *Annu. Rev. Cell Dev. Bi.* **2017**, *33*, 319-342.
- [132] N. Liu, T. Pan, RNA epigenetics. *Transl. Res.* **2015**, *165*, 28-35.
- [133] S. D. Agarwala, et al., RNA Methylation by the MIS Complex Regulates a Cell Fate Decision in Yeast. *PLOS Genet.* **2012**, *8*, e1002732.
- [134] S. Geula, et al., Stem cells. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **2015**, *347*, 1002-1006.
- [135] J. Liu, et al., A METTL3-METTL14 complex mediates mammalian nuclear RNA N(6)-adenosine methylation. *Nat. Chem. Biol.* **2014**, *10*, 93-95.
- [136] X.-L. Ping, et al., Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* **2014**, *24*, 177-189.
- [137] S. Schwartz, et al., Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* **2014**, *8*, 284-296.
- [138] P. Ślędź, M. Jinek, Structural insights into the molecular mechanism of the m6A writer complex. *eLife* **2016**, *5*, e18434.
- [139] P. Wang, et al., Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. *Mol. Cell* **2016**, *63*, 306-317.
- [140] X. Wang, et al., Structural basis of N6-adenosine methylation by the METTL3–METTL14 complex. *Nature* **2016**, *534*, 575.
- [141] K. Horiuchi, et al., Wilms' tumor 1-associating protein regulates G(2)/M transition through stabilization of cyclin A2 mRNA. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17278-17283.
- [142] N. Liu, et al., N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **2015**, *518*, 560-564.
- [143] Y. Yue, et al., RNA N(6)-methyladenosine methylation in post-transcriptional gene expression regulation. *Gene Dev.* **2015**, *29*, 1343-1355.
- [144] K. Horiuchi, et al., Identification of Wilms' Tumor 1-associating Protein Complex and Its Role in Alternative Splicing and the Cell Cycle. *J. Biol. Chem.* **2013**, *288*, 33292-33302.
- [145] D. P. Patil, et al., m6A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **2016**, *537*, 369.
- [146] J. A. Bokar, et al., Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *J. Biol. Chem.* **1994**, *269*, 17697-17704.
- [147] D. P. Patil, et al., Reading m6A in the Transcriptome: m6A-Binding Proteins. *Trends Cell Biol.* **2018**, *28*, 113-127.
- [148] P. Batista, et al., m6A RNA modification controls cell fate transition in mammalian embryonic stem cells. *Cell Stem Cell* **2014**, *15*, 707-719.
- [149] X. Wang, et al., m(6)A-dependent regulation of messenger RNA stability. *Nature* **2014**, *505*, 117-120.
- [150] R. R. Edupuganti, et al., N6-methyladenosine (m6A) recruits and repels proteins to regulate mRNA homeostasis. *Nat. Struct. Mol. Biol.* **2017**, *24*, 870.
- [151] O. Penagarikano, et al., The Pathophysiology of Fragile X Syndrome. *Annu. Rev. Genom. Hum. G.* **2007**, *8*, 109-129.
- [152] V. Brown, et al., Microarray Identification of FMRP-Associated Brain mRNAs and Altered mRNA Translational Profiles in Fragile X Syndrome. *Cell* **2001**, *107*, 477-487.
- [153] J. C. Darnell, et al., FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **2011**, *146*, 247-261.
- [154] X. Wang, et al., N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* **2015**, *161*, 1388-1399.
- [155] Z. Zhang, et al., The YTH Domain Is a Novel RNA Binding Domain. *J. Biol. Chem.* **2010**, *285*, 14701-14710.

- [156] W. Xiao, et al., Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **2016**, *61*, 507-519.
- [157] C. R. Alarcón, et al., N(6)-methyl-adenosine (m6A) marks primary microRNAs for processing. *Nature* **2015**, *519*, 482-485.
- [158] E. Kierzek, R. Kierzek, The thermodynamic stability of RNA duplexes and hairpins containing N(6)-alkyladenosines and 2-methylthio-N(6)-alkyladenosines. *Nucleic Acids Res.* **2003**, *31*, 4472-4480.
- [159] C. Roost, et al., Structure and Thermodynamics of N6-Methyladenosine in RNA: A Spring-Loaded Base Modification. *J. Am. Chem. Soc.* **2015**, *137*, 2107-2115.
- [160] J. T. Lee, Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Gene Dev.* **2009**, *23*, 1831-1842.
- [161] K. D. Meyer, et al., 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **2015**, *163*, 999-1010.
- [162] G. Jia, et al., N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **2011**, *7*, 885.
- [163] G. Zheng, et al., ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* **2013**, *49*, 18-29.
- [164] T. Brückl, et al., Parallel Isotope-Based Quantification of Modified tRNA Nucleosides. *Angew. Chem. Int. Ed.* **2009**, *48*, 7932-7934.
- [165] K. O'Brien, et al., The Biflavonoid Isoginkgetin Is a General Inhibitor of Pre-mRNA Splicing. *J. Biol. Chem.* **2008**, *283*, 33147-33154.
- [166] M. Huranová, et al., The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. *J. Cell Biol.* **2010**, *191*, 75-86.
- [167] J. E. Carette, et al., Global gene disruption in human cells to assign genes to phenotypes. *Nat. Biotechnol.* **2011**, *29*, 542-546.
- [168] C. Schweingruber, et al., Nonsense-mediated mRNA decay — Mechanisms of substrate mRNA recognition and degradation in mammalian cells. *BBA-Gene Regul. Mech.* **2013**, *1829*, 612-623.
- [169] M. Fakruddin, et al., Cdk5rap1-mediated 2-methylthio-N6-isopentenyladenosine modification is absent from nuclear-derived RNA species. *Nucleic Acids Res.* **2017**, *45*, 11954-11961.
- [170] M. Toompuu, et al., Polyadenylation and degradation of structurally abnormal mitochondrial tRNAs in human cells. *Nucleic Acids Res.* **2018**, *46*, 5209–5226.
- [171] T. Pfaffeneder, et al., Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **2014**, *10*, 574-581.
- [172] A. C. Drohat, C. T. Coey, Role of Base Excision “Repair” Enzymes in Erasing Epigenetic Marks from DNA. *Chem. Rev.* **2016**, *116*, 12711-12729.
- [173] C. Popp, et al., Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by Aid deficiency. *Nature* **2010**, *463*, 1101-1105.
- [174] P. Hajkova, et al., Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **2010**, *329*, 78-82.
- [175] T. Lindahl, R. D. Wood, Quality Control by DNA Repair. *Science* **1999**, *286*, 1897-1905.
- [176] F. Agius, et al., Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11796-11801.
- [177] J. Penterman, et al., DNA demethylation in the *Arabidopsis* genome. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 6752-6757.
- [178] A. Maiti, A. C. Drohat, Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites. *J. Biol. Chem.* **2011**, *286*, 35334-35338.
- [179] K. Wiebauer, J. Jiricny, Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 5842-5845.
- [180] R. Rahimoff, et al., 5-Formyl- and 5-Carboxydeoxycytidines Do Not Cause Accumulation of Harmful Repair Intermediates in Stem Cells. *J. Am. Chem. Soc.* **2017**, *139*, 10359-10364.

- [181] A. C. Drohat, A. Maiti, Mechanisms for enzymatic cleavage of the N-glycosidic bond in DNA. *Org. Biomol. Chem.* **2014**, *12*, 8367-8378.
- [182] B. Delatte, F. Fuks, TET proteins: on the frenetic hunt for new cytosine modifications. *Brief. Funct. Genomics* **2013**, *12*, 191-204.
- [183] Z. Gong, J.-K. Zhu, Active DNA demethylation by oxidation and repair. *Cell Res.* **2011**, *21*, 1649-1651.
- [184] F. Guo, et al., Active and Passive Demethylation of Male and Female Pronuclear DNA in the Mammalian Zygote. *Cell Stem Cell* **2014**, *15*, 447-459.
- [185] M. Bachman, et al., 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* **2015**, *11*, 555-557.
- [186] H. Kim, et al., Modulation of β -catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal. *Nat. Commun.* **2013**, *4*, 2403-2403.
- [187] A. N. Lane, T. W. M. Fan, Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Res.* **2015**, *43*, 2466-2485.
- [188] M. Vogeser, C. Seger, Pitfalls Associated with the Use of Liquid Chromatography–Tandem Mass Spectrometry in the Clinical Laboratory. *Clin. Chem.* **2010**, *56*, 1234-1244.
- [189] A. Tsumura, et al., Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **2006**, *11*, 805-814.
- [190] M. Hooper, et al., HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **1987**, *326*, 292.
- [191] F. Lu, et al., Role of Tet proteins in enhancer activity and telomere elongation. *Gene Dev.* **2014**, *28*, 2103-2119.
- [192] S. Schiesser, et al., Deamination, Oxidation, and C–C Bond Cleavage Reactivity of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine. *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599.
- [193] K. Iwan, et al., 5-Formylcytosine to cytosine conversion by C–C bond cleavage in vivo. *Nat. Chem. Biol.* **2017**, *14*, 72.
- [194] M. R. Waterman, G. I. Lepesheva, Sterol 14 α -demethylase, an abundant and essential mixed-function oxidase. *Biochem. Biophys. Res. Co.* **2005**, *338*, 418-422.
- [195] T. Y. Hargrove, et al., Substrate Preferences and Catalytic Parameters Determined by Structural Characteristics of Sterol 14 α -Demethylase (CYP51) from *Leishmania infantum*. *J. Biol. Chem.* **2011**, *286*, 26838-26848.
- [196] G. I. Lepesheva, M. R. Waterman, Sterol 14 α -Demethylase Cytochrome P450 (CYP51), a P450 in all Biological Kingdoms. *Biochim. Biophys. Acta* **2007**, *1770*, 467-477.
- [197] A. Z. Shyadehi, et al., The Mechanism of the Acyl-Carbon Bond Cleavage Reaction Catalyzed by Recombinant Sterol 14 α -Demethylase of *Candida albicans* (Other Names Are: Lanosterol 14 α -Demethylase, P-45014DM, and CYP51). *J. Biol. Chem.* **1996**, *271*, 12445-12450.
- [198] N. Li, et al., Conversion of Fatty Aldehydes to Alka(e)nes and Formate by a Cyanobacterial Aldehyde Decarboxylase: Cryptic Redox by an Unusual Dimetal Oxygenase. *J. Am. Chem. Soc.* **2011**, *133*, 6158-6161.
- [199] K. G. Aukema, et al., Cyanobacterial aldehyde deformylase oxygenation of aldehydes yields n-1 aldehydes and alcohols in addition to alkanes. *ACS Catal.* **2013**, *3*, 2228-2238.
- [200] J. A. Smiley, et al., Radioactivity-Based and Spectrophotometric Assays for Isoorotate Decarboxylase: Identification of the Thymidine Salvage Pathway in Lower Eukaryotes. *Anal. Biochem.* **1999**, *266*, 85-92.
- [201] S. Xu, 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.* **2013**, *23*, 1296.
- [202] M. Fujihashi, et al., Orotidine Monophosphate Decarboxylase – A Fascinating Workhorse Enzyme with Therapeutic Potential. *J. Genet. Genomics* **2015**, *42*, 221-234.
- [203] T. H. Bestor, The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **2000**, *9*, 2395-2402.

- [204] M. Okano, et al., DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **1999**, *99*, 247-257.
- [205] J. C. Morales, et al., Review of Poly (ADP-ribose) Polymerase (PARP) Mechanisms of Action and Rationale for Targeting in Cancer and Other Diseases. *Crit. Rev. Eukar. Gene* **2014**, *24*, 15-28.
- [206] J. C. Amé, et al., The PARP superfamily. *BioEssays* **2004**, *26*, 882-893.
- [207] D. D'Amours, et al., Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochem. J.* **1999**, *342*, 249-268.
- [208] K. W. Caldecott, et al., XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Res.* **1996**, *24*, 4387-4394.
- [209] C. Flohr, et al., Poly(ADP-ribosyl)ation accelerates DNA repair in a pathway dependent on Cockayne syndrome B protein. *Nucleic Acids Res.* **2003**, *31*, 5332-5337.
- [210] E. L. Jacobson, et al., Cellular recovery of dividing and confluent C3H10T1/2 cells from N-methyl-N'-nitro-N-nitrosoguanidine in the presence of ADP-ribosylation inhibitors. *Carcinogenesis* **1985**, *6*, 715-718.
- [211] G. G. Poirier, et al., Poly(ADP-ribosyl)ation of polynucleosomes causes relaxation of chromatin structure. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 3423-3427.
- [212] E. Seto, M. Yoshida, Erasers of Histone Acetylation: The Histone Deacetylase Enzymes. *CSH Perspect. Biol.* **2014**, *6*, a018713.
- [213] S.-i. Imai, et al., Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **2000**, *403*, 795.
- [214] A. Nagy, et al., Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8424-8428.
- [215] M. W. McBurney, et al., The Mammalian SIR2 α Protein Has a Role in Embryogenesis and Gametogenesis. *Mol. Cell. Biol.* **2003**, *23*, 38-54.
- [216] H. Lei, et al., De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **1996**, *122*, 3195-3205.