Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Investigation of the role and regulation of modified DNA and RNA nucleosides

Caterina Brandmayr

aus

Verona, Italien

2013

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 15. November 2013

Caterina Brandmayr

Dissertation eingereicht am	25 November 2013
1. Gutachter:	Prof. Dr. Thomas Carell
2. Gutachter:	Dr. Stylianos Michalakis
Mündliche Prüfung am	11 February 2014

Parts of this thesis were published:

Publications:

T. Pfaffeneder,[1] F. Spada,[1] M. Wagner,[1] C. Brandmayr, S. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Muller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schar, M. Muller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574-581. "*Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.“*

B. Steigenberger,[1] S. Schiesser,[1] B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366-369. "*Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction.*"

T. Carell, C. Brandmayr, A. Hienzsch, M. Muller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131. "*Structure and Function of Noncanonical Nucleobases.*" (Review Article)

C. Brandmayr,[1] M. Wagner,[1] T. Bruckl,[1] D. Globisch, D. Pearson, A. C. Kneuttinger, V. Reiter, A. Hienzsch, S. Koch, I. Thoma, P. Thumbs, S. Michalakis, M. Muller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 11162-11165. "*Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with Translational Efficiency*."

Further publications:

A. F. Wait, C. Brandmayr, S. T. Stripp, C. Cavazza, J. C. Fontecilla-Camps, T. Happe, F. A. Armstrong, *J. Am. Chem. Soc.* **2011**, *133*, 1282-1285. "*Formaldehyde-A Rapid and Reversible Inhibitor of Hydrogen Production by FeFe-Hydrogenases*."

S. T. Stripp, G. Goldet, C. Brandmayr, O. Sanganas, K. A. Vincent, M. Haumann, F. A. Armstrong, T. Happe, *Proc. Natl Acad. Sci. U. S. A.* **2009**, *106*, 17331-17336. "*How oxygen attacks FeFe hydrogenases from photosynthetic organisms.*"

G. Goldet, C. Brandmayr, S. T. Stripp, T. Happe, C. Cavazza, J. C. Fontecilla-Camps, F. A. Armstrong, *J. Am. Chem. Soc.* **2009**, *131*, 14979-14989. "*Electrochemical Kinetic Investigations of the Reactions of FeFe-Hydrogenases with Carbon Monoxide and Oxygen: Comparing the Importance of Gas Tunnels and Active-Site Electronic/Redox Effects*."

---

[1] These authors contributed equally to this work.

Acknowledgments

First of all, I would like to thank Prof. Dr. Thomas Carell for giving me the opportunity to work in his research group. Throughout the PhD he has been extremely supportive and has given me an invaluable chance to explore many new subjects and aspects of research. What I most appreciated was his openness and curiosity for new ideas, as well as his ability to look at my failures as inspiration for new experiments and to always motivate me in my projects.

A warm thank you also goes to Dr. Markus Müller. He has always been extremely helpful in all respects, whether it was related to management of the daily lab life, teaching of new techniques, or discussing together bigger scientific questions. I really appreciated his strong interest in science and research and his critical thinking.

I would also like to thank Dr. Stylianos Michalakis and Dr. Heinrich Leonhardt for the fruitful collaborations and scientific discussions.

The Boehringer Ingelheim Fonds have been invaluable not only for their financial support during my PhD, but also for giving me the opportunity to broaden my scientific knowledge and meet many interesting and interested people.

The organizational and administrative aspects for the PhD definitely would not have been as smooth without the contribution of Mrs. Slava Gärtner, Mrs Sabine Voß and Kerstin Kurz.

Past and present members of the group also deserve a special mention. First of all, I thank Dr. David Pearson, our beloved kiwi, with whom I worked on the MnmC project and shared a lot of hilarious moments together. I also really enjoyed working with Benjamin Hackner and the epigenetics group, the work together has been really interesting and stimulating. Dr. Danila Fazio, Dr. Antje Hienzsch and Dr. Ulrike Lischke have been of great support and company throughout my PhD.

I am deeply thankful to Mirko Wagner, who has been an invaluable colleague and a dear friend. He has been extremely supportive both for work and life, and I wish him all the best for his future.

Finally, a warm thank you goes to all the friends scattered in Munich and elsewhere, and to my family. They have been the basis for my happiness and motivation.

Table of contents

Summary V
1 Introduction 1
1.1 RNA modifications 1
1.1.1 Transfer RNA 2
1.1.2 Functions of tRNA modifications 3
1.1.2.1 Modulation of codon-anticodon interaction 3
1.1.2.2 Folding and structural stabilisation 5
1.1.2.3 tRNA recognition 6
1.1.3 Biogenesis of tRNAs and tRNA modifications 6
1.1.4 Regulation of tRNA modifications 8
1.1.5 tRNA modifications in evolution 8
1.2 DNA modifications 9
1.2.1 Epigenetic modulation of gene expression 9
1.2.2 Chromatin structure and histone modifications 10
1.2.3 DNA methylation 11
1.2.3.1 Distribution of DNA methylation 11
1.2.3.2 DNA Methyltransferases 13
1.2.4 Oxidized cytosine derivatives 14
1.2.4.1 Tet family proteins 15
1.2.4.2 Distribution of DNA oxidized cytosine derivatives 17
1.2.4.3 Putative roles of oxidized cytosine derivatives 18
2 Aims of the Project 21
3 Isotope-based analysis of modified tRNA nucleosides correlates modification density with translational efficiency 23
4 Investigation of yeast tRNA modification content and turnover during growth into stationary phase 29
4.1 Introduction 30
4.2 Results 30
4.2.1 Variation in tRNA content per cell during growth 31
4.2.2 Quantification of tRNA modifications during growth into stationary phase 32
4.2.3 Quantification of light and heavy labelled modifications during growth 36
4.2.4 Evaluation of CH_3- and CD_3-modification content over time 40
4.3 Discussion 41
4.4 Materials and Methods 47
4.4.1 Growth and handling of yeast 47
4.4.2 tRNA extraction 47

4.4.3 tRNA purification by ion-exchange chromatography 48
4.4.4 Enzymatic digestion of tRNA samples....48
4.4.5 HPLC-ESI-MS 48
4.4.6 Quantification of modified nucleosides and evaluation of light to heavy modification ratio....49
4.5 Quantification data 51
4.5.1 tRNA content per OD unit during growth of *S. cerevisiae* 51
4.5.2 Quantification of modified nucleosides....51
4.5.3 Evaluation of CH_3/CD_3 ratio during entry into stationary phase....54
4.5.4 Evaluation of CH_3- and CD_3-modification content 55
5 Investigation of the structure-function relationships for the bifunctional tRNA-modifying enzyme MnmC in complex with substrate tRNA....57
5.1 Introduction 58
5.2 Results and discussion.... 60
5.2.1 Purification of MnmC....60
5.2.2 Purification of undermodified *E.coli* $tRNA^{Glu}$....61
5.2.3 Characterization of the FAD-binding domain and design of inactive mutants for crystallization studies....62
5.2.3.1 Cloning, purification and enzymatic activity assay of MnmC mutants R567A, S617A and R618A....63
5.2.3.2 Cloning, purification and enzymatic activity assay of MnmC mutants Y312F, C500A, Y504F, Y521A and double mutant R567A/R618A 65
5.2.4 Circular dichroism studies of MnmC mutants....67
5.2.5 Spectroscopic properties of MnmC and comparison with mutants....67
5.2.6 Binding studies with T1 and T2 68
5.2.7 Preliminary crystallization screening of the MnmC-T2 complex 70
5.3 Conclusion.... 71
5.4 Materials and methods.... 71
5.4.1 Cloning of pET-28a and pET-30a MnmC....71
5.4.2 Mutagenesis of pET-28a and pET-30a....72
5.4.3 Protein expression and purification....73
5.4.4 tRNA expression and purification....74
5.4.5 Mutant activity assay....75
5.4.6 Circular dichroism....75
5.4.7 UV-Vis Spectroscopy....75
5.4.8 Electrophoretic mobility shift assay for the protein-tRNA complex....75
5.4.9 Protein crystallization screening....76

6 Investigation of putative pathways of cytosine demethylation involving C-C bond cleavage or replication-coupled dilution 77
6.1 Introduction 78
6.2 Results and discussion 79
6.2.1 Substrates 79
6.2.2 *In vitro* experiments using cell lysate samples 81
6.2.3 *In vitro* experiments using protein-enriched samples 82
6.2.4 Dnmt1 82
6.2.4.1 Purification of recombinant Dnmt1 82
6.2.4.2 Methylation assay 84
6.2.4.3 Interaction with oxidised cytosine derivatives 85
6.2.4.4 Decarboxylation and deformylation assays 87
6.3 Conclusion 87
6.4 Materials and methods 87
6.4.1 *In vitro* experiments using cell lysate samples 87
6.4.2 *In vitro* experiments using protein pull-down samples 89
6.4.3 Enzymatic digestion of DNA samples 90
6.4.4 HPLC-ESI-MS analysis 90
6.4.5 Overexpression and purification of mDnmt1 90
6.4.6 Methylation assay 91
6.4.7 Covalent trapping of mDnmt1 in the presence of modified cytosine derivatives 91
6.4.8 Decarboxylation and deformylation assays 92
7 Synthesis of 5-hydroxymethyl-, 5-formyl- and 5-carboxycytidine-triphosphates and their incorporation into oligonucleotides by Polymerase Chain Reaction 95
8 Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA 101
9 Outlook 113
10 Literature 115
11 Contributions 123
12 Abbreviations 124
13 Supplementary Information 127
13.1 Supplementary material for Chapter 3 127
13.2 Supplementary material for Chapter 7 159
13.3 Supplementary material for Chapter 8 177

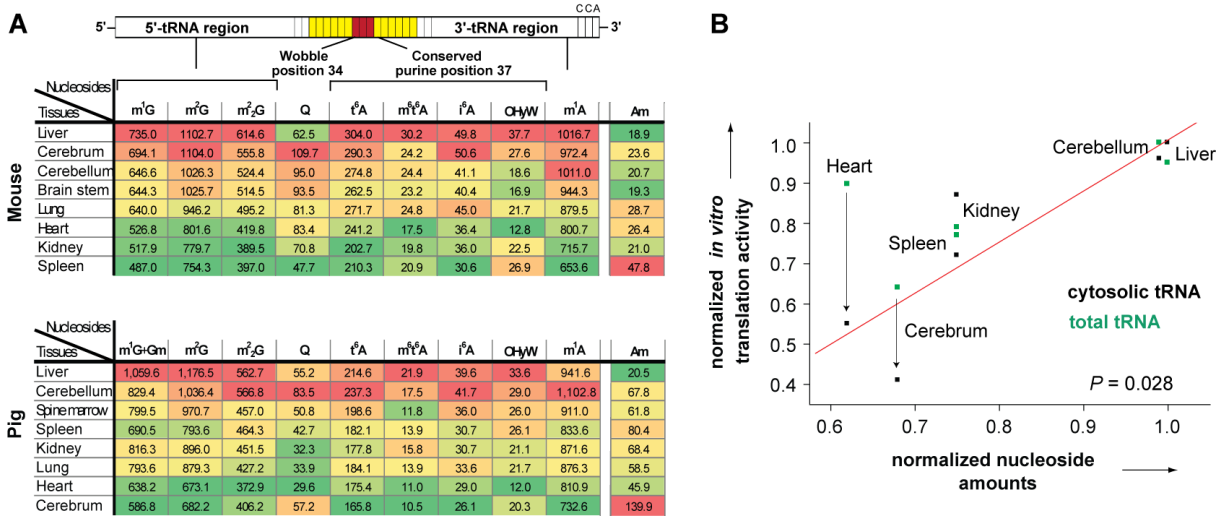
Summary

Nucleic acids are the fundamental molecules of life. Although both DNA and RNA are generally seen as composed of four canonical bases, they instead include a range of modified nucleosides which are involved in key epigenetic processes and in modulation of protein translation.

Analysis of modified tRNA nucleosides

The greatest variety of modified nucleosides is found in tRNA, where modifications can range from simple methylations to complex hypermodified nucleosides such as wybutosine. To date, a large number of studies have led to structural elucidation and identification of the biosynthetic pathways for a number of these modifications. However, there is still little information concerning the systemic behaviour of modified tRNA nucleosides, largely due to the fact that for a long time there was no method available for a parallel quantification of tRNA modification sets.

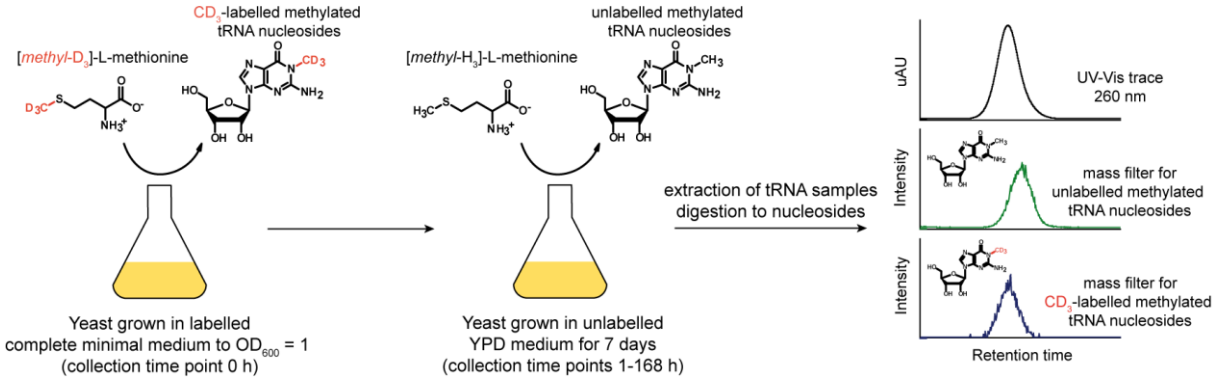
To this end, the Carell group recently developed an LC-MS based quantification method using synthetic isotope standards which provided us with an effective tool for the investigation of tRNA modifications in a system-based manner. We were interested in investigating the tissue-specific distribution of modified tRNA nucleosides and therefore quantified 11 representative RNA modifications in 8 different tissues from mouse and pig. Interestingly, tRNA populations were found to be differently modified depending on the tissue, with some tissues having overall highly modified tRNA and others displaying less modified tRNA pools, suggesting that different modification levels are required to satisfy different translational needs (Summary Figure 1). In support of this hypothesis, we obtained a positive correlation between modification levels and *in vitro* protein translation rates for cytosolic tRNA samples.



Summary Figure 1 A) Quantification of modified tRNA nucleosides in mouse and pig organs; B) Correlation analysis of tRNA modification content and *in vitro* translation activity.

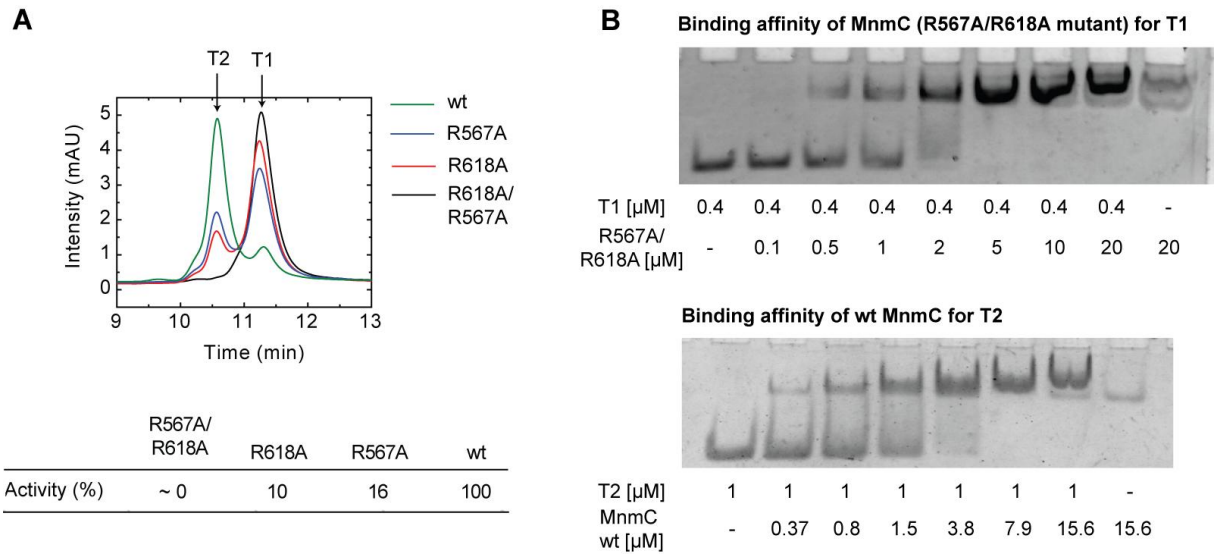
In a related project, we were interested in elucidating the regulation of tRNA modification levels in yeast cells entering stationary phase. In a previous study done in the Carell group, yeast was found to

display an increasingly modified tRNA population when grown to quiescence. We therefore decided to monitor the fate of tRNA modifications by pulse-chase labelling using labelled methionine during initial growth, followed by growth in unlabelled medium over 7 days (Summary Figure 2). Monitoring of the heavy (labelled) modification content over time revealed low modification turnover rates. Furthermore, analysis of the light to heavy ratios of the methylated tRNA nucleosides revealed a higher proportion than expected of the light (unlabelled) modification compared to the corresponding heavy (labelled) modification starting from day 1. This suggests that the observed increase in modification content per tRNA might result from increased modification activity over time, possibly linked to low tRNA turnover rates. Although we cannot exclude that there might be some extent of tRNA degradation, this process seems to preferentially target hypomodified tRNA species. We suggest that the increase in modification content observed during entry into stationary phase might be necessary for effective protein synthesis during stress response and it might additionally ensure a rapid start of translation upon exit from quiescence.



Summary Figure 2 Experimental procedure for monitoring tRNA modifications during yeast growth.

In addition, we were interested in investigating the molecular features which ensure the selective modification of uridine at the wobble position 34 of *E.coli* $tRNA^{Glu}$, $tRNA^{Lys}$ and $tRNA^{Arg}$. More specifically, we focused on the last two biosynthetic steps leading to the 5-methylaminomethyl-2-thiouridine (mnm^5s^2U) base, both of which are performed by the bifunctional enzyme MnmC. With the final goal of elucidating the biochemical and structural features of this enzyme in complex with substrate tRNAs for both the first and second modification steps, we identified the key residues involved in the first modification step and further characterized the binding affinity of MnmC to both substrate RNA molecules (Summary Figure 3).

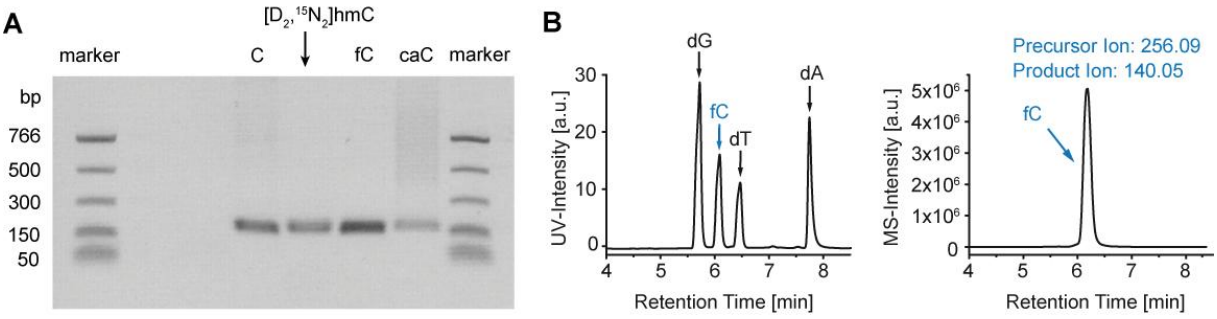


	R567A/ R618A	R618A	R567A	wt
Activity (%)	~ 0	10	16	100

Summary Figure 3 A) Activity assay of wild type and mutant MnmC with cmnm^{5}s^{2}U-containing substrate tRNA; B) Investigation of binding affinity of MnmC for substrate tRNA for both the first and second modification steps.

Analysis of modified DNA nucleosides

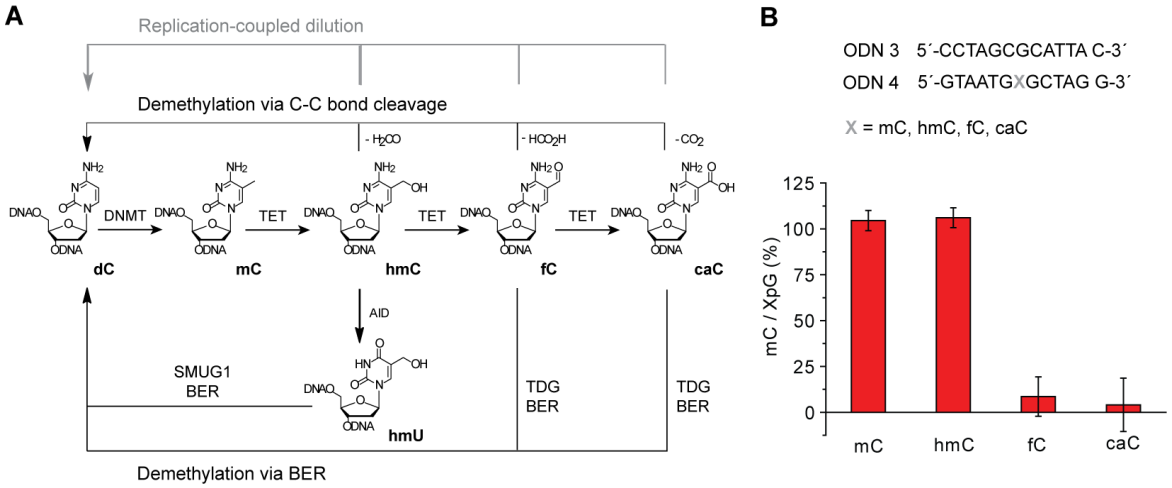
DNA also comprises a number of modified nucleosides. Most importantly, methylcytosine (mC) is a key player in epigenetic regulation of gene expression. Since 2009, additional modified cytosine derivatives have been discovered: these are hydroxymethylcytosine (hmC), formylcytosine (fC) and carboxycytosine (caC), all found to be generated by the Tet protein family. In the course of this study we were interested in elucidating whether hmC, fC and/or caC are involved in a pathway of active demethylation by C-C bond cleavage. In order to selectively monitor putative demethylation reactions, heavy atom-isotopologues of these cytosine derivatives were used in this work to generate DNA strands by Polymerase Chain Reaction (Summary Figure 4). These strands, containing the isotopically labelled cytosine derivatives, were subsequently used as substrates for *in vitro* assays performed using stem cell nuclear extracts. Up to now the protein or complex involved in this putative pathway of demethylation remains elusive and further experiments are ongoing.



Summary Figure 4 A) PCR products analyzed by gel electrophoresis; B) UV and mass trace of digested fC-containing PCR product.

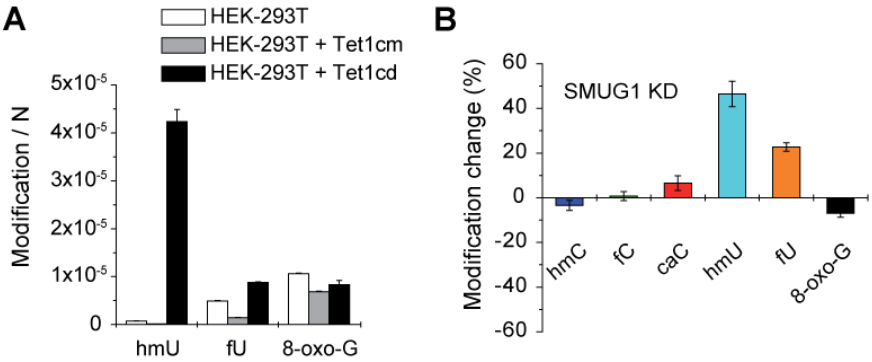
The purification of a recombinant mouse Dnmt1 protein fragment enabled to assess the *in vitro* methylation activity of Dnmt1 in the presence of fC and caC in a hemi-modified CpG context. This

assay revealed that maintenance methylation cannot take place in the presence of fC and caC, supporting a role for these modifications in removal of the methylation pattern via replication-dependent dilution (Summary Figure 5).



Summary Figure 5 A) Summary of demethylation pathways; B) Methylation activity of a recombinant mouse Dnmt1 fragment in the presence of a hemi-modified CpG (plotted as mC/XG (%), XG = modified CpG site, X = mC, hmC, fC or caC).

Finally, we were interested in investigating the deamination of hmC to hmU as an alternative pathway of active demethylation where deamination is coupled to base excision repair (BER). In the course of this study however, we observed that hmU is generated by enzymatic activity of Tet1 on T nucleobases. In fact, overexpression of the catalytic domain of Tet1 and its inactive equivalent in HEK-293T cells revealed that the generation of hmU is dependent on Tet1 activity (Summary Figure 6). The same experiment was repeated in the presence of labelled methionine: labelling was detected in the case of mC and hmC, as expected, but was not observed in the case of hmU. This confirms that hmU is not generated by deamination of hmC, but rather by oxidation of T. We further observed a Vitamin C-dependent increase of hmU production by Tet1 in HEK-293T cells and confirmed the removal of hmU in an A:T context by SMUG1. These results reveal that hmU is an additional nucleobase generated by Tet1 and suggest that hmU, up to now regarded only as an oxidative damage, might instead have a broader function.



Summary Figure 6 A) hmU generation in HEK-293T cells upon overexpression of Tet1 catalytic domain and inactive mutant; B) Effect of SMUG1 knock down on DNA modification levels.

Zusammenfassung

Nukleinsäuren sind die zentralen Moleküle des Lebens. Obwohl sowohl DNA als auch RNA gemeinhin als aus vier kanonischen Basen bestehend betrachtet werden, enthalten sie tatsächlich zusätzlich eine Vielzahl von modifizierten Nukleosiden, die entscheidend an epigenetischen Prozessen und der Modulation der Proteintranslation beteiligt sind.

Analyse modifizierter tRNA-Nukleoside

Die größte Vielfalt an modifizierten Nukleosiden findet man in der tRNA, wo neben einfachen Methylierungen auch komplexere hypermodifizierte Nukleoside wie z.B. Wybutosin vorkommen. Mittlerweile sind viele dieser Modifikationen sowohl im Hinblick auf ihre strukturellen Eigenschaften als auch bezüglich ihrer Biosynthese gut charakterisiert. Trotzdem gibt es nur wenige Informationen über das systemische Verhalten der modifizierten tRNA-Nukleoside als Gesamtheit, hauptsächlich weil bisher eine effektive Methode für die parallele Quantifizierung von tRNA-Modifikationssets fehlte. Aus diesem Grunde wurde kürzlich von Carell und Mitarbeitern eine LC-MS-basierte Quantifizierungsmethode für RNA-Modifikationen entwickelt, die auf der Verwendung von synthetischen Isotopologen als Massenstandards beruht und eine systembasierte Untersuchung von tRNA-Modifikationen erlaubt.

Mit dieser Methode sollte zunächst die Verteilung von modifizierten tRNA-Nukleosiden in Geweben unterschiedlicher Organe untersucht werden. Dazu haben wir 11 repräsentative RNA-Modifikationen ausgewählt und diese in 8 verschiedenen Gewebetypen, jeweils von Maus und Schwein, quantifiziert. Interessanterweise zeigte sich dabei, dass die tRNA-Populationen der verschiedenen Gewebetypen unterschiedlich stark modifiziert sind. Z.B. enthalten die tRNA-Sets von Leber und Kleinhirn alle untersuchten Nukleosidmodifikationen in großer Menge, andere Gewebe haben dagegen ein generell niedriges tRNA-Modifikationsniveau (Abbildung 1). Dies deutet darauf hin, dass unterschiedliche tRNA-Modifikationsgrade notwendig sind, um die speziellen Anforderungen bezüglich der Translation in den verschiedenen Geweben zu befriedigen. In Übereinstimmung mit dieser Hypothese konnten wir durch *in-vitro*-Experimente zeigen, dass eine positive Korrelation zwischen dem Modifikationsgrad cytosolischer tRNA-Sets und den entsprechenden Proteintranslationsraten besteht.

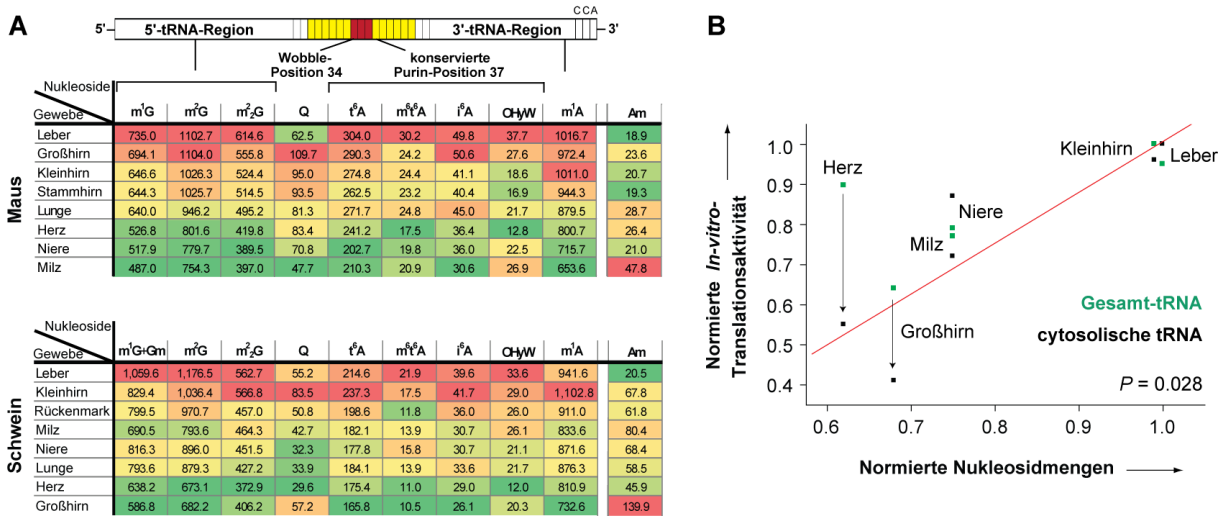


Abbildung 1 A) Quantifizierung modifizierter tRNA-Nukleoside in Mäuse- und Schweineorganen; B) Korrelationsanalyse zwischen tRNA-Modifikationsgehalt und *in-vitro*-Translationsaktivität.

In einem ähnlichen Projekt haben wir die Regulation des tRNA-Modifikationsniveaus in Hefezellen beim Erreichen der stationären Wachstumsphase untersucht. In einer früheren Arbeit von Carell und Mitarbeitern wurde gezeigt, dass das tRNA-Modifikationsniveau ansteigt, sobald sich die Hefezellen dem stationären Zustand nähern. Wir wollten daher das Schicksal der einzelnen tRNA-Modifikationen durch *Pulse-Chase-Labelling*-Experimente verfolgen. Dazu wurde dem Nährmedium während der ersten Wachstumsphase Methyl-D_3-markiertes Methionin zugesetzt, was zu einer CD_3-Isotopenmarkierung der neugebildeten methylierten tRNA-Nukleoside führte. Dann wurden die Hefezellen sieben Tage lang in unmarkiertem Medium weiterkultiviert (Abbildung 2). Zu verschiedenen Zeitpunkten wurden Proben entnommen und das Verhältnis von leichtem (unmarkiertem) zu schwerem (markiertem) Isotopolog bei den methylierten tRNA-Nukleosiden analysiert. Eine Bestimmung der Absolutmengen der markierten Modifikationen zu verschiedenen Zeitpunkten ergab geringe Umsatzraten für die verschiedenen Modifikationen. Zusätzlich war ab Tag 1 bei allen untersuchten Modifikationen der Anteil an unmarkierten Nukleosiden im Vergleich zum Anteil an markierten Nukleosiden größer als erwartet. Diese Ergebnisse deuten darauf hin, dass der beobachtete Anstieg der Modifikationsdichte wahrscheinlich aus einer gesteigerten Modifikationsaktivität resultiert, eventuell in Verbindung mit niedrigen tRNA-Umsatzraten. Obwohl wir nicht ausschließen können, dass es bei einem Teil der tRNAs zu einer Degradation kommt, scheint dieser Prozess selektiv die hypomodifizierten tRNA-Spezies zu betreffen. Wir vermuten, dass die beobachtete Zunahme der Modifikationsdichte beim Eintritt der Hefezellen in die stationäre Phase notwendig ist, um eine effektive Proteinsynthese auch bei ungünstigen Wachstumsbedingungen zu gewährleisten. Zusätzlich könnte die hohe tRNA-Modifikationsdichte auch einen schnellen Neustart der Translation beim Ausgang aus dem Ruhezustand ermöglichen.

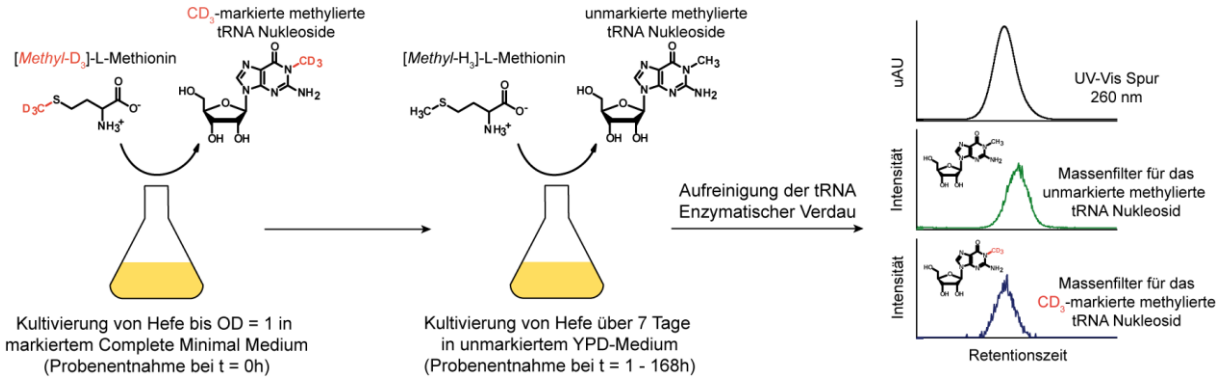
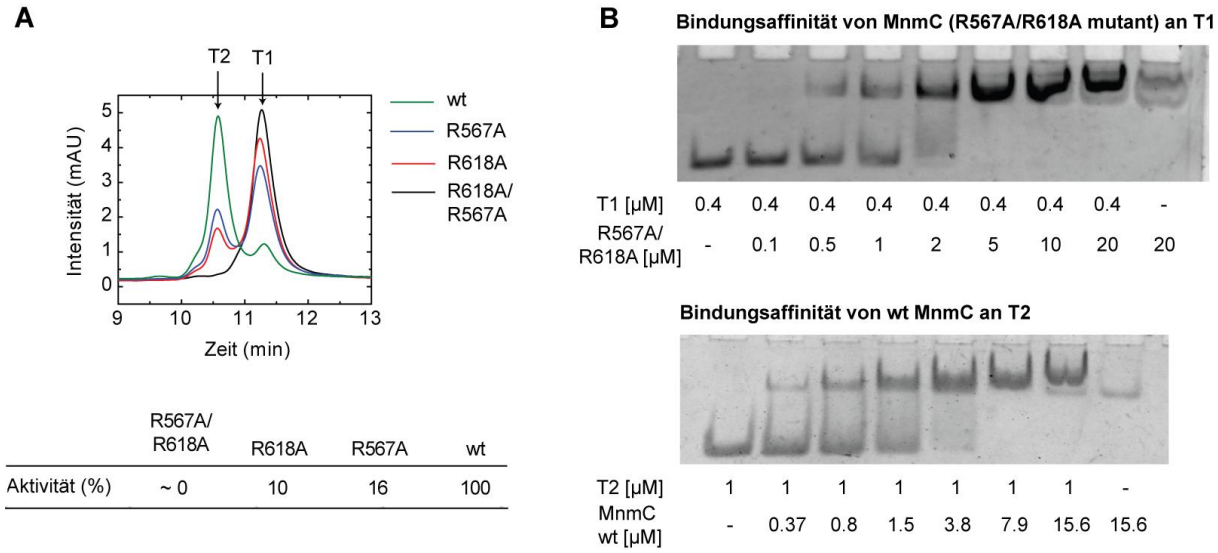


Abbildung 2 Experiment zur Untersuchung der modifizierten tRNA-Nukleoside von Hefe in verschiedenen Wachstumsphasen.

In einem weiteren Projekt wollten wir die molekularen Prozesse aufklären, die eine selektive Modifikation des Uridins an der Wobble-Position (Position 34) in den *E. coli*-tRNAs $tRNA^{Glu}$, $tRNA^{Lys}$ und $tRNA^{Arg}$ ermöglichen. Dabei haben wir uns auf die letzten beiden Stufen der mnm^5s^2U-(5-Methylaminomethyl-2-thiouridin)-Biosynthese konzentriert, die beide von dem bifunktionellen Enzym MnmC katalysiert werden. Das Ziel dabei war, die biochemischen und strukturellen Eigenschaften von MnmC aufzuklären, auch im Komplex mit den Substrat-tRNAs sowohl des ersten als auch des zweiten Modifikationsschrittes. Im Rahmen dieses Projektes konnten wir die Bindungsaffinität von MnmC zu beiden Substrat-tRNAs bestimmen und die für den ersten Modifikationsschritt entscheidenden Aminosäuren identifizieren (Abbildung 3).



	R567A/ R618A	R618A	R567A	wt
Aktivität (%)	~ 0	10	16	100

Abbildung 3 A) Aktivitätsassay von Wildtyp- und mutiertem MnmC mit $cmnm^5s^2U$-enthaltender Substrat-tRNA; B) Untersuchung der Bindungsaffinität von MnmC zu den Substrat-tRNAs für den ersten und den zweiten Modifikationsschritt.

Analyse modifizierter DNA-Nukleoside

Auch die DNA enthält mehrere modifizierte Nukleoside; sehr wichtig ist z.B. das Methylcytosin (mC) als Hauptakteur in der epigenetischen Regulation der Genexpression. Seit 2009 wurden weitere Cytosinderivate als Bestandteil der DNA entdeckt. Es handelt sich hierbei um Hydroxymethylcytosin (hmC), Formylcytosin (fC) und Carboxycytosin (caC), die alle von den Tet-Enzymen generiert werden.

Während der Prozess der Methylierung von Cytosin detailliert untersucht ist, gibt es praktisch nur wenige Erkenntnisse darüber, wie diese Methylierung wieder rückgängig gemacht werden kann. Wir wollten daher zunächst untersuchen, ob hmC, fC und / oder caC als Zwischenstufen an einer aktiven Demethylierung durch Spaltung der C-C-Bindung zwischen der substituierten Methylgruppe und dem C5-Atom des Cytosins in Frage kommen. Um mögliche Demethylierungsreaktionen selektiv verfolgen zu können, wurden im Rahmen dieser Studie synthetische Isotopologe dieser Cytosinderivate durch eine Polymerase-Kettenreaktion (PCR) in DNA-Stränge eingebaut (Abbildung 4). Diese DNA-Stränge wurden dann für *in-vitro-Assays* verwendet, bei denen sie mit Zellkernextrakten von Stammzellen inkubiert wurden. In diesen Experimenten konnte jedoch bis jetzt keine aktive Demethylierung nachgewiesen werden, und auch ein für die postulierte aktive Demethylierung verantwortliches Enzym oder Komplex konnte noch nicht identifiziert werden.

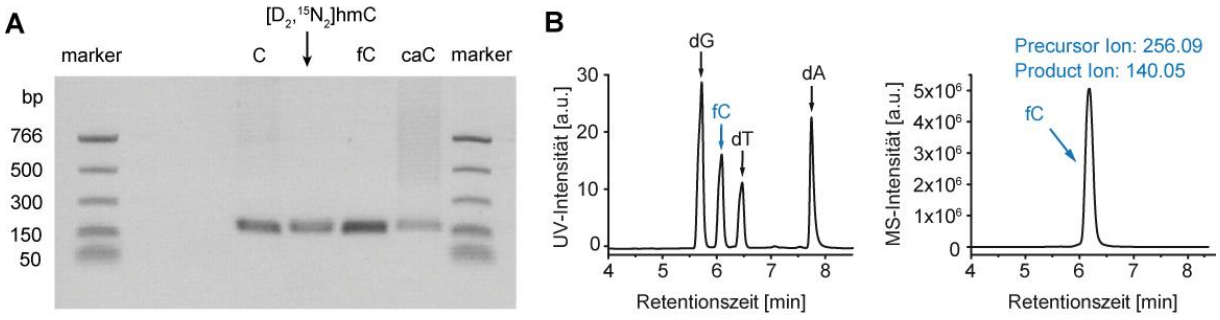


Abbildung 4 A) Analyse der PCR-Produkte durch Gelelektrophorese; B) UV- und Massenspur des verdauten, fC-enthaltenden PCR-Produktes.

Das Enzym Dnmt1 ist bei der DNA-Replikation zuständig für das Übertragen des Methylierungsmusters auf den neugebildeten Strang. Die Methylierungsaktivität von Dnmt1 beim Vorhandensein von fC oder caC in der hemimethylierten DNA konnte von uns *in-vitro* durch die erfolgreiche Reindarstellung eines rekombinanten Dnmt1-Proteinfragments der Maus untersucht werden. Befindet sich fC oder caC im DNA-Strang, so findet keine Methylierung des gegenüberliegenden DNA-Abschnittes durch Dnmt1 statt, auch wenn es sich dabei um einen CpG-Abschnitt handelt (Abbildung 5). Dies deutet auf eine mögliche Rolle von fC und caC im Rahmen einer passiven DNA-Demethylierung durch replikationsgekoppelte Verdünnung hin.

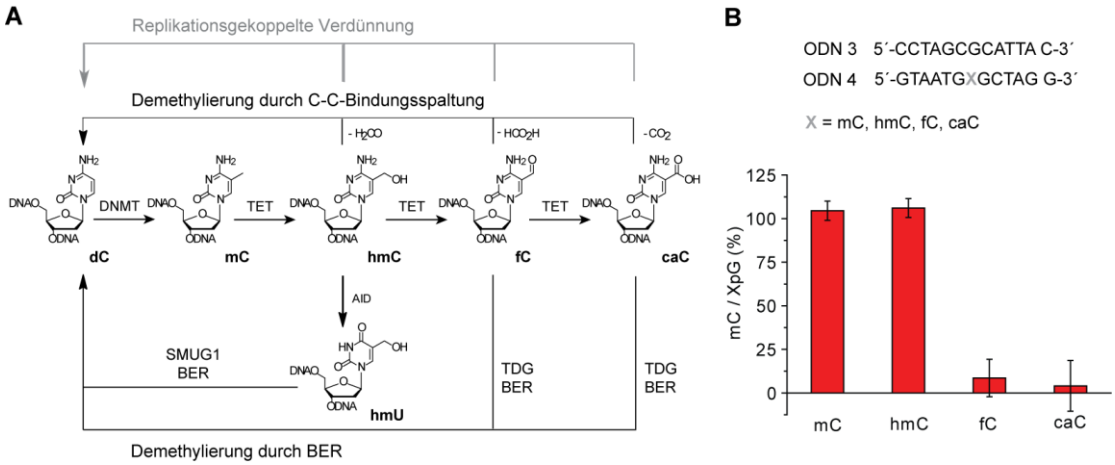


Abbildung 5 A) Übersicht über mögliche Demethylierungswege; B) Methylierungsaktivität eines rekombinanten Dnmt1-Proteinfragments der Maus bezüglich halbmodifiziertem CpG (Dargestellt als mC/XG (%), XG = modifiziertes CpG, X = mC, hmC, fC oder caC).

Eine alternative Hypothese über den Ablauf der aktiven DNA-Demethylierung beinhaltet die Deaminierung von hmC zu Hydroxymethyluracil (hmU), gefolgt von einer Basen-Exzisionsreparatur (*base excision repair*, BER), bei der hmU durch C ersetzt wird. Unsere Untersuchungen zur Deaminierung von hmC ergaben jedoch, dass diese Reaktion praktisch nicht stattfindet. Stattdessen konnten wir zeigen, dass auch hmU durch die enzymatische Aktivität von Tet1 gebildet wird, wobei Thymin (T) dem Enzym als Substrat dient. Dazu haben wir sowohl die katalytische Domäne (Tet1cd) als auch eine inaktive Variante von Tet1 (Tet1cm) in HEK-293T-Zellen überexprimiert und die modifizierten DNA-Nukleoside *via* LC-MS/MS quantifiziert. Dabei war das hmU-Level bei Überexpression von Tet1cd um den Faktor 60 größer als bei einer Expression von Tet1cm (Abbildung 6). Bei einer Wiederholung dieses Experiments in Gegenwart von markiertem Methionin konnten wir zeigen, dass sich die Markierung wie erwartet zwar bei mC und hmC, nicht aber bei hmU findet, d.h. hmU wird nicht durch Deaminierung von hmC, sondern durch Oxidation von T gebildet. Wir konnten weiterhin einen Vitamin C-abhängigen Anstieg der hmU-Produktion durch Tet1 nachweisen und bestätigen, dass hmU durch die Glycosylase SMUG1 in einem A:T- bzw. A:hmU-Kontext aus der DNA entfernt wird. Unsere Resultate zeigen, dass hmU eine weitere durch Tet1 gebildete Nukleobase ist und nicht nur, wie bisher vermutet, ein oxidativer DNA-Schaden.

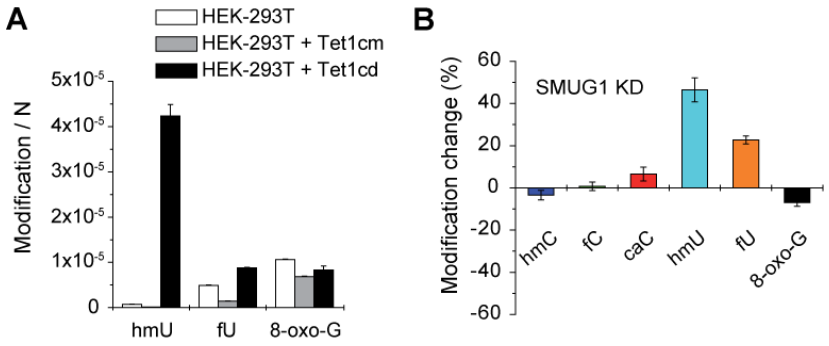


Abbildung 6 A) hmU-Level in HEK-293T-Zellen bei Überexpression der katalytischen Domäne von Tet1 (Tet1cd) bzw. einer katalytisch inaktiven Variante (Tet1cm); B) Effekt eines SMUG1-*Knockdowns* auf die DNA-Modifikationslevel.

1 Introduction

Nucleosides are fundamental molecules of life which constitute the building blocks of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is responsible for storage of the genetic information, while RNA, transcribed from DNA, performs a wide range of functions, from transport and translation of the genetic information to catalysis and modulation of gene expression.

The four canonical DNA nucleosides are deoxyadenosine (dA), deoxycytosine (dC), deoxyguanosine (dG) and deoxythymidine (dT), while RNA is composed of the corresponding canonical ribonucleosides A, C, G and U. However, in both cases there are additional non-canonical nucleosides. In the case of RNA these range from simple methylations to highly complex modifications and they play a role in fine-tuning of the RNA structure and its catalytic properties. In the case of DNA, the structural complexity of the modified nucleosides is reduced compared to RNA, but, nevertheless, modified DNA nucleosides play an essential function in mammalian epigenetic gene regulation.

1.1 RNA modifications

So far, more than 100 RNA modifications have been identified. Modifications have been found for all four canonical nucleosides and can range from simple methylations, such as for N^1-methylguanosine (m^1G), to more complex hypermodified nucleosides, such as wybutosine (yW) or queuosine (Q).[1] The majority of RNA modifications are found in tRNA, where modifications also display the largest diversity and complexity of structures. However, modified RNA nucleosides have also been found in rRNA and mRNA, as well as in small non-coding RNA species. Recently, the development of more sensitive detection methods, as well as the improvement of the RNA isolation protocols, have allowed the detection of 14 new RNA modifications, including agmatidine (C^+) and 8-methyladenosine (m^8A). Further technological advances will most likely lead to the discovery of new modified nucleosides in the coming years.[2]

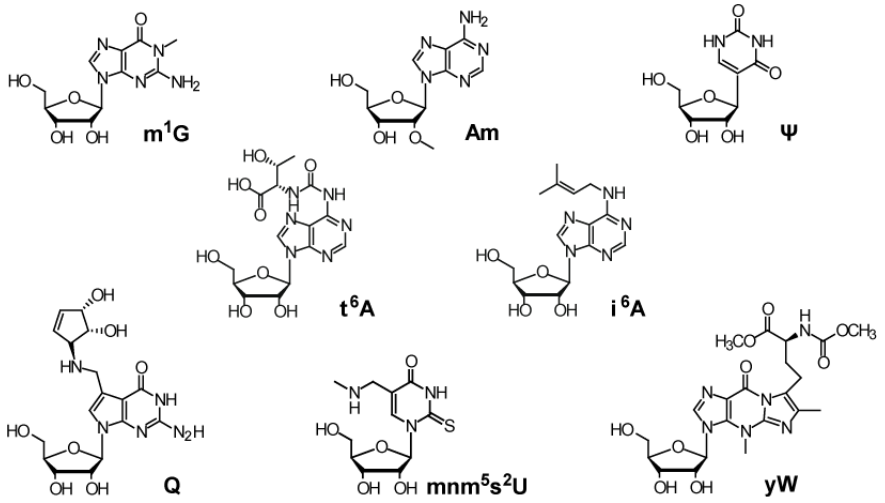


Figure 1 Examples of naturally occurring RNA modifications.

1.1.1 Transfer RNA

Transfer RNA (tRNA) is an adapter molecule that links each mRNA codon triplet to the corresponding amino acid during protein translation in the ribosome. tRNA molecules are composed of three main loops, an amino acid acceptor stem and a variable loop. All tRNAs adopt a characteristic L-shaped tri-dimensional structure which allows for optimal fitting in the ribosomal active site (see Figure 2). The CCA tail, located at the 3´-end of the tRNA molecule, is crucial for amino acid loading, while the dihydrouridine stem and loop (DSL) and the thymidine stem and loop (TSL) contribute to the 3D structure.[3] The anticodon stem and loop (ASL) is found at the end of the L-shaped structure and it contains the three-nucleotide sequence which is required for codon recognition and protein synthesis.

Modifications are found in various positions of the tRNA molecule, with an average of 10-15% of the tRNA nucleosides being modified, and they are essential for correct tRNA folding and maturation.[4] Most modifications outside of the anticodon loop are simple modifications such as methylated nucleosides or pseudouridines (Ψ). On the contrary, the anticodon loop (position 32 to 38) displays the largest variety of modifications, in particular at position 34, known as the wobble position, and at position 37, directly adjacent to the anticodon triplet (Figure 2).[5] These two positions are crucial for modulation of codon-anticodon interactions at the ribosome A- and P-sites.[6] Position 32 is a conserved pyrimidine position, generally unmodified or containing simple modifications (e.g. methylations or Ψ), while position 33 is almost exclusively occupied by a uridine.[5]

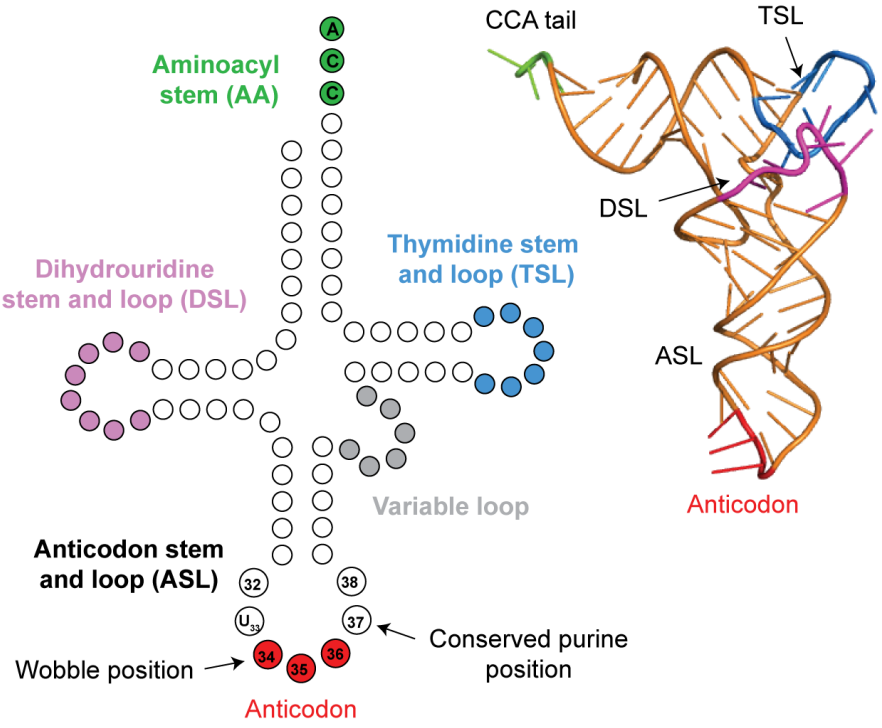


Figure 2 Schematic representation of tRNA cloverleaf structure and 3D structure of yeast tRNA^{Phe}.[7] In both representations characteristic features of the tRNA molecule are highlighted: DSL (violet), TSL (blue), CCA-tail (green), anticodon (red). Structural data for yeast tRNA^{Phe} was derived from the crystal structure 1EHZ.[7]

1.1.2 Functions of tRNA modifications

RNA modifications, depending on their nature and location in the tRNA molecules, have three main functions: modulation of the codon-anticodon interaction, folding and structural stabilization and, finally, recognition of the specific tRNA molecule.

1.1.2.1 Modulation of codon-anticodon interaction

The fast rate of protein synthesis (20-40 peptide bonds per second[8]) and the high fidelity of mRNA translation (1 error per $1x10^3$–$1x10^4$ amino acids[9]) requires precise and efficient cognate tRNA selection. Binding of the correct tRNA at the ribosome A-site is verified by kinetic and induced-fit proofreading mechanisms. Upon entering of the aminoacyl-tRNA, elongation factor and GTP in the ribosome A-site, the anticodon forms a minihelix with the codon triplet. The minihelix is evaluated by formation of specific H-bonds between ribosomal A1492, A1493 and G530 and the backbones of the codon and anticodon RNA strands. Incorrect H-bond formation arising from non-cognate tRNA results in release of the tRNA molecule, while correct codon-anticodon pairing (or pairing to near-cognate tRNA) leads to GTP hydrolysis, release of the elongation factor and translocation of the tRNA from the A- to the P-site of the ribosome.[10]

Modifications involved in codon-anticodon interaction contribute to cognate-codon selection in the A-site, as well as to stabilization of the codon-anticodon binding and frameshift prevention.[11] These modifications are generally located in the anticodon loop, and the major contributions are given by the modifications at position 34 and position 37.

Position 34, known as the wobble position, is involved in decoding of degenerate codons which correspond to a single amino acid. In fact, some amino acids are coded for by as many as six codon triplets, while others are coded by only one or two triplets. Translation is performed by a limited set of tRNAs, which in some cases recognize more than one codon sequence by wobbling of the third codon-anticodon base pair. Wobble base-pairing accounts for up to 30-40 % of all codon recognition.[12] The wobble position often consists of a large variety of uridine derivatives, inosine (I), 2´-*O*-methylated nucleosides and hypermodified nucleosides belonging to the queuosine family.[5, 13] Depending on the nature of the modification, wobble base pairing can be extended to all four nucleobases or restricted to only two. An interesting example is provided by the uridine derivatives. In fact, while s^2U34 enhances base pairing to A and xm^5U34 (such as mnm^5U34) allows for efficient pairing to codons ending in A and G, xo^5U34 (e.g. cmo^5U34) enables a single tRNA isoacceptor to read three or four synonymous codons, i.e. NNX (X = A/C/G/U).[12, 14] The different wobbling properties have been explained in terms of dynamic conformations of the modified nucleotides, where in this case the 2-thio group limits the uridine to *anti*, 3´-endo, gauche$^+$ conformation, promoting binding to adenosine, while 5-oxy derivatives can adopt both the C2´-endo as well as the C3´-endo conformation, allowing a wider pairing ability.[15] In support of the major role played by wobble base

modifications, lack of modification at the U34 position of ASLs of glutamine, lysine, arginine and cysteine results in impaired codon binding.[11a, 16]

Purines found at position 37 of the ASL are generally modified and play a key role in codon-anticodon interaction.[6] Prominent modifications are N^6-threonylcarbanyladenosine (t^6A), N^2-methylguanosine and 2-methylthio-N^6-isopentenyladenosine (ms^2i^6A).[1, 5] Upon codon-binding in the A-site of the ribosome, purine 37 stacks on top of the first codon-anticodon base pair, as shown for t^6A in Figure 3, where the modified nucleoside stacks both with the base at position 38 as well as with the first base of the mRNA codon.[17] This stabilisation effect is particularly important for U/A-rich codon-anticodon pairs, and t^6A37 in the ASL of $tRNA^{Lys(UUU)}$ is an essential requirement for codon-binding in the A-site.[16b, 17]

The enhanced stability of the codon-anticodon interaction afforded by modifications at position 37 (and/or at the wobble position) also prevents +1 frameshift events by favouring cognate tRNA selection. Furthermore, it reduces pausing in the A-site and prevents slippage of the peptidyl-tRNA in the P-site of the ribosome.[18] For example, unmodified A37 in $tRNA^{Tyr(QUA)}$ resulted in up to 9-fold increase in frameshifting compared to tRNA containing the modification ms^2io^6A37. Similarly, lack of the mnm^5 modification at U34 in $tRNA^{Lys(UUU)}$ resulted in nearly three-fold increase in +1 frameshifting when paired with codon AAA, and almost six-fold increase in frameshifting when paired with the synonymous codon AAG.[18a]

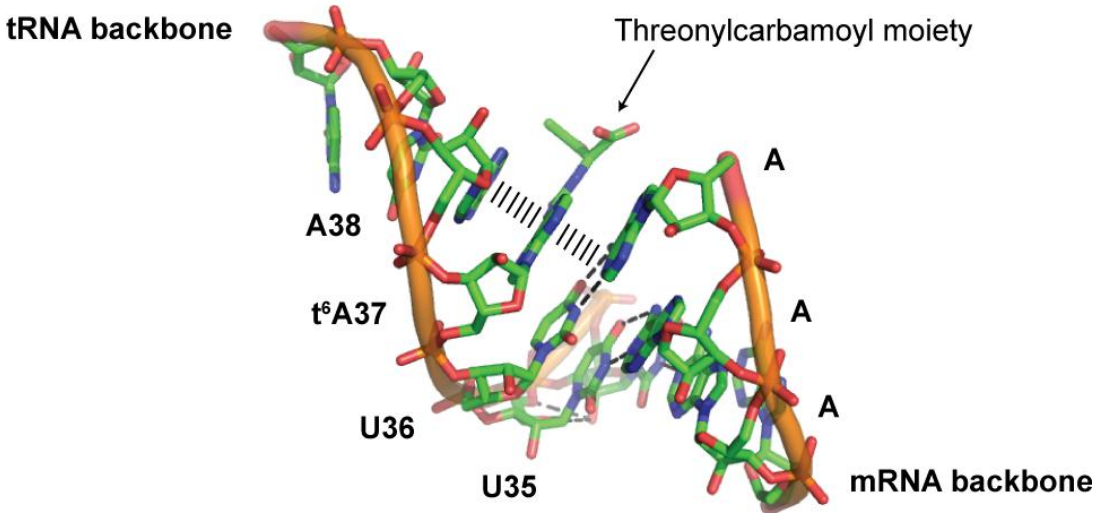


Figure 3 Modulation of codon-anticodon interaction. Modification t^6A at position 37 of $tRNA^{Lys(UUU)}$ enhances interaction with codon AAA in the A site of the ribosome. Stacking of the purine moiety of A37 with neighbouring bases is shown by a wide dotted line. Data was derived from crystal structure 1XMQ.[17]

Another important function of modified purine 37 nucleosides is to structure the anticodon loop for optimal fitting in the ribosomal A- and P-sites. In fact, as shown in Figure 4, modification at position 37 prevents intra-loop H-bonding, ensuring an ´open loop´ conformation of the ASL.[19] The absence of intra-loop base pairing leads to a reduction of the melting temperature of the ASL compared to that

of unmodified RNA. However, as evident from solution-structure analyses and X-ray crystallographic structures of tRNAs,[11b] it also results in a constrained architecture which is optimal for fitting in the ribosomal A-site, therefore reducing the entropic penalty upon binding of the tRNA. This, in turn, leads to enhanced accuracy and faster rates of cognate tRNA selection, given that the optimal ASL as well as the stabilising effect of modified nucleosides in the extended anticodon domain ensure more efficient selection of the tRNA molecules.[11b]

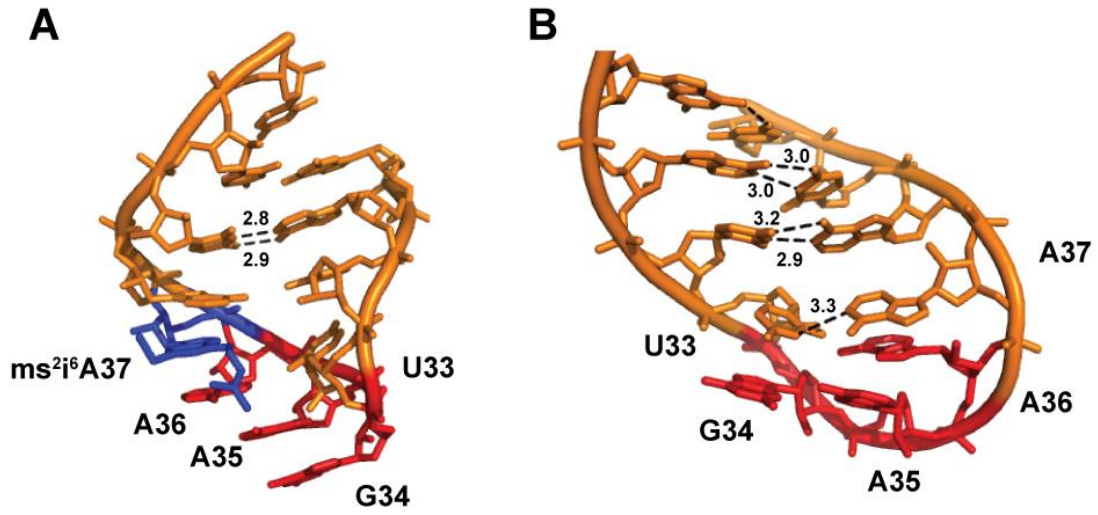


Figure 4 Structural role of modifications at position 37. A) *E.coli* tRNA^{Phe} anticodon loop conformation in the presence of ms^{2}i^{6}A modification at position 37 (PDB 318G).[20] B) *E.coli* tRNA^{Phe} unmodified ASL showing intraloop Watson-Crick bases between unmodified A37 and U33, resulting in a compact three nucleotide loop (PDB 1J4Y).[21]

1.1.2.2 Folding and structural stabilisation

tRNA modifications also contribute to folding and stabilisation of the tRNA molecule in various ways. As mentioned in the previous chapter, an important contribution to conformational stability is introduced by modifications at purine 37, which prearrange the ASL for optimal fitting in the A site.[11b] Modifications located in the core of the tRNA, especially 2´-O-methylations, pseudouridinylations or thiolations, further stabilise the tRNA molecule by favouring the 3´-endo conformation of the ribose, which is the preferred conformation adopted upon hybridization to A-form helices.[4b, 22] The additional stabilisation afforded by modifications also results in an increase of the melting temperature (Tm),[23] as exemplified by studies with yeast tRNA^{Phe}, where modifications led to a Tm of 6 °C higher compared to unmodified tRNA.[24] In some cases, modifications contribute to larger structural effects, as in the case of mitochondrial tRNA^{Lys}, where m^{1}A at position 9 is essential for correct folding of the tRNA molecule.[25]

Presence of specific modifications is also necessary for maturation of tRNA molecules, and absence of these can lead to rapid degradation of the tRNA. This is the case for m^{1}A at position 58 of tRNA$_i^{Met}$, where this modification is required for maturation and stability of the corresponding pre-tRNA, which is otherwise rapidly degraded by a *TRF4/RRP6*-dependent nuclear surveillance pathway.[26]

Furthermore, absence of certain modification pairs in yeast tRNA, such as m^7G46 and m^5C,[4a] results in temperature-sensitive growth and degradation of $tRNA^{Val(AAC)}$ by a rapid tRNA degradation (RTD) pathway, again suggesting a key role for specific modification networks in maturation and stabilisation of the tRNA.[4a, 27]

1.1.2.3 tRNA recognition

Modifications also contribute in defining the identity of tRNA molecules in order to ensure selective interaction with enzymes involved in tRNA processing and recruitment to the translational machinery.[28] In particular, various modifications have been implicated in the selective recognition by amino acyl synthetases. For examples, m^1G strongly reduces misacylation of $tRNA^{Asp}$ with an Arg by arginyl-tRNA synthetase (ArgRS).[29] Another interesting example is that of lysidine k^2C34 of *E.coli* $tRNA^{Ile(CAU)}$. Selective interaction of IleRS with the lysidine modification at the anticodon ensures correct charging with isoleucine, while the unmodified tRNA would be mischarged with methionine by MetRS.[30]

Modifications also play a role in discrimination between initiator and elongator $tRNA^{Met}$ in fungi and plants. The steric bulk resulting from the modification 2´-O-ribosyladenosine Ar (p) or 2´-O-ribosylguanosine Gr(p) located at position 64 of initiator $tRNA_i^{Met}$ prevents ternary complex formation with elongator factor eEF-1α and GTP, therefore preventing $tRNA_i^{Met}$ from contributing to the elongation process.[31]

1.1.3 Biogenesis of tRNAs and tRNA modifications

tRNAs and their modifications have a complex, stepwise biosynthetic pathway. In yeast, pre-tRNAs are transcribed in the nucleolus,[32] where they also undergo 5´-processing.[33] The 3´ trailer sequence is subsequently cleaved, most likely in the nucleoplasm, followed by addition of the CCA sequence at the 3´-end of the molecule.[34] Furthermore, intron-containing tRNAs undergo splicing either in the nucleus or in the cytoplasm depending on the organism.[34a]

Modification of tRNA molecules occurs in a stepwise manner following a strict order.[35] The main factors governing the tRNA modification process are substrate specificity as well as subcellular localization of the modification enzymes.[34a]

Substrate specificity dictates the timing of the modification process depending on the extent of 5´- and 3´-end processing and on the presence or absence of an intron. Certain modifying enzymes depend specifically on the local architecture of intron-containing tRNAs, as in the case of pseudouridines at positions 34, 35 and 36 as well as m^5C34 of various eukaryotic tRNAs.[36] Similarly, studies performed with yeast $tRNA^{Phe}$ revealed the intron-dependent formation of m^5C40. The same study also identified a subset of modifications which requires prior intron splicing, specifically the anticodon modifications Cm32, Gm34 and m^1G37, and another subset which is introduced independently of the presence of an intron, which includes a number of modifications outside of the anticodon stem-loop.[37]

Timing of tRNA modification is also affected by the cellular distribution of modifying enzymes, as in the case of yeast $tRNA^{Phe}$ (Figure 5).[28] In this case, after initial processing and modification in the nucleus, the intron-containing tRNA is exported to the cytoplasm, where splicing takes place at the mitochondrial outer membrane.[38] Modifications m^2G10, Cm32 and Gm34 are subsequently introduced in the cytoplasm, prior to import of the tRNA into the nucleus, where m^1G37 is generated. $tRNA^{Phe}$ is then re-exported to the cytoplasm where biosynthesis of yW at position 37 takes place and the mature tRNA can be amino acylated and recruited to the translational machinery.[34b] In relation to tRNA transport, it should be noted that retrograde transport of tRNA molecules into the nucleus is not only occurring during tRNA maturation, but it has also been observed to involve mature tRNAs, particularly in response to nutrient deprivation.[39] Retrograde transport of mature tRNAs is reversible and it is suggested to play a role in proofreading of tRNAs or in downregulation of protein translation.[39-40]

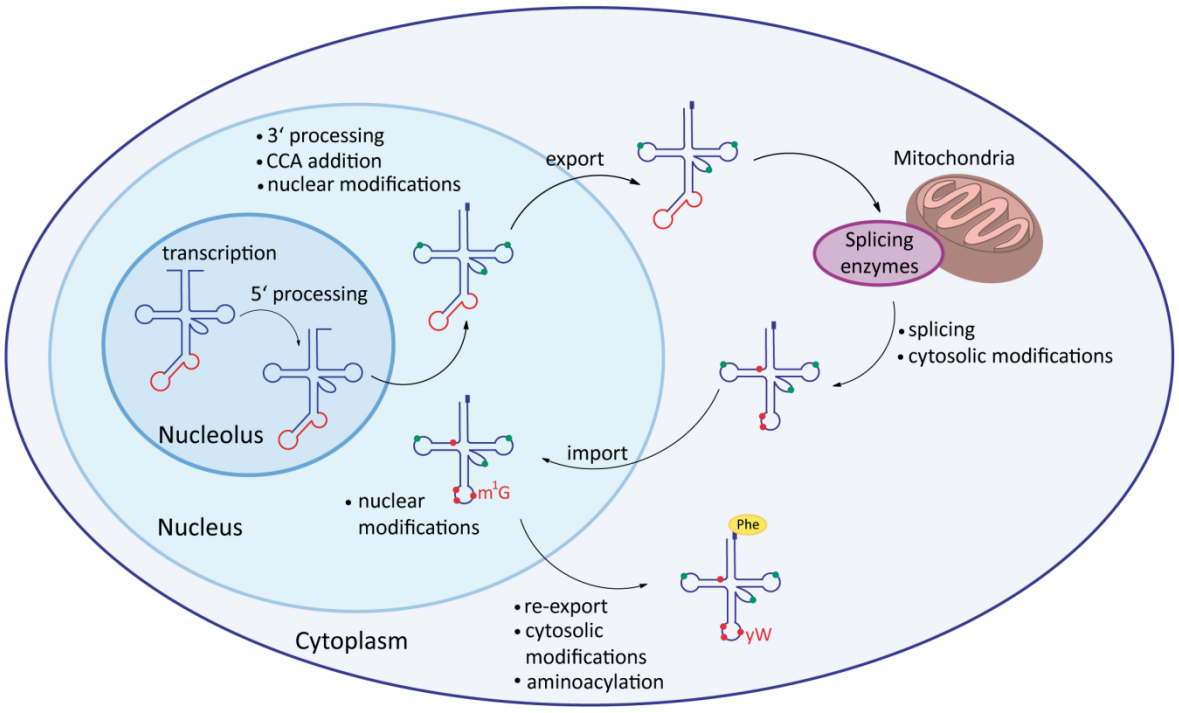


Figure 5 Biosynthesis of yeast $tRNA^{Phe}$. Maturation of $tRNA^{Phe}$ in *S.cerevisiae* depicting transport between the nucleus and the cytosol, alongside the various processing steps (5´ and 3´ end processing, splicing, modifications).[28]

However, tRNA modifications can also be governed by subtle architectural aspects, such as the presence of other specific modifications in the tRNA molecule.[4c] One such example is found in *Trypanosoma brucei*, where C to U editing at the first position of the anticodon of $tRNA^{Trp}$ was found to be negatively regulated by the presence of s^2U33, suggesting that both the CCA and the edited UCA anticodons might have a distinct biological role.[41]

Incorrect processing and maturation of tRNA molecules results in degradation of the tRNA. Various degradation pathways have been described,[34a] one involving the TRAMP complex, which functions

in nuclear surveillance, and one being the Rapid tRNA Decay (RTD), which specifically targets certain hypomodified tRNAs.[4a, 34a] Furthermore, cells have been found to respond to various stress conditions by inducing endonucleolytic cleavage of tRNAs at the anticodon loop.[42] The resulting tRNA halves have been suggested to play a role in modulation of protein synthesis, either by blocking the translational machinery, or by acting as siRNA- or miRNA-like templates for selective mRNA degradation. Yet, the exact role of tRNA endonucleolytic cleavage and of the resulting tRNA fragments is still unclear.[43]

1.1.4 Regulation of tRNA modifications

Emerging evidence suggests a role for tRNAs and tRNA modifications in cellular response to stress.[43-44] A study performed by Chan *et al*. revealed a highly dynamic reprogramming of tRNA modifications in response to various stresses, with some modifications increasing and other being depleted.[44a] In a subsequent report, further investigation of the response of yeast cells to H_2O_2 exposure highlighted the crucial role of Trm4 and the resulting m^5C34 modification for modulation of the response mechanism.[45] In fact, oxidative stress triggers expression of a subset of genes enriched in TTG codons. The selective increase in m^5C modification at the wobble position of $tRNA^{Leu(CAA)}$ ensures efficient translation of this gene pool. Loss of Trm4 resulted in hypersensitivity to oxidative stress, confirming the essential role of the modification m^5C34 in modulating stress response.[45] Similarly, Trm9, which catalyses the methylation of uridine derivative mcm^5U34 and mcm^5s^2U34, was found to contribute to stress response against methyl methanesulfonate (MMS).[46] Here again, Trm9-dependent modifications were observed to modulate efficient translation of arginine and glutamic acid codons which are selectively enriched in genes actively transcribed during the response process.[46] In line with these findings, recent studies highlighted that specific non-optimal codon compositions are enriched in subsets of genes involved in the same pathway, as an optimization strategy to enhance coordinate protein synthesis for these genes. Such variations in codon content are suggested to be involved in modulation of cell-cycle dependent fluctuations in protein levels[47] as well as in the specific transcription of genes involved in stress response.[46] Modulation of tRNA modifications to selectively favour translation of certain codons might therefore provide an avenue for efficient translation of these subsets of genes.[48]

Specific tRNA modifications are also enhanced for structural stabilization in response to heat stress in thermophiles.[4b, 49] Examples include the modifications Gm, $m^2{}_2Gm$, ac^4Cm, s^2T and m^1A. The enhanced thermal stabilization conferred by these modifications allows growth at elevated temperatures, which is otherwise impaired in null mutants.[50]

1.1.5 tRNA modifications in evolution

tRNAs from different organisms contain a variety of different RNA modifications at different positions, suggesting that the nature of transfer RNA modifications as well as their patterns in the tRNA molecule might depend on the origin of the tRNA. However, certain modifications are found

across all three domains, Eukarya, Bacteria and Archaea,[5] suggesting a very early function of these modifications in evolution.[51] Efficient modulation of protein translation and maintenance of the reading frame is an essential feature for survival of the organism, and must have therefore exerted a strong selective pressure in optimization of the tRNA molecules.[52] Indeed, modification m^1G37, which is found across all three kingdoms, was identified as an essential feature for cellular life throughout evolution and was suggested to have evolved prior to divergence of the kingdoms.[52] m^1G37 was in fact shown to play a major role in maintaining the reading frame and lack of this modification is known to result in severe growth impairment in both bacteria and eukaryotes.[18a, 52] A similar scenario was proposed by Urbonavicius *et al*., which suggested that modifications involved in preventing +1 frameshifting must have played a crucial role throughout evolution.[18a] However, while specific modifications involved in modulation of translation might have indeed evolved for optimized protein synthesis, Phizicky *et al*. have pointed out that other modifications, particularly those outside of the anticodon loop, might have arisen from overlapping reactivity and poor substrate specificity of the modification enzymes rather than as an evolutionary demand.[4c] This might also explain the lack of phenotype observed in mutant strains lacking certain modifications in the tRNA core structure.[4a, 4c]

Interestingly, a recent study further elucidated the role of two tRNA-modifying enzymes in the evolution of tRNA isoacceptor populations across the three kingdoms. According to this report, appearance of eukaryotic tRNA-dependent adenosine deaminases (hetADAT), which catalyse conversion of adenosine 34 to inosine 34, and of bacterial uridine methyltransferases (UMs) which modify xo^5U34 derivatives, was found to have contributed to divergence of the tRNA populations of Eukarya and Bacteria with respect to the genome of Archaea. In fact, the improved decoding ability of specific tRNA isoacceptors enabled by these modifications led to an enrichment of these isoacceptors, therefore contributing to the evolution of genomic codon composition and to differences in tRNA gene populations.[53]

1.2 DNA modifications

Modified DNA nucleosides display a lower degree of structural complexity compared to RNA nucleosides. Of crucial importance are the modified cytosine derivatives found in eukaryotic genomic DNA, which include methylcytosine (mC), hydroxymethylcytosine (hmC), formylcytosine (fC) and carboxycytosine (caC). In particular, mC is known to be a key modification involved in epigenetic regulation of gene expression, while the growing number of studies focusing on hmC, fC and caC are slowly uncovering their roles as potential epigenetic markers and/or intermediates in various pathways of demethylation.

1.2.1 Epigenetic modulation of gene expression

Multicellular organisms are composed of different cell types with vastly different characteristics despite all cells sharing the same genetic information. These characteristics are defined by different

sets of proteins expressed and therefore by different sets of genes activated depending on the cell type. A crucial step in the regulation of gene expression is transcription, which is generally regulated by sequence-specific transcription factors as well as by cis-acting regulatory elements such as promoters and enhancers. Nevertheless, given the complexity of the mammalian gene expression system, there is need for an extra layer of regulation which ensures durable gene expression patterns and cellular identity, while at the same time allowing for cells to respond to environmental and developmental stimuli. This additional level of regulation is introduced by histone modifications, which are responsible for short-term modulation of gene-expression, while DNA methylation confers long-term gene silencing.[54] These regulatory features, histone and DNA modifications, are by definition epigenetic, given that they alter gene expression without changing the DNA sequence.

1.2.2 Chromatin structure and histone modifications

Eukaryotic DNA is organized in a hierarchy of structural levels closely associated with specific DNA binding proteins. This combination of DNA and proteins is known as chromatin and it is essential for packaging of the DNA as well as for modulation of gene expression. The smallest packaging unit is known as nucleosome, where a 147-bp DNA segment is coiled in 1.65 turns around a histone octamer, composed by two H2A-H2B heterodimers and by a H3-H4 tetramer.[55] Core histones are globular proteins with an unstructured N-terminal tail which can be covalently modified. Nucleosomes are separated by about 50-bp-long DNA segments. This loosely packed chromatin form is known as ´beads on a string´ and it is generally associated with active transcription. Further packaging is introduced by the linker histone H1, which binds between nucleosomes generating the so called ´30-nm fiber´, and nucleosomes are finally organized in various higher order chromatin architectures.[56] According to its structure, chromatin is divided into heterochromatin, which is more densely packed and generally (but not exclusively) transcriptionally inactive, and euchromatin, more dispersed and commonly regarded as transcriptionally active.[57] Most importantly, chromatin packaging is a highly dynamic process, with profound implications for gene expression.

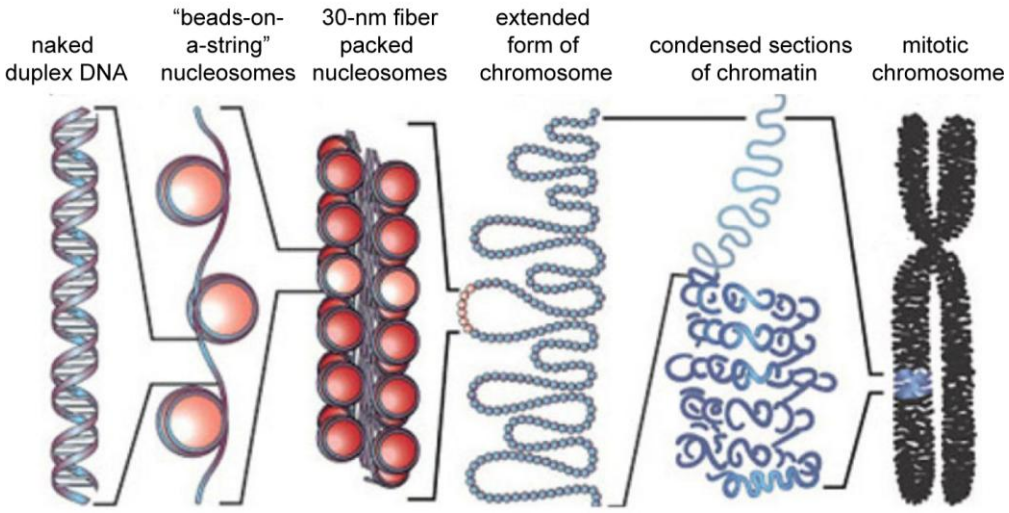


Figure 6 Structural organization of eukaryotic DNA. (Adapted from Felsenfeld and Groundine, 2003[56b]**)**

Histones play a crucial role in modulation of the dynamics of chromatin accessibility. Histones can in fact be dynamically functionalized with a variety of post-translational modifications. These include acetylation, methylation, phosphorylation as well as more complex modifications such as SUMOylation or ADP-ribosylation. These modifications have important roles in gene expression, replication and DNA repair and they collectively contribute to defining the ´histone code´.[58] Histone modifications affect chromatin packaging by two main mechanisms: they can either disrupt nucleosome interactions, or they can mediate the recruitment of specific non-histone proteins.[59] For example, acetylation of lysine residues leads to charge neutralization, therefore weakening the charge-dependent nucleosome-DNA interactions and increasing chromatin accessibility. Charge-effects are also suggested to take place upon phosphorylation.[58a] On the other hand, various proteins have been found to interact with or to be affected by specific histone modifications, contributing to modulation of downstream events.[60] For example, DNMT3L-dependent targeting of *de novo* DNA methylation by interaction with H3 tails is inhibited by H3K4me.[61] In addition, specific histone methyltransferases were found to target DNA methylation to defined genomic sections by recruitment of DNA methyltransferases[62] while, on the other hand, DNA methylation was similarly observed to direct histone modifications.[63]

In general, transcriptionally active chromatin is associated with acetylation of histones H3 and H4 and trimethylation of H3K4, H3K36 and H3K79, while heterochromatin generally correlates with low levels of acetylation and high content of methylated H3K9, H3K27 and H4K20.[64] Bivalent domains have also been found, possessing both activating and repressive modifications.[65] In fact, histones can be modified at multiple sites simultaneously, contributing to a complex crosstalk between different epigenetic marks.

1.2.3 DNA methylation

1.2.3.1 Distribution of DNA methylation

In the genome, cytosine methylation takes place predominantly in a CpG dinucleotide context. CpG dinucleotides are underrepresented in the genome, possibly because they constitute mutation hot spots,[66] and they tend to cluster in regions known as CpG islands (CGIs). GCIs are defined as DNA segments of at least 200 bp with a high C/G content (minimum 50 %) and a ratio of observed to statistically predicted CpG frequency of at least 0.6.[67] CGIs are associated with about 60-70 % of human gene promoters. Of these, the majority is usually unmethylated,[68] while about 6 % of them is methylated in a tissue-specific manner at early stages of development or in differentiated tissues.[69]

In general, promoter CGIs methylation inversely correlates with gene expression. Methylation-dependent transcriptional repression can be induced by two main mechanism: methylation can lead to the recruitment of proteins which specifically bind methylated CpG sites, therefore preventing binding of transcription factors at these sites;[70] alternatively, methyl-CpG-binding proteins can themselves

recruit chromatin remodelling factors and histone modifiers which lead to repression of transcription.[71] Methylation of promoter CGIs is generally associated with long-term silencing[72] and it plays an essential role in genomic imprinting[73] and X-chromosome inactivation.[74]

However, DNA methylation does not only occur at CGIs and it does not exclusively correlate with transcriptional silencing. In fact, sparsely distributed CpGs found in gene bodies are generally highly methylated[72] and methylation at these sites was found to positively correlate with active transcription.[75] DNA methylation can also occur in regions in close proximity to CGIs, known as CpG island shores, and in this case it is closely associated with transcriptional repression.[76] Furthermore, DNA methylation has also been observed in a CHH and CHG context, where H is A, C or T.[77] In fact, about one quarter of the methylated cytosines present is embryonic stem cells is found in a non-CpG context, with enrichment of these methylated cytosines at gene bodies, showing a positive correlation with gene expression, and depletion at protein binding sites and enhancers.[77b] Interestingly, levels of non-CpG methylcytosine were found to decrease during differentiation and to be restored in induced pluripotent stem cells, suggesting that stem cells might rely on a different mechanism for modulation of gene expression.[77b] Finally, methylation is not only involved in regulation of gene transcription, but also in genome stabilization. In particular, high methylation levels found at repetitive elements are thought to prevent reactivation of endoparasitic sequences that would otherwise result in translocations and chromosomal instability.[78]

DNA methylation plays a key role in mammalian development and large changes in methylation patterns are observed throughout differentiation.[79] A first wave of genome-wide epigenetic reprogramming is observed after fertilization, when the genome of the male pronucleus undergoes rapid demethylation (as a result of genome-wide oxidation of mC), followed by passive demethylation of the maternal genome. A similar wave of demethylation is observed during specification of primordial germ cells.[79b] Subsequently, during development, establishment of methylation patterns by *de novo* methyltransferases is essential for correct differentiation, contributing to the long-term repression of key pluripotency genes *Oct4* and *Nanog* and to tissue-specific modulation of gene expression.[79a]

Importantly, misregulation of methylation patterns is strongly linked to diseases.[80] Cancer cells feature genome-wide hypomethylation, which promotes chromosomal instability,[81] as well as local hypermethylation at CpG islands, which often leads to inactivation of genes involved in a variety of essential pathways, such as DNA repair, cell cycle control and apoptosis.[82] Misregulation of DNA methylation has also been implicated in a variety of other diseases, including ICF (Immune deficiency, centromeric instability and facial abnormalities) and Rett syndrome.[80] Interestingly, inhibitors of DNA methyltransferases, such as azacytidine, have been approved as drugs for cancer treatment, once again highlighting the essential role of correct methylation in maintenance of the normal cell functioning.[83]

1.2.3.2 DNA Methyltransferases

Methylation of cytosine is performed by DNA methyltransferases (Dnmts) in the presence of cofactor *S*-adenosylmethionine (SAM). In mammals, three Dnmts display DNA methyltransferase activity, namely Dnmt1, Dnmt3a and Dnm3b (Figure 7). A fourth member, Dnmt3L, lacks catalytic activity, although it was shown to interact and stimulate activity of Dnmt3a and Dnmt3b.[84] Finally, Dnmt2, despite sharing all the conserved catalytic motifs, has only limited activity as DNA methyltransferase[85] and was instead reported to act as a tRNA methyltransferase.[86]

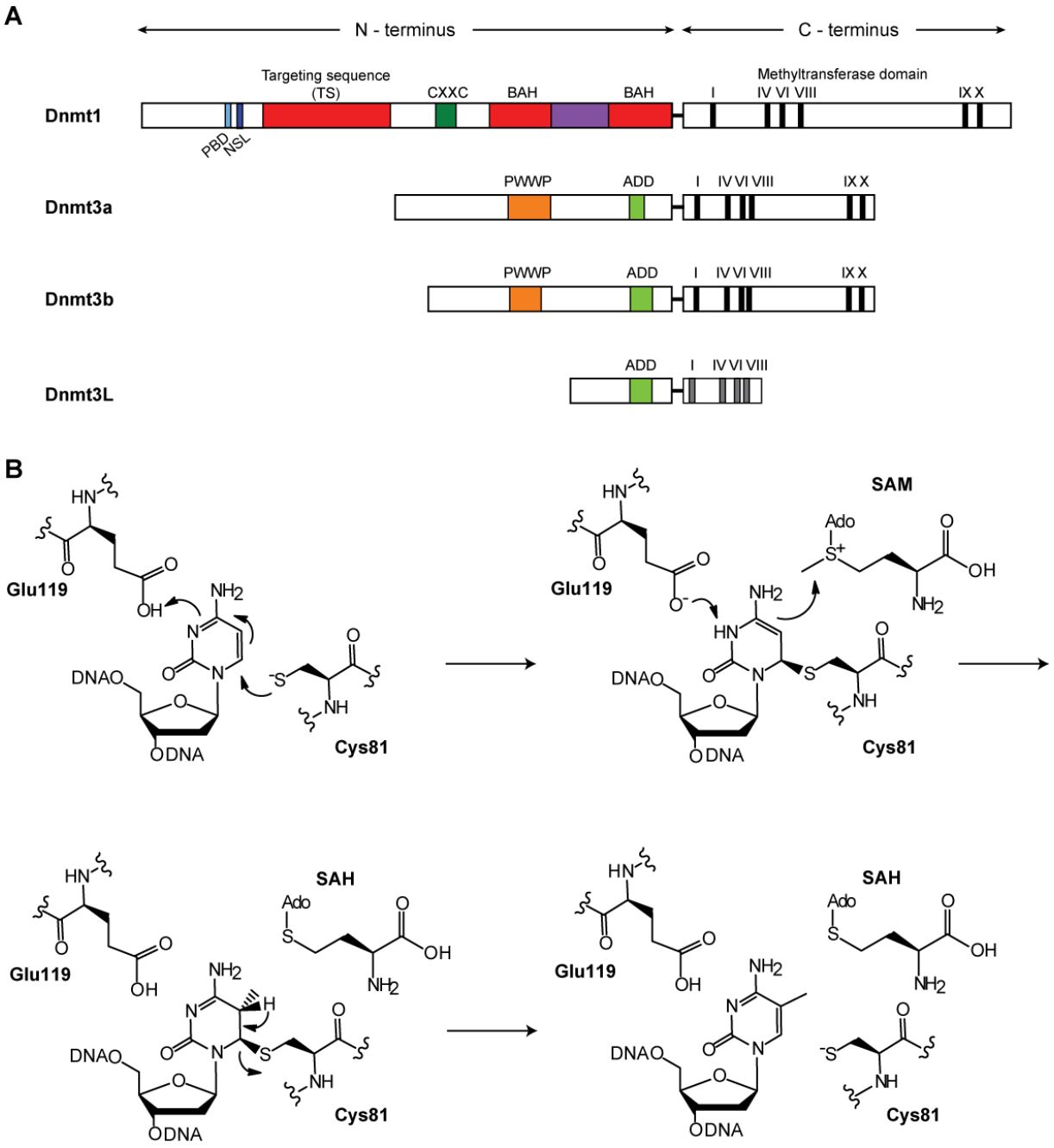


Figure 7 A) Schematic representation of domain structure of mammalian DNA methyltransferases and B) their catalytic mechanism. Mechanism of methyl transfer reaction is based on that of the bacterial methyltransferase M.Hha1.[87] SAM: *S*-adenosylmethionine; SAH: *S*-adenosylhomocystein.

Dnmt3a and Dnmt3b are known as *de novo* methyltransferases and are responsible for the establishment of the DNA methylation pattern during embryogenesis and development of germ cells.[88] Expression levels of the *de novo* methyltransferases are elevated in embryonic stem cells, while differentiation leads to their down-regulation and only low expression levels are observed in somatic tissues.[88] Both Dnmt3a and Dnmt3b are essential for mammalian development. *Dnmt3b* null mouse embryos show developmental defects after E 9.5 and no viable embryo can develop to term.[88a] On the contrary, *Dnmt3a* null embryos are able to develop to term, but show significant growth defects and generally die four weeks after birth.[88a] The different phenotypes observed for the two *de novo* methyltransferases are suggestive of different roles for these two Dnmts. Dnmt3b plays a major role during early development, while Dnmt3a appears to be crucial at later developmental stages.[88a] Similarly, *Dnmt3a* conditional mutant mice show imprinting defects in the germline, while no apparent phenotype is observed for *Dnmt3b* conditional mutants, suggesting that Dnmt3a is crucial for the establishment of maternal and paternal imprints.[89]

After establishment of DNA methylation by *de novo* methyltransferases, maintenance of this pattern is performed by Dnmt1.[90] Dnmt1 has a higher affinity for hemimethylated DNA[91] and is associated with the replication machinery to ensure efficient maintenance of methylation.[90, 92] Furthermore, the CXXC domain of Dnmt1 is reported to selectively bind to unmethylated CpG sites[93] and to result in a conformation which prevents catalysis on unmethylated DNA, thus providing an autoinhibitory mechanism that ensures methylation only in the presence of hemimethylated CpGs.[94] While *Dnmt1* null mutant stem cells are viable, with low but stable levels of methylation, loss of *Dnmt1* in mice results in embryonic lethality, highlighting the importance of maintenance of methylation for normal development.[95]

DNA methyltransferases are characterized by a large multidomain N-terminus, which is responsible for nuclear localization and modulation of protein-protein and protein-DNA interactions, and by a highly conserved C-terminus which harbours the catalytic methyltransferase domain and which is responsible for cofactor and substrate DNA binding and catalysis (Figure 7A).[96] All active methyltransferases share a common catalytic mechanism (Figure 7B):[97] initially, a conserved cysteine residue performs a nucleophilic attack at the C6 position of the cytosine (which is flipped out of the DNA double helix and buried into the hydrophobic pocket of the active site);[87] this reaction, which is facilitated by protonation at N3 by a conserved glutamate residue, is followed by attack at the methyl group of the SAM cofactor and by final re-aromatization of the methylated cytosine nucleobase via a *syn*-elimination of the enzyme cysteine.[87, 97]

1.2.4 Oxidized cytosine derivatives

In 2009 two groups independently reported the detection in mammalian genomic DNA of a further modified cytosine species, 5-hydroxymethylcytosine (hmC), which is now widely accepted as the sixth genomic nucleoside in higher eukaryotes.[98] A previous report had also documented the

detection of hmC in mammalian DNA, although the levels of hmC documented were substantially higher compared to more recent studies.[99] Recent reports additionally identified formylcytosine (fC) and carboxycytosine (caC) in genomic DNA of higher eukaryotes, bringing the total number of modified cytosines to four.[100]

A computational search to identify homologues of the trypanosome thymine hydroxylases JBP1 and JBP2 led to the discovery of the ten-eleven translocation (Tet) protein family,[98b] whose members were shown to catalyze the conversion of methylcytosine (mC) to hmC,[98b, 101] as well as the sequential oxidation of hmC to generate fC and caC.[100a, 100b]

The discovery of the Tet protein family and of the oxidized cytosine derivatives highlighted the possibility for a dynamic, Tet-mediated regulation of methylcytosine levels in genomic DNA and, since the first report by Tahiliani *et al*., it has been the focus of intensive research worldwide.[102]

1.2.4.1 Tet family proteins

In jawed vertebrates the Tet protein family is composed of Tet1, Tet2 and Tet3.[98b, 101] In all three proteins, the catalytic domain, composed of a Cys-rich region followed by a double-stranded β-helix (DSBH) domain, is located at the C-terminus and it displays 2-oxoglutarate (2-OG)- and iron (II)-dependent dioxygenase activity (Figure 8).[98b, 101] Additionally, Tet1 and Tet3 posses a CXXC domain at their N-terminus, while in the case of Tet2 a chromosomal inversion event led to the detachment of the CXXC domain.[103]. The CXXC domain of Tet1 has been reported to bind CpG-containing DNA in the presence of unmodified cytosine as well as of mC and hmC, a feature which might explain the enrichment of Tet1 at CpG-dense genomic regions,[104] while the CXXC domain of Tet3 is able to bind unmodified cytosine regardless of whether it is followed by a guanine.[105]

Tet1 is highly expressed in mouse ES cells and downregulated during differentiation and it is suggested to play a role in regulation of pluripotency and differentiation. In fact, a report by Ito *et al*. showed that shRNA-mediated knockdown of Tet1 in mouse ES cells results in downregulated expression of the pluripotency factor *Nanog* by increased methylation at the *Nanog*-proximal promoter,[101] and similar observations were made for other pluripotency-related genes,[106] although there are still discrepancies concerning these results.[107] Furthermore, Tet1 was suggested to be important for inner cell mass specification, as depletion of Tet1 levels in pre-implantation embryos leads to biased differentiation towards trophectoderm and mesoendoderm lineage.[101, 106-107] *Tet1*-null mice are viable and fertile, but tend to display smaller body size, suggesting a role of Tet1 in regulation of development.[108]

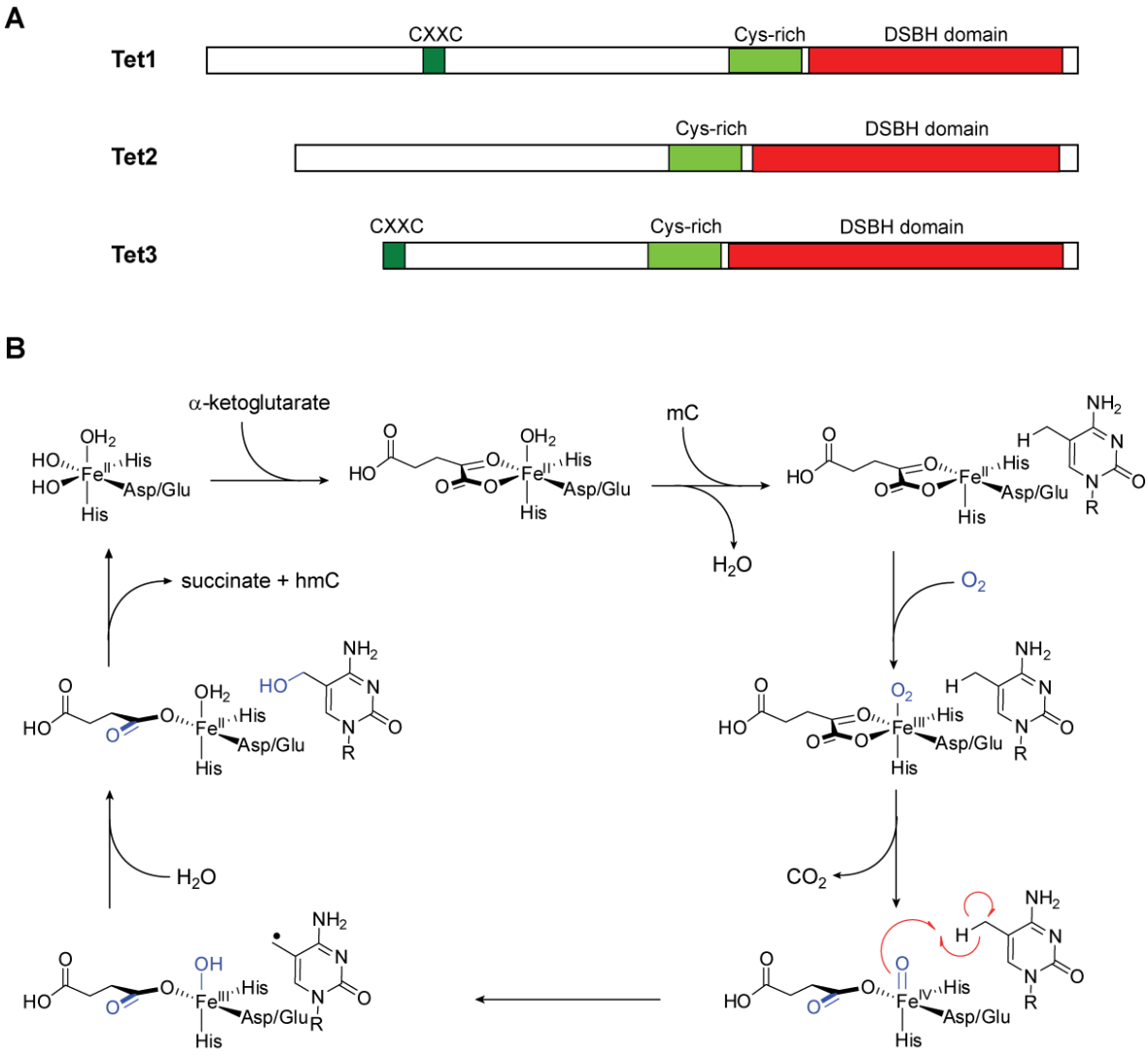


Figure 8 A) Summary of domain structure of mammalian Tet proteins and B) their catalytic mechanism.[109] DSBH: dioxygenase activity, metal binding domain.

Genome-wide mapping of Tet1 in ES cells highlighted a dual role of this protein in the regulation of transcription. Tet1 is enriched at gene promoters and exons,[106, 110] and it has been suggested to contribute to maintenance of a hypomethylated state for CpG-rich promoters.[104a, 106] However, although Tet1-dependent promoter hypomethylation is necessary for expression of a subset of transcriptionally active genes, Tet1 also plays a repressive role, especially for genes involved in the differentiation process.[110b, 111] In fact, Tet1 contributes to gene silencing by favouring recruitment of the Polycomp Repressive Complex 2 (PRC2) through maintenance of a hypomethylated state at the target genes (since binding of PRC2 is inhibited by the presence of mC).[111] Similarly, Tet1 was found to display a significant overlap of target genes with the SIN3A co-repressor complex, further supporting a role of Tet1 in transcriptional repression.[110b] The dual role of Tet1 highlighted by these studies is thought to be essential for correct maintenance of pluripotency and timely coordination of the differentiation process.

Tet2 is also expressed in ES cells, although its depletion results in milder effects compared to Tet1[101, 107] and it was not found to associate with SIN3A.[110b] Interestingly, Tet2 is also expressed in hematopoietic stem cells (HSCs) and it is suggested to play a crucial role in regulation of hematopoietic differentiation.[112] In fact, although *Tet2*-null mice are viable, they display a lower genomic hmC content and an enlarged HSCs pool, with a tendency to develop hematopoietic malignancies.[112a] Furthermore, mutations in *Tet2*, both leading to catalytic inactivation or unrelated to enzymatic activity, are amongst the most frequently encountered mutations in myeloid malignancies.[113]

In contrast to Tet1 and Tet2, Tet3 is poorly expressed in ES cells but is instead enriched in oocytes and zygotes. Here, Tet3 was shown to be responsible for the genome-wide oxidation of mC to hmC, fC and caC in the male pronucleus after fertilization.[114] *Tet3*-deficient zygotes fail to reduce global levels of mC in the paternal genome and depletion of Tet3 in the female germ line results in reduced fecundity and increased probability of developmental failure of embryos.[114a] Studies performed in *Xenopus laevis* also highlighted a role of Tet3 in early eye and neural development,[105] while depletion of Tet2 and Tet3 in developing mouse cortex resulted in impaired differentiation of neural progenitor cells.[115]

Tet proteins have also been reported to interact with the enzyme OGT (*O*-linked β-D-*N*-acetylglucosamine (*O*-GlcNAc) transferase), responsible for the *in vivo* addition of *O*-GlcNAc to serine and threonine residues. OGT, which is known to modify a variety of proteins including chromatin interactors and modifiers, was shown to be enriched at promoters in a Tet-dependent manner and to specifically trigger *O*-GlcNAcylation of histones and histone modifiers, therefore contributing to modulation of transcriptional activity.[116]

1.2.4.2 Distribution of DNA oxidized cytosine derivatives

Hydroxymethylcytosine has been detected in a variety of cell types, ranging from ES cells[98b] to Purkinje neurons[98a], as well as in mammalian tissues where hmC levels are found to be highest in the brain.[117] Formylcytosine and carboxycytosine have also been found in a range of mammalian cells, but while hmC levels are around 5-10 % of the mC content in ES cells, fC and caC are found in much lower amounts, about 0.1 and 0.01 % of the mC content, respectively.[100] Since 2009, the development of genome-wide mapping of hmC, fC and caC, both using precipitation-based and single-base resolution methods, has provided invaluable tools for the study of the distribution of modified cytosines in genomic DNA.[118]

Mapping of hmC revealed that it is mostly associated with euchromatic regions. Enrichment was detected at gene bodies, where gene expression is generally found to positively correlate with hmC levels, as well as at enhancers and at promoter regions with intermediate CpG content,[106, 110a, 119] Furthermore, base-resolution mapping showed that hmC is asymmetrically distributed at CpG

sites.[118b] Mapping of fC in ES cells revealed enrichment mostly at the same set of promoters as hmC,[120] while depletion of TDG in ES cells showed increased presence of fC and caC at these promoters and at enhancers regions, suggesting the possible occurrence of Tet-mediated demethylation at these sites.[118d, 121]

1.2.4.3 Putative roles of oxidized cytosine derivatives

Interest in the Tet proteins and in the oxidized cytosine derivatives has mostly focused on the possibility that these nucleosides could serve as intermediates in a pathway of DNA demethylation. In fact, methylation patterns have been found to display a dynamic nature, with both genome-wide and gene-specific changes being observed. To date, several mechanisms have been proposed, involving both passive and active demethylation.[102a, 122]

Passive demethylation, which is up to now the most widely accepted mechanism in mammals, involves erasure of the methylation pattern by preventing maintenance methylation by Dnmt1 and its partner Uhrf1. Dnmt1 was in fact found to have reduced methylation activity at hemi-hmC CpG sites, therefore suggesting a role for hmC in supporting passive demethylation (although coupled to oxidation) during transcription.[123] Passive demethylation is thought to take place during the two waves of genome-wide demethylation observed in the paternal pronucleus after fertilization and during specification of primordial germ cells. Rapid demethylation of the male pronucleus has in fact been found to correspond to genome-wide oxidation of mC to hmC, fC and caC by Tet3, and current evidence suggests that the modified cytosine patterns are subsequently lost through replication.[114] A similar phenomenon is thought to take place in primordial germ cells, where the drop in methylation levels observed between E9.5 and E10.5 was in fact found to coincide with massive oxidation of mC to hmC, in this case performed by Tet1 and Tet2, followed by replication-dependent dilution of hmC.[124]

Various mechanisms of active demethylation have been suggested and can mostly be classified as involving the DNA repair machinery, namely base excision repair (BER), or direct C-C bond cleavage at the C5 position of the modified cytosine (Figure 9).

Demethylation involving BER is well established in plants, where the Demeter family of glycosylases is responsible for direct excision of mC in *Arabidopsis thaliana*.[125] In mammals, BER is thought to either involve direct removal of the oxidizes cytosine derivatives, or to be coupled to deamination. The former mechanism was shown to take place in the case of fC and caC, which can both be excised by Thymine DNA glycosylase (TDG).[100a, 126] The latter pathway was proposed to involve deamination of hmC by activation-induced cytidine deaminase (AID)/apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC) family enzymes, followed by excision of the deamination product, hmU, by the glycosylases TDG and single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1).[127] In support of this pathway, TDG was shown to interact with AID and with the damage

response protein Gadd45,[127a] and previous findings have reported the coupling of deamination and BER in the removal of mC via conversion to T followed by base excision.[128] However, this pathway remains controversial, since studies investigating the activity of AID and APOBEC family proteins reported a reduced activity in the presence of increasingly C5-modified cytosine derivatives.[129]

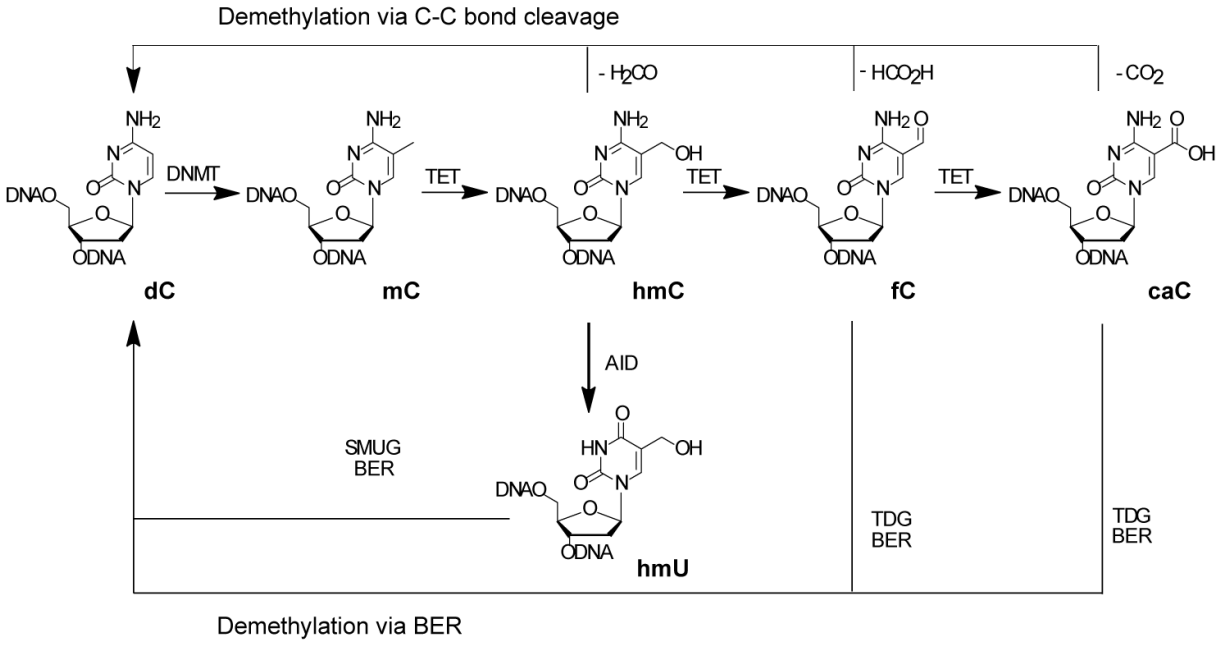


Figure 9 Summary of active demethylation pathways involving hmC, fC and caC.

On the other hand, active demethylation is proposed to take place via direct C-C bond cleavage at the C5 position of the modified cytosine.[122b] Hints for this mechanism come from the *E. coli* 2-oxoglutarate (2OG)-dependent dioxygenase AlkB, known to carry out the oxidative demethylation of 1-methyladenine and 3-methylcytosine by initial oxidation followed by release of formaldehyde.[130] Similarly, enzymes involved in the thymidine salvage pathway were shown to perform three consecutive oxidation steps to generate iso-orotate from T, followed by decarboxylation. However, the thymine 7-hydroxylase and the iso-orotate decarboxylase involved in this pathway were only isolated from fungi, and no homologue has yet been found in mammals.[131] Up to now, evidence for such a pathway was provided by a report published by Schiesser *et al*., who documented the presence of a decarboxylation activity in the ES nuclear extracts,[132] while Dnmt3a and Dnmt3b were shown to catalyze the removal of the hydroxymethyl group from hmC *in vitro*.[133] Nevertheless, the identity of the enzyme or complex involved in active demethylation via C-C bond cleavage *in vivo* remains elusive.

Apart from a possible role as intermediates in various pathways of DNA demethylation, oxidized cytosine derivatives might also act as specific DNA markers which can influence transcription and chromatin structure. Mass spectrometry analysis of modified cytosine interactors reveals that mC and its oxidized derivatives bind distinct subsets of proteins, suggesting specific biological roles particularly in the case of mC and hmC, while fC and caC largely recruit proteins involved in DNA

repair.[134] Most importantly, oxidized cytosines might be involved in antagonizing the mC-dependent silencing effect. For example, methyl CpG-binding protein 1 (MBD1), MBD2 and Kaiso, which are responsible for recruitment of repressive histone modifying enzymes, specifically interact with mC but not with hmC, suggesting that Tet-mediated oxidation to hmC might result in chromatin remodelling and in transcriptional activation.[123, 135] On the other hand, oxidized cytosine can affect the processivity of Pol II, and especially fC and caC were reported to stall the polymerase *in vitro*,[136] while hmC was reported to lower the melting temperature of the double helix compared to mC, therefore possibly promoting transcriptional elongation.[137]

In summary, since the discovery of the Tet proteins and of the oxidised methylcytosine derivatives hmC, fC and caC there has been great interest in these novel modified DNA nucleosides. While the large number of studies already publish has helped to shed light on aspects related to these nucleosides, many fundamental questions regarding their role and removal remain unanswered.

2 Aims of the Project

The distribution and dynamic regulation of tRNA modifications in eukaryotic systems is essential for the modulation of protein translation. Most of the studies on this topic reported to date have investigated the role and biosynthesis of individual tRNA modifications, while there is little information available concerning their systemic behaviour. We were therefore interested in investigating the global tRNA modification levels both in higher eukaryotes as well as in the unicellular organism *S. cerevisiae*, using an LC-MS-based quantification method. On the one hand we wanted to quantify the tRNA modification levels in various mammalian tissues in order to evaluate their tissue-dependent distribution (Chapter 3). On the other hand, we were interested in elucidating the nature of the variation in modification content in yeast cells during growth into stationary phase, with the aim of distinguishing between active increase of modification content versus passive enrichment of modified tRNA molecules. (Chapter 4).

Additionally, we wanted to elucidate the biochemical and structural features which enable the selective modification of U34 in some *E.coli* tRNAs. Biochemical studies were therefore planned to characterize the bifunctional enzyme MnmC, responsible for the two final modification steps leading to mnm^5s^2U34, and to analyze its interaction with substrate tRNA (Chapter 5).

In a second part of this work we focused on the analysis of modified DNA nucleosides and more specifically on the investigation of pathways involved in the removal of methylcytosine (mC). Since 2009, the number of modified cytosine derivatives has expanded to include hydroxymethylcytosine (hmC), formylcytosine (fC) and carboxycytosine (caC). These cytosine derivatives are suggested to be involved in pathways of active demethylation, but despite the growing number of studies on this subject there is still controversy concerning the mechanism of mC removal. We therefore wanted to investigate putative demethylation pathways by C-C bond cleavage between the hydroxymethyl-/formyl-/carboxy- group and the C5 atom on the cytosine base, using selectively labelled heavy atom-isotopologues of hmC, fC and caC (Chapter 6). For this purpose, DNA strands containing modified cytosine derivatives had to be generated by Polymerase Chain Reaction (Chapter 7), and subsequently implemented as substrates for *in vitro* activity assays. Additionally, we were interested in determining whether maintenance methylation is possible in the presence of hemi-modified fC- and caC-containing strands, and we therefore planned to investigate the methylation activity of mouse Dnmt1 in the presence of these cytosine derivatives (Chapter 6).

In a related project, we wanted to investigate an alternative pathway of active demethylation involving the putative deamination of hmC to yield hydroxymethyluracil (hmU). Deamination of hmC is in fact highly contested and our aim was to determine whether this process is indeed involved in the removal of mC. Interestingly, in the course of this study we observed that hmU generation is largely dependent

on oxidation of T by Tet1, rather than on deamination of hmC. We subsequently focused on confirming this finding (Chapter 8).

3 Isotope-based analysis of modified tRNA nucleosides correlates modification density with translational efficiency

Modified RNA nucleosides

DOI: 10.1002/anie.201203769

Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with Translational Efficiency**

*Caterina Brandmayr, Mirko Wagner, Tobias Brückl, Daniel Globisch, David Pearson, Andrea Christa Kneuttinger, Veronika Reiter, Antje Hienzsch, Susanne Koch, Ines Thoma, Peter Thumbs, Stylianos Michalakis, Markus Müller, Martin Biel, and Thomas Carell**

Transfer RNAs (tRNAs) are adapter molecules needed to translate genetic information into a peptide sequence.[1] At the ribosome, the anticodon of each tRNA reads the corresponding codon of the messenger RNA. This anticodon–codon interaction allows the ribosome's large subunit to catalyze amide-bond formation between the cognate amino acids present at the 3′ terminus of aminoacyl-tRNAs and the growing peptide chain.[2] The tRNA adapters required for this process display a surprisingly large chemical diversity.[3] Aside from the four canonical nucleosides A, C, G, and U, more than 100 modified nucleosides are key constituents (Figure 1).[4] The most diverse and complex chemical structures are found in the anticodon stem-loop either in the anticodon at the wobble position or directly adjacent to the 3′ position of the anticodon,[5] suggesting that here the chemical complexity is necessary to establish translational fidelity.[6] The ribosome seems to need the modified anticodon region to better distinguish correctly base-paired tRNA from mispaired interactions in order to prohibit, for example, codon-slippage processes that would lead to frameshifts.[7]

In order to investigate how the set of nucleoside modifications influences the translational efficiency we quantified the tRNA modifications individually in various tissues by an isotope-dilution-based LC–MS method. (Details on the materials and methods are given in the Supporting Information, Table S1, and Figure S1). The quantified modification levels were correlated with the translational efficiency by means of an in vitro translation system. For the experiments, 11 representative tRNA modifications (Figure 1) were chemically synthesized as isotope-labeled derivatives.[8] A majority of the investigated nucleosides are located inside the extended anticodon,[9] the other synthesized modified nucleosides are found at various other positions.[10]

As biological material for the analysis we chose a range of different organ tissues from mouse and pig. Porcine tissues were used because they are available in large amounts, while murine tissues were analyzed at a later stage to confirm the results in a genetically more defined organism. For pig, 5–10 g of tissue from two animals was sampled from each organ, while murine samples were obtained from two sets of five animals of which whole organs were analyzed. After total tRNA extraction and complete enzymatic hydrolysis to nucleosides, a mixture of the isotope-labeled tRNA modifications was added and the solution was subjected to LC–MS analysis. The ratios of the mass peak integrals from natural to isotope-labeled nucleosides were determined and calibration curves, which were previously measured for each investigated modified nucleoside, then allowed exact parallel quantification of the respective modifications (see Figure S2 in the Supporting Information).[8a] LC–MS quantification was performed at least in triplicate and results were averaged for each tissue. The error margin of the experiment was in this way limited to around 5%.

The obtained quantitative values for mouse and pig samples are shown color coded in Figure 2 A and B, respectively, together with the approximate positions of the measured modifications in the tRNA sequence. The values represent the measured number of each modification per 1000 tRNA molecules (‰) (exact values are listed in Tables S2–7 in the Supporting Information). Therefore, rather than yielding the absolute concentration of a modification in a given tissue, the data show directly the extent to which the analyzed tRNA set is modified. For representative murine and porcine tissues, an additional quantification of the ubiquitous tRNA modifications m^5C and Ψ was performed (see Figure S4).

The data show that each tissue type incorporates different amounts of a specific modified nucleoside into the respective tRNA ensemble. While the tRNAs in liver tissue contain large numbers of modified nucleosides (colored red), those isolated from lung and kidney tissue feature far fewer

[*] M. Chem. C. Brandmayr,[+] Dipl.-Chem. M. Wagner,[+] Dr. T. Brückl,[+] Dr. D. Globisch, Dr. D. Pearson, M. Sc. A. C. Kneuttinger, Dipl.-Chem. V. Reiter, Dr. A. Hienzsch, M. Sc. I. Thoma, Dipl.-Chem. P. Thumbs, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@cup.lmu.de
Homepage: http://www.carellgroup.de
Dipl.-Biol. S. Koch, Dr. S. Michalakis, Prof. Dr. M. Biel
Center for Integrated Protein Science at the Department of Pharmacy, Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)

[+] These authors contributed equally to this work.

[**] We thank the excellence cluster CIPSM and the SFBs 646 and 749 for generous support. Further support from the Fonds der Chemischen Industrie and Bayer Pharma AG is gratefully acknowledged. C.B. is grateful to the Boehringer Ingelheim Fonds for a predoctoral fellowship.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201203769.

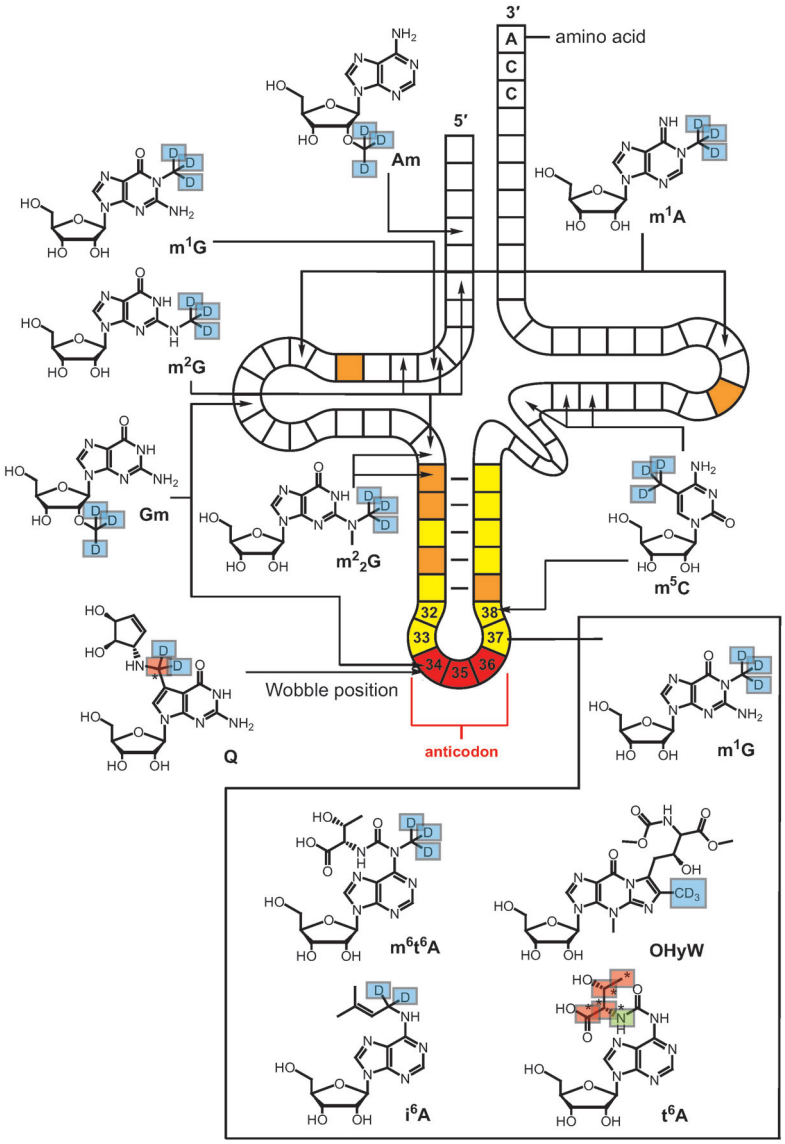


Figure 1. Isotope-labeled tRNA nucleosides present in eukaryotic tRNA and positions where these modifications are typically found. The introduced isotope labels are marked in color: D in a blue box: deuterium; * in a red box: ^{13}C; * in a green box: ^{15}N. The anticodon is highlighted in red, the remainder of the anticodon stem-loop is in yellow, and positions of Ψ are marked in orange. Abbreviations are explained in Table S1 in the Supporting Information.

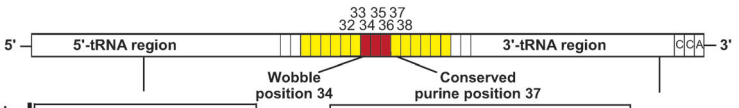
modifications (colored yellow and green). It is known that the amounts of individual tRNA species vary between tissues,[11] but these variations are small in comparison to the detected changes in modification levels, arguing that tissues modify their tRNA to different extents (for a detailed discussion about the influence of codon bias see the Supporting Information). Most important in this respect is the observation that the levels of the m^1A, m^2G, m^5C, and Ψ modifications (Figure 2 and Figure S4) follow the overall trend. These modifications are present in almost all tRNA species, and hence if tRNA composition would bias the quantitative data then the levels of these modifications would be expected to stay constant (or at least not follow the trend strongly).

Consistently, murine and porcine tissues show similar trends with liver, characterized by a high metabolic activity, having tRNA in both cases highly modified, while the tRNA from muscle tissue such as heart shows a rather low modification content.[12] Divergence between the two organisms is observed for some tissues such as cerebrum and spleen. This might be due to species-specific variation in tissue metabolism, or it might arise from genetic variation between species, as previously observed for bacteria.[13] Surprising is the observation that the data for Am follow a different trend, with higher Am levels found in tissues with largely unmodified tRNA (see Figure 2 and Table S8 in the Supporting Information). Because 2′-*O*-methylation stabilizes RNA,[4,14] the observed pattern could reflect the role of this modification in stabilizing hypomodified tRNA. Furthermore, while levels of queuosine (Q) generally fit the overall trend, this nucleoside has unexpectedly high levels in brain tissue both in mouse and pig, suggesting a more complex specialized role in those tissues.[15]

In order to confirm the results, we next measured the modification content in a sequence context. To this end we carried out a parallel LC–MS analysis of partial tRNA digests (RNase A) from two representative porcine tissues, liver and heart. From the digests we obtained a number of defined tRNA fragments (small oligomers) resulting from selective cleavage after C and U. Out of the obtained fragments we determined 10 for which we were able to detect the unmodified and modified sequences using mammalian tRNA sequences from the Sprinzl tRNA database.[3b] We then determined the relative amounts of the modified versus the unmodified tRNA fragments. The extent of modification of the respective tRNA sequence was calculated directly from the ratio between the areas of the specific mass peaks for the modified and the unmodified fragments (see Figure 3 and Table S9 in the Supporting Information).[16] The results show that the representative modified nucleosides m^1G, m^1A, m^2G, i^6A, and t^6A are indeed more abundant in tRNA fragments derived from liver, supporting the data from the isotope-dilution-based direct nucleoside quantification.

Based on the data we concluded that tissues mature their tRNA differently to satisfy individual translational needs. In support of this hypothesis we observed that published data for the rates of protein synthesis in vivo in different mammalian organs show a good level of correlation with our quantitative data for pig tissues (Figure S5 in the Supporting Information),[12,17] suggesting that higher overall tRNA modification content might be linked to faster rates of protein translation in a certain tissue. In order to test this hypothesis directly, we analyzed the translational efficiency of tissue-specific tRNA

A



Mouse: Tissues \ Nucleosides	m^1G	m^2G	$m^2{}_2G$	Q	t^6A	m^6t^6A	i^6A	OHyW	m^1A	Am
Liver	735.0	1102.7	614.6	62.5	304.0	30.2	49.8	37.7	1016.7	18.9
Cerebrum	694.1	1104.0	555.8	109.7	290.3	24.2	50.6	27.6	972.4	23.6
Cerebellum	646.6	1026.3	524.4	95.0	274.8	24.4	41.1	18.6	1011.0	20.7
Brain stem	644.3	1025.7	514.5	93.5	262.5	23.2	40.4	16.9	944.3	19.3
Lung	640.0	946.2	495.2	81.3	271.7	24.8	45.0	21.7	879.5	28.7
Heart	526.8	801.6	419.8	83.4	241.2	17.5	36.4	12.8	800.7	26.4
Kidney	517.9	779.7	389.5	70.8	202.7	19.8	36.0	22.5	715.7	21.0
Spleen	487.0	754.3	397.0	47.7	210.3	20.9	30.6	26.9	653.6	47.8

B

Pig: Tissues \ Nucleosides	m^1G+Gm	m^2G	$m^2{}_2G$	Q	t^6A	m^6t^6A	i^6A	OHyW	m^1A	Am
Liver	1059.6	1176.5	562.7	55.2	214.6	21.9	39.6	33.6	941.6	20.5
Cerebellum	829.4	1036.4	566.8	83.5	237.3	17.5	41.7	29.0	1102.8	67.8
Spine marrow	799.5	970.7	457.0	50.8	198.6	11.8	36.0	26.0	911.0	61.8
Spleen	690.5	793.6	464.3	42.7	182.1	13.9	30.7	26.1	833.6	80.4
Kidney	816.3	896.0	451.5	32.3	177.8	15.8	30.7	21.1	871.6	68.4
Lung	793.6	879.3	427.2	33.9	184.1	13.9	33.6	21.7	876.3	58.5
Heart	638.2	673.1	372.9	29.6	175.4	11.0	29.0	12.0	810.9	45.9
Cerebrum	586.8	682.2	406.2	57.2	165.8	10.5	26.1	20.3	732.6	139.9

Figure 2. Quantitative data for the investigated tRNA modifications in various murine (A) and porcine (B) tissues. All tRNA nucleoside values are given per 1000 tRNA molecules (‰). These data reveal a similar, tissue-dependent extent of modification for all investigated modified nucleosides except Am. Color codes in (A) and (B) are based on quantile calculations; red: highest value, yellow: 50% quantile, green: lowest value. For intermediate values appropriate intermediate shades were calculated. Despite the slight variation in the absolute quantification values, trends in modification content were conserved across different biological samples (both in mouse and pig, see Tables S2–S5 in the Supporting Information).

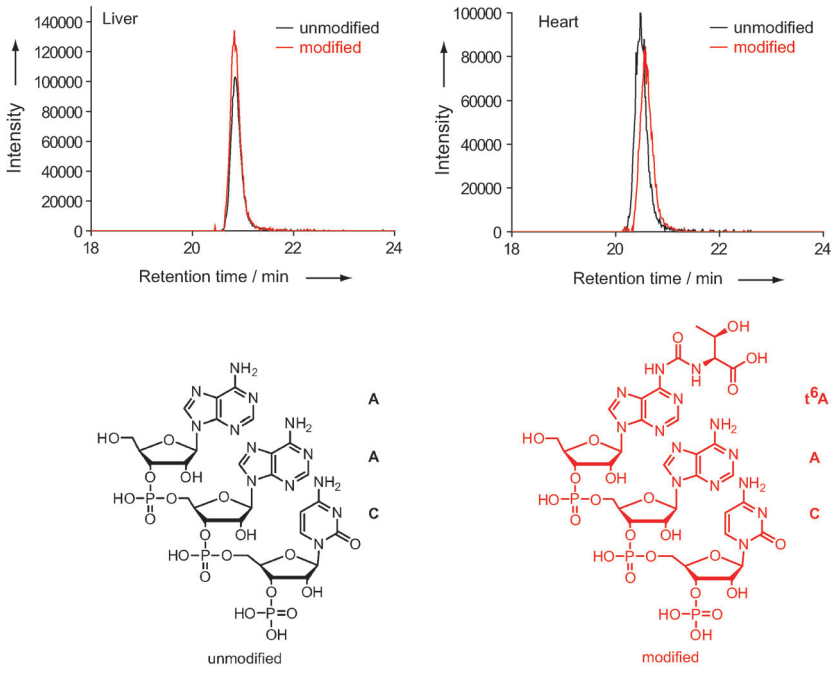


Figure 3. Representative qualitative comparison of amounts of unmodified RNA fragments AAC and the corresponding modified t^6AAC in the RNase A digests of liver and heart tRNA. Overlayed LC–MS chromatograms showing ions detected at the calculated masses of the AAC (m/z = 489.5682–489.5742) and t^6AAC (m/z = 562.0863–562.0933) fragments (z = −2) and the corresponding structures. The ratio of the peak areas of modified to unmodified fragments for liver can be seen to be higher than that for heart. Further identified fragment ratios are listed in Table S9 in the Supporting Information.

ensembles using an in vitro coupled transcription/translation reticulocyte lysate system.[18] The original tRNAs present in the system were removed chromatographically using an ethanolamine–Sepharose column.[18] Subsequently, the tissue-extracted tRNA ensembles were added. Translational efficiency was measured by observing the increase in luminescence linked to the production of the protein luciferase (see Figure S7). The slopes of the plotted curves from at least three repeated experiments were normalized to the most efficient ensemble.

In a first set of experiments, total tRNA ensembles from porcine tissues were used. Figure 4 plots the measured rates against the corresponding normalized modification levels calculated based on LC–MS data presented in Figure 2B (exact values are listed in Figure S8 in the Supporting Information). From Figure 4 it can be seen that the overall modification content correlates with the translational efficiency of the isolated tRNA ensemble, but the correlation is far from optimal.

We noted that specifically the values obtained from tissues known to be rich in mitochondria (heart in particular) deviate from the expected trend.[19] Since mitochondrial tRNA features its own set of modifications,[13] we therefore removed in a second experiment the mitochondria from the porcine tissues before the tRNA extraction (Figure S6 in the Supporting Information).[20] The obtained data for cytosolic tRNA show indeed a higher degree of correlation between the translational activity and the modification content (black squares in Figure 4; $r = 0.861$, $P = 0.028$), indicating that the modification level of cytosolic tRNAs is one factor that influences the efficiency of translation. This correlation was further confirmed for mouse using total tRNA ensembles extracted from tissues known to have relatively low mitochondrial tRNA content (see Figure S9). Our results are

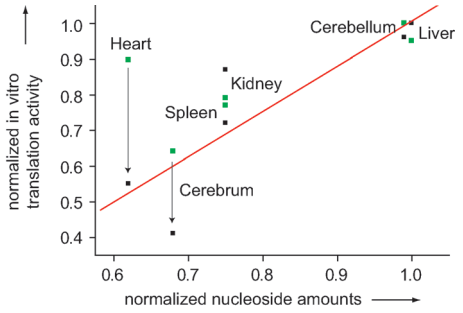


Figure 4. Translation activity of tRNA extracts isolated from different porcine tissues with (black squares) and without (green squares) removal of mitochondrial tRNAs. A plot of the linear fit of relative in vitro translation activity and normalized nucleoside levels shows a significant correlation after removal of mitochondria (red line; $r = 0.861$, $P = 0.028$).

consequently in good agreement with the common idea that specific noncanonical bases fine-tune the binding of tRNAs to the ribosome. As the translation rate is determined by the competition between near-cognate and cognate aminoacyl-tRNAs,[21] a high modification level increases the affinity of the correct tRNA to the ribosome, which may allow faster discrimination.[22] This reduces the ribosome step time, which in turn may increase protein synthesis rates.

In summary, we have reported the parallel quantification of 12 modified nucleosides in tRNA ensembles from various porcine and murine tissues and showed that the overall modification content varies substantially. Furthermore, we provide evidence that the modification level correlates with the in vitro protein synthesis capacity, suggesting that the extent to which the tRNA ensemble is chemically modified modulates the translational efficiency. Our data show that the tRNA modification level is another layer of information that programs cells in terms of their translational potency.

Received: May 16, 2012
Revised: July 26, 2012
Published online: October 4, 2012

Keywords: isotopic labeling · mass spectrometry · RNA modification · translation · tRNA

[1] M. Ibba, D. Söll, *Science* **1999**, *286*, 1893 – 1897.
[2] a) N. Fischer, A. L. Konevega, W. Wintermeyer, M. V. Rodnina, H. Stark, *Nature* **2010**, *466*, 329 – 333; b) T. M. Schmeing, V. Ramakrishnan, *Nature* **2009**, *461*, 1234 – 1242.
[3] a) H. Grosjean, *Fine-Tuning of RNA Functions by Modification and Editing*, Springer, Heidelberg, **2005**; b) F. Jühling, M. Mörl, K. Hartmann Roland, M. Sprinzl, F. Stadler Peter, J. Pütz, *Nucleic Acids Res.* **2009**, *37*, D159 – D162.
[4] Y. Motorin, M. Helm, *Biochemistry* **2010**, *49*, 4934 – 4944.
[5] a) P. F. Agris, F. A. Vendeix, W. D. Graham, *J. Mol. Biol.* **2007**, *366*, 1 – 13; b) A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* **2007**, *3*, 29 – 35.
[6] a) L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, *Nat. Struct. Mol. Biol.* **2010**, *17*, 555 – 560; b) A. L. Konevega, N. G. Soboleva, V. I. Makhno, A. V. Peshekhonov, V. I. Katunin, *Mol. Biol.* **2006**, *40*, 597 – 610; c) K. Nakanishi, L. Bonnefond, S. Kimura, T. Suzuki, R. Ishitani, O. Nureki, *Nature* **2009**, *461*, 1144 – 1148; d) L. A. Sylvers, K. C. Rogers, M. Shimizu, E. Ohtsuka, D. Söll, *Biochemistry* **1993**, *32*, 3836 – 3841.
[7] J. F. Atkins, G. R. Björk, *Microbiol. Mol. Biol. Rev.* **2009**, *73*, 178 – 210.
[8] a) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem.* **2009**, *121*, 8074 – 8077; *Angew. Chem. Int. Ed.* **2009**, *48*, 7932 – 7934; b) J. J. Dalluge, T. Hashizume, J. A. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 3242 – 3245.
[9] M. Yarus, *Science* **1982**, *218*, 646 – 652.
[10] a) E. M. Phizicky, J. D. Alfonzo, *FEBS Lett.* **2010**, *584*, 265 – 271; b) M. Helm, *Nucleic Acids Res.* **2006**, *34*, 721 – 733.
[11] a) J. B. Plotkin, H. Robins, A. J. Levine, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12588 – 12591; b) K. A. Dittmar, J. M. Goodenbour, T. Pan, *PLoS Genet.* **2006**, *2*, 2107 – 2115.
[12] J. C. Waterlow, *Exp. Physiol.* **1984**, *69*, 409 – 438.
[13] D. Globisch, D. Pearson, A. Hienzsch, T. Brückl, M. Wagner, I. Thoma, P. Thumbs, V. Reiter, A. C. Kneuttinger, M. Müller, S. A. Sieber, T. Carell, *Angew. Chem.* **2011**, *123*, 9913 – 9916; *Angew. Chem. Int. Ed.* **2011**, *50*, 9739 – 9742.
[14] a) Y. Motorin, M. Helm, *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 611 – 631; b) J. A. Kowalak, J. J. Dalluge, J. A. McCloskey, K. O. Stetter, *Biochemistry* **1994**, *33*, 7869 – 7876.
[15] a) T. J. Siard, J. R. Katze, W. R. Farkas, *Neurochem. Res.* **1989**, *14*, 1159 – 1164; b) L. Szabo, S. Nishimura, W. R. Farkas, *Biofactors* **1988**, *1*, 241 – 244.
[16] C. M. Castleberry, P. A. Limbach, *Nucleic Acids Res.* **2010**, *38*, e162.
[17] a) A. Suryawan, P. M. J. O'Connor, J. A. Bush, H. V. Nguyen, T. A. Davis, *Amino Acids* **2009**, *37*, 97 – 104; b) D. F. Goldspink, S. E. Lewis, F. J. Kelly, *Biochem. J.* **1984**, *217*, 527 – 534; c) M. Winick, A. Noble, *Dev. Biol.* **1965**, *12*, 451 – 466; d) J. W. Frank, J. Escobar, A. Suryawan, S. R. Kimball, H. V. Nguyen, L. S. Jefferson, T. A. Davis, *J. Nutr.* **2005**, *135*, 1374 – 1381.
[18] R. J. Jackson, S. Napthine, I. Brierley, *RNA* **2001**, *7*, 765 – 773.
[19] V. Reiter, D. M. S. Matschkal, M. Wagner, D. Globisch, A. C. Kneuttinger, M. Müller, T. Carell, *Nucleic Acids Res.* **2012**, *40*, 6235 – 6240.
[20] C. Frezza, S. Cipolat, L. Scorrano, *Nat. Protoc.* **2007**, *2*, 287 – 295.
[21] J. Ling, N. Reynolds, M. Ibba, *Annu. Rev. Microbiol.* **2009**, *63*, 61 – 78.
[22] a) A. L. Konevega, N. G. Soboleva, V. I. Makhno, Y. P. Semenikov, W. Wintermeyer, M. V. Rodnina, V. I. Katunin, *RNA* **2004**, *10*, 90 – 101; b) O. Allnér, L. Nilsson, *RNA* **2011**, *17*, 2177 – 2188; c) V. Dao, R. Guenther, A. Malkiewicz, B. Nawrot, E. Sochacka, A. Kraszewski, J. Jankowska, K. Everett, P. F. Agris, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2125 – 2129; d) J. Urbonavičius, J. M. B. Durand, G. R. Björk, *J. Bacteriol.* **2002**, *184*, 5348 – 5357.

4 Investigation of yeast tRNA modification content and turnover during growth into stationary phase

4.1 Introduction

tRNA modifications are known to play a crucial role during protein translation by affecting the tRNA binding affinity to the ribosome and by reducing frame-shifting and misincorporation.[11a, 11b, 18a] Furthermore, recent studies have revealed that modified tRNA nucleosides are also involved in cellular stress response mechanisms by altering the translational efficiency of specific protein subsets and by influencing tRNA stability.[4a, 44a, 46] Previous studies performed by Dr. Antje Hienzsch, which focused on variations in tRNA modification content in *S. cerevisiae* in response to a variety of stress conditions,[138] revealed that yeast cells grown into stationary phase display an increasingly modified tRNA population. A similar observation was made by Preston *et al.*,[139] who documented an increase in m^5C during entry of *S. cerevisiae* into stationary phase. However, there are yet no data available concerning how this increase in tRNA modifications is afforded, and whether it is a regulated response to nutrient depletion.

4.2 Results

The study presented in this chapter aims at investigating the origin of the observed modification increase during entry of yeast cells into stationary phase. According to our hypothesis, three main scenarios could explain the change in modification content: on the one hand, fast turnover of tRNA species would imply that the increase in modification is due to new, more modified tRNAs which are generated during entry into stationary phase, implying a regulation of tRNA modifications in response to the stress conditions; on the other hand, slow tRNA turnover would link modification increase to a longer RNA half-life, suggesting that tRNA species present in the cell can be increasingly modified over time. Finally, a third possibility is that, regardless of the turnover rate, hypomodified tRNAs might be selectively degraded leading to an enrichment of the modified tRNA pool.

In order to distinguishing between these possible scenarios we decided to monitor tRNA modification turnover during entry into stationary phase by pulse-chase labelling. The approach relied on initial labelling of the tRNA modifications during exponential growth and subsequent monitoring of these modifications during growth into stationary phase to evaluate their turnover and to assess the timing and regulation of RNA modification.

In order to label the tRNA modifications, yeast *S. cerevisiae* was grown in medium containing [*methyl*-D_3]-L-methionine. The labelled methionine is then converted into [*methyl*-D_3]-*S*-adenosyl methionine ([*methyl*-D_3]-SAM), which serves as a common methyl donor, leading to labelling of methylated RNA nucleosides. In order to monitor the fate of RNA modifications throughout different growth phases, yeast was initially grown in labelled complete minimal medium to an optical density of $OD_{600} = 1$, at which point cells were gently pelleted and resuspended in rich YPD medium (non-labelled) and allowed to grow for 7 days in the same medium (see Figure 10). Samples were removed prior to medium exchange (time point 0 h), and at time points 1, 2, 4, 20, 24, 48, 72, 96, 120, 144,

168 h after exchange to non-labelled medium. At each collection point the optical density was measured and tRNA was extracted using the procedure detailed in section 4.4.2. tRNA samples were subsequently digested to nucleosides, spiked with labelled samples to quantify non-methylated RNA modifications and subjected to LC-MS analysis.

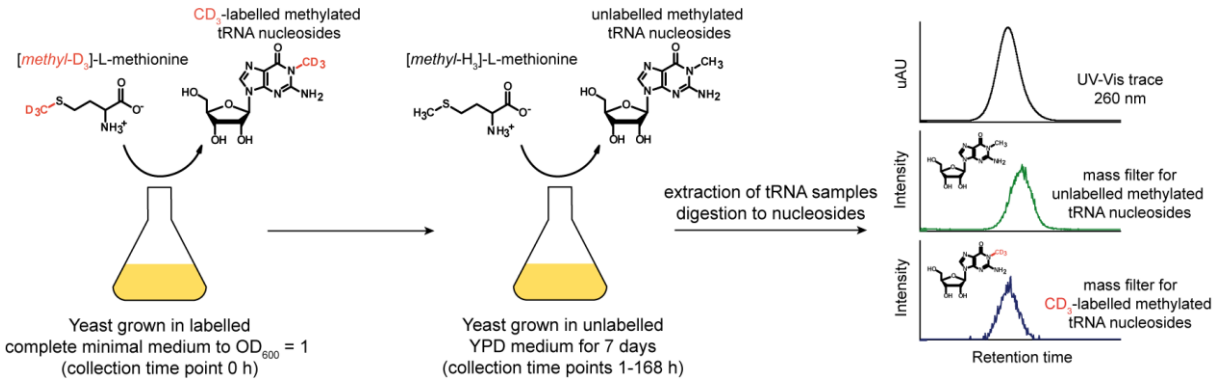


Figure 10 Overview of the experimental procedure.

The analysis of samples collected at time points 0–168 h proceeded in three steps: first, evaluation of the tRNA and tRNA modification content, in order to confirm the increase in modification levels during entry into stationary phase;[138] second, evaluation of the ratio of light to heavy modification content over time and comparison with expected ratios based on cell growth; third, evaluation of heavy labelled modification content and its variation over time.

4.2.1 Variation in tRNA content per cell during growth

Analysis of the tRNA content throughout the 7 days revealed that the overall tRNA population changes over time. As shown in Figure 11 (Panel A), upon exchange to unlabelled YPD medium, optical density followed the expected growth profile: initially, we observe a rapid increase in cell density, corresponding to log-phase (first 24 h), where cell growth is based on fermentation of glucose; this stage is followed by a more moderate growth phase known as post-diauxic phase (day 1-5), where metabolism shifts from fermentation to respiration, mostly based on ethanol and other non-fermentable carbon sources. Once all carbon sources have been depleted from the medium, cells reach saturation and cease proliferation at about day 5.

Interestingly, tRNA content per OD unit increases during the first four time points, reaching a maximum at time point 4 h, and it then slowly decreases starting from day 1 until day 7 (see Figure 11, Panel B). Note that in the absence of further time points between time 4 h and 20 h we cannot exclude that during this time the tRNA content per OD unit might further increase. The observed decrease in tRNA content starting from day 1 is in line with previous findings published by Sethy *et al.*,[140] which showed that tRNA transcription is down-regulated during transition from exponential to postdiauxic phase. In their study, the growth-dependent change in transcription was attributed to depletion of $TFIIIB_{70}$ and possibly also affected by TFIIIC, though to a lesser extent. $TFIIIB_{70}$ is a subunit of transcription factor TFIIIB which is stoichiometrically limiting for Pol III transcription. Later findings

also documented Maf1 negative regulation of Pol III transcription levels during shift from fermentation to respiration.[141]

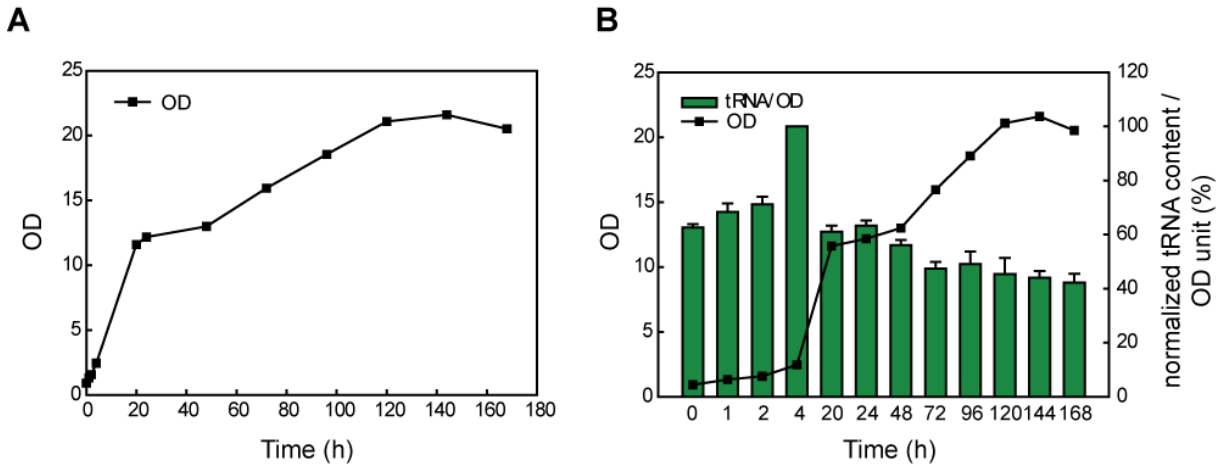


Figure 11 Yeast growth profile and tRNA content/OD unit after transfer to unlabelled YPD medium. A) Variation in optical density (OD_{600}) during growth in YPD over 7 days. B) Normalized tRNA content per OD unit (%). Time points 1-168 h refer to sampling time after medium exchange to unlabelled YPD (note that the Time axis in plot B is not linear). (exact values listed in Table 2).

4.2.2 Quantification of tRNA modifications during growth into stationary phase

In order to monitor the extent of tRNA modification and to confirm that it increases during growth into stationary phase after initial growth in labelled medium, quantification of the modified nucleosides was carried out using two different methods depending on the nature of the modification. In the case of methylated modifications (m^1A, m^1G, m^2G, m^2_2G, m^5C, Am), quantification was performed by UV absorbance, since the endogenous labelled nucleosides did not permit quantification using synthetic standards. In the case of non-methylated modifications (Ar, i^6A, t^6A), isotope standards were added to the nucleoside mixture for LC-MS quantification.[142] Therefore, accurate quantification could be obtained for the modifications m^1G, m^2G, m^1A, m^5C, Am, Ar, i^6A, t^6A and m^2_2G. In the case of Cm, m^7G and Gm, as well as the initial 4 time points of m^2_2G, the modified nucleosides could not be quantified by UV absorbance due to poor resolution by column chromatography, and only approximate quantification was obtained from comparison of the ion count signals as detailed below. yW could not be quantified due to low abundance and poor absorbance of the hypermodified nucleoside. Most importantly, in all cases quantification of the modified nucleosides was normalized with respect to the content of A nucleoside (quantified by UV absorbance). Therefore, changes in modification levels reflect changes in the proportion of modified nucleosides in the overall tRNA population.

As an example, Figure 12B shows the normalized variation in m^1G content over 7 days, a trend which could be observed for most other modifications as well. Exceptions were observed in the case of i^6A, t^6A and Am, and they will be discussed separately.

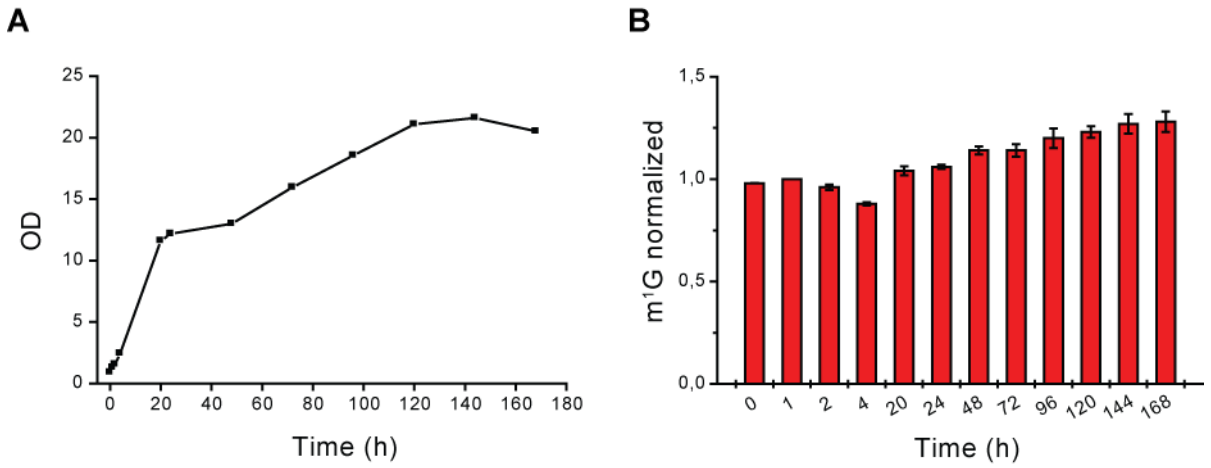


Figure 12 A) Variation in optical density (see Figure 11) and B) Relative change in m^1G content during 7 days. Quantification of m^1G was performed by means of UV absorbance. Values were normalizes to time point 1 h. Time points 1-168 h refer to sampling time after medium exchange to unlabelled YPD. Note that in B) the Time axis is not linear. For further details, see Table 3.

As shown in the plot, variation in modification content can be divided in two phases: an initial phase, during the first 24 h, and a subsequent phase starting from day 2. Throughout day 1, modifications vary only very slightly, but there is a distinctive trend showing an initial decrease in modification content which reaches a minimum at 4 h (for individual modifications, see Figure 13). This feature is observed for m^1G, m^2G, m^1A, m^5C and Ar. Interestingly, this minimum correlates with a maximum in tRNA content as highlighted in Figure 11B. This correlation could be explained in terms of a fast tRNA production required to sustain log phase exponential growth, which cannot be matched by a sufficiently fast tRNA modification rate, therefore yielding a partially undermodified tRNA population at time point 4 h.

In the case of Am, modification content follows exactly the opposite trend, reaching a maximum at time 4 h. Methylation at the 2´-O- position is known to stabilize RNA[4b, 50a] and, in the case of mammalian tissues, higher overall tRNA modification content correlates with a lower Am content (see Chapter 3), suggesting that this modification is introduced to stabilize less modified tRNA samples. Therefore, at time point 4 h (as shown is Figure 13), a higher content of Am modification might be present in order to compensate for the lower extent of modification. In the case of i^6A, no distinctive trend is observed, while t^6A shows a decreasing trend during the first 24 h (except for time point 0 h).

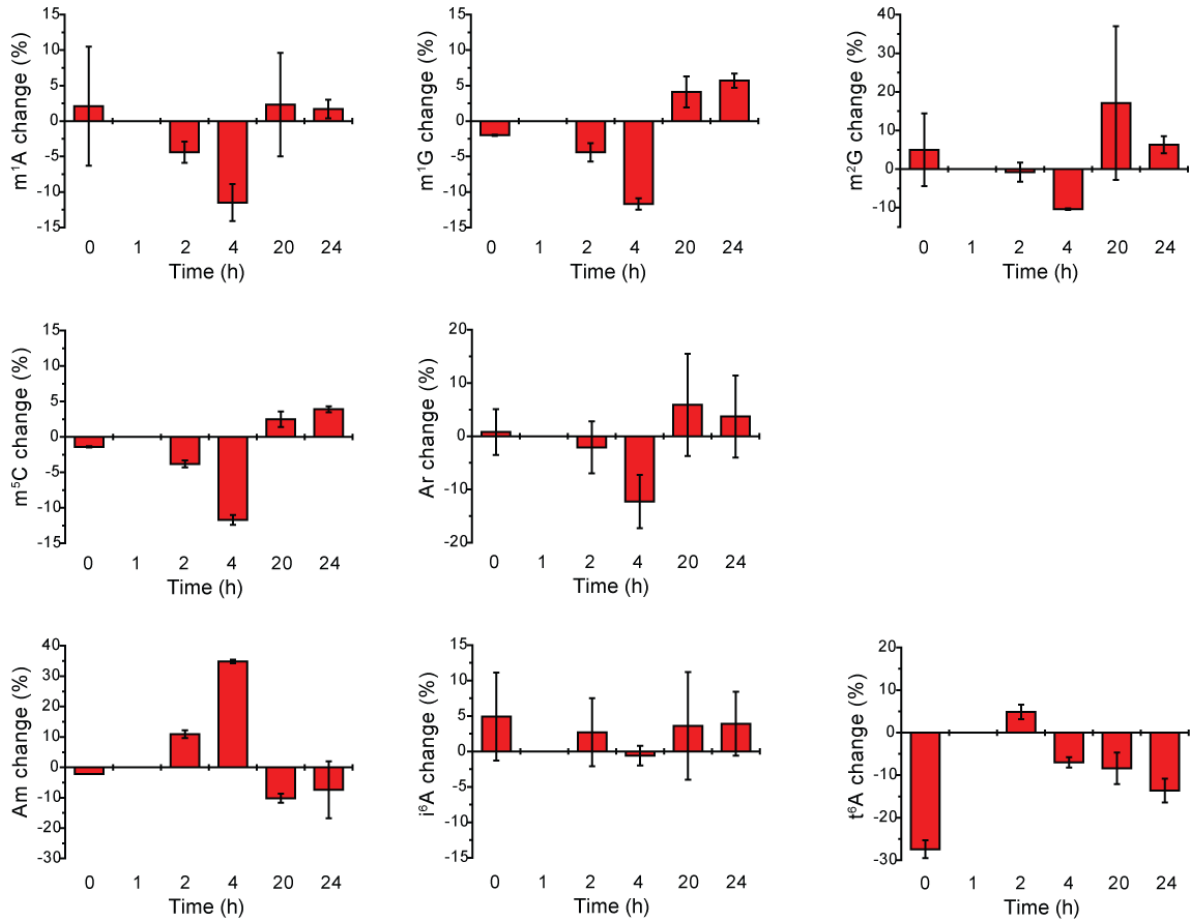
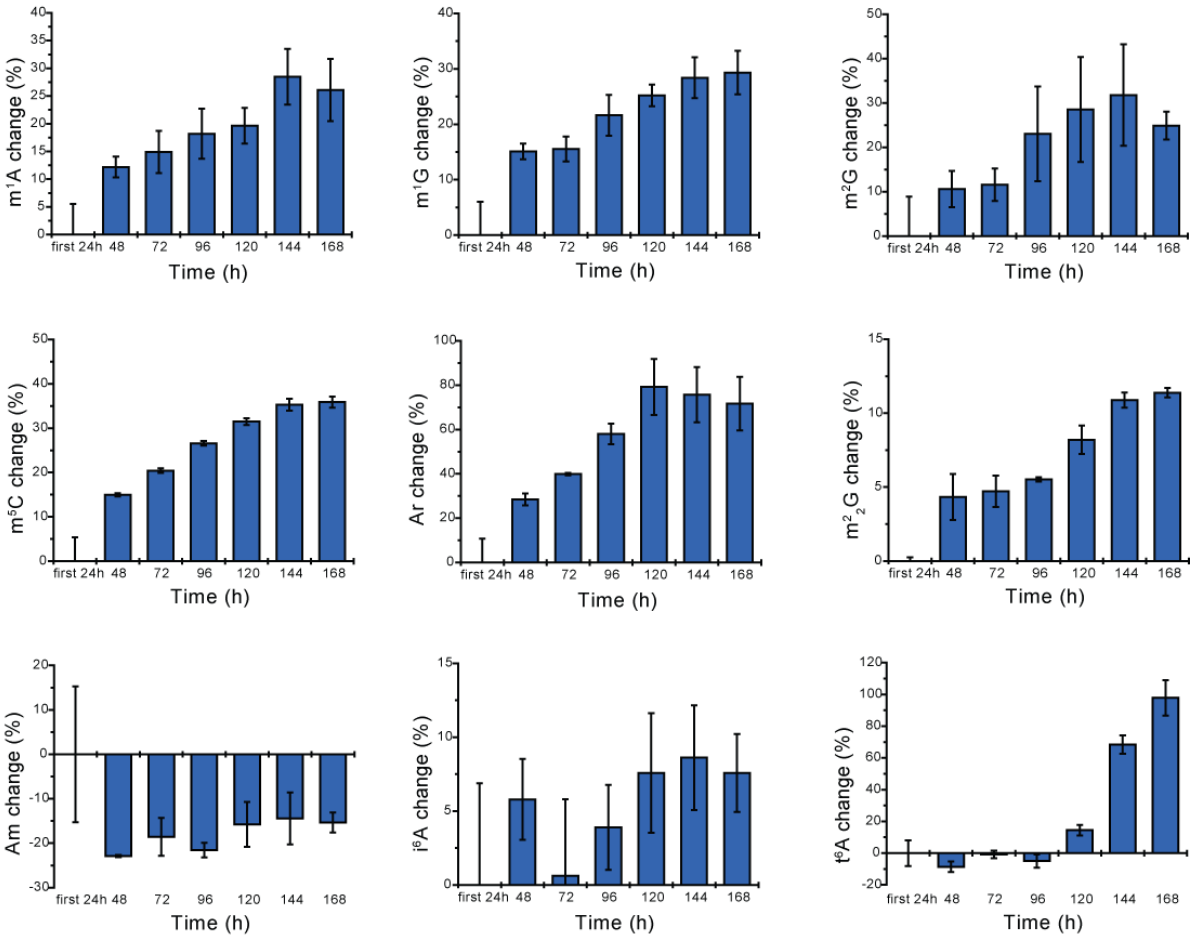


Figure 13 Percental change in modification content during day 1. Plots show the variation percent in modification content for the nucleosides m^1A, m^1G, m^2G, m^5C and Am quantified by UV absorbance, and for Ar, i^6A and t^6A quantified by LC-MS using labelled isotope standards (in all cases quantification was normalized to A nucleoside content). Values were normalized to time point 1 h (see Table 4).

After the first 24 h, in agreement with results previously described by Dr. Antje Hienzsch, most modifications increase during growth towards saturation.[138] As shown in Figure 14 (Panels A and B), which details the percental change for each modification with respect to day 1, m^1A, m^1G, m^2G, m^5C, Ar and $m^2{}_2G$ all increase from day 2 to day 7. In some cases the increase is very marked, such as for m^5C and Ar, while in other cases only mild increase is observed, such as in the case of $m^2{}_2G$. Once again, the 2´-O-methylated nucleoside Am behaves differently, showing a rather constant or slightly decreasing trend with respect to the first 24 h. As previously discussed, this might be explained in terms of stabilization of the tRNA species, which for the increasingly modified tRNA samples observed over time results therefore in reduced Am content. In the case of i^6A, modification content remains constant over time, which is to some extent in disagreement with previous findings described by Dr. Antje Hienzsch, who instead observed an increase of this modification. A possible reason for such discrepancy might arise from the different initial samples, since in the original study by Dr. Antje Hienzsch yeast samples at time 0 were collected at a cell density of $OD_{600} = 0.6$, while in this case the density at time 0 h was $OD_{600} = 1$. Alternatively, it might be due to the different starting conditions of yeast cells in minimal labelled medium with respect to cells grown only in rich YPD medium (as performed by Dr. Antje Hienzsch). Finally, t^6A modification content remains relatively constant

during the first 4 days, followed by a dramatic increase starting from day 5. This could either result from a larger availability of L-threonine (required for the biosynthesis) at later growth stages, or might instead be linked to a higher activity or expression of the biosynthetic enzyme Sua5. Dependence of the t^6A modification content on L-threonine availability might also explain the low t^6A content at time point 0 h (see Figure 13), since at this time point yeast is grown in minimal medium where we expect a lower availability of amino acids.

A



B

	tRNA scaffold modifications					structural and anticodon modifications		modifications at position 37	
	Am	m^2_2G	m^1A	m^2G	Ar	m^5C	m^1G	t^6A	i^6A
first 24h	0	0	0	0	0	0	0	0	0
48	-22.88	4.33	12.19	10.62	28.47	14.97	15.10	-8.68	5.79
72	-18.59	4.72	14.92	11.58	39.88	20.42	15.53	-0.82	0.62
96	-21.55	5.52	18.19	23.06	57.93	26.59	21.62	-5.02	3.90
120	-15.77	8.20	19.63	28.55	79.19	31.48	25.21	14.48	7.58
144	-14.46	10.89	28.47	31.80	75.63	35.29	28.40	68.42	8.62
168	-15.35	11.38	26.07	24.90	71.65	35.88	29.32	97.88	7.58

Figure 14 Percental change in modification content with respect to day 1. A) Plots show the variation percent in modification content during days 2 to 7 with respect to day 1. Average values for day 1 were calculated by averaging the first six time points (from 0 to 24 h). B) Summary color-coded table displaying the percental change for each modification depending on the position in the tRNA molecule. Color code: red = highest value; yellow = 0 % change; green = lowest value (see Table 5).

In the case of Gm, Cm and m^7G, an approximate quantification of the relative change was performed by comparison of the integral of the ion count signals of light and heavy modification with respect to either m^1G, in the case of Gm, or with respect to m^5C, for Cm and m^7G. m^1G and m^5C were selected for comparison because they elute at similar retention times. These ratios were normalized with respect to the UV absorbance of the reference nucleosides and of the nucleoside A (for details, see Table 1). The resulting change in modification content is displayed in Figure 15. Approximate quantification reveals that, especially in the case of Gm, there is a large increase in modification content stating from day 2, in agreement with other modifications (Figure 14).

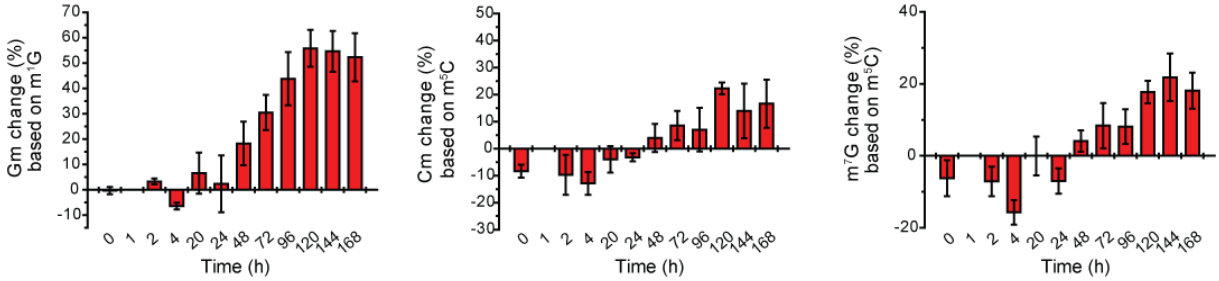


Figure 15 Percental change in Gm, Cm and m^7G. Plots show the percental change of Gm (based on m^1G), Cm and m^7G (based on m^5C) with respect to the modification content at time point 1 h. Changes in modification content were estimated from comparison of the ion count integrals and normalized to A nucleoside content (see Table 6).

In summary, yeast tRNAs show a slight decrease in the extent of modification during the first 24 h, followed by an overall increase in modification content during entry into stationary phase, in line with previous findings.[138] Exceptions are observed particularly for Am, which shows the opposite trend. This modification is generally thought to stabilize poorly modified tRNA molecules. Variations in Am modification content might therefore reflect a different need for stabilization which inversely correlates with the overall extent of modification of the tRNA species.

4.2.3 Quantification of light and heavy labelled modifications during growth

In order to monitor the fate of heavy (CD_3)-labelled modifications, ratios of light (CH_3)- to heavy (CD_3)-labelled modifications was evaluated over time. The first sample (time 0 h) was collected prior to exchanging the medium to rich YPD medium (non-labelled) and was used to evaluate the maximal extent of labelling. Subsequent samples collected after medium exchange displayed gradually decreasing labelling, as expected in response to culture growth in non-labelled medium. Note that even in labelled medium, the extent of heavy modification did not reach 100 %, due to endogenous methionine biosynthesis. Endogenous methionine is non-labelled because it is generated from homocysteine and 5-methyltetrahydropteroyltri-L-glutamate. The result is a mixed SAM pool which is only partially labelled and which results therefore in a partial labelling of methylated nucleosides.[143] Upon changing of the medium, the mixed SAM pool containing [*methyl*-D_3]-SAM is expected to convert to fully unlabelled SAM within few minutes (labelled SAM expected to drop to half within 20 s).[144] Therefore the extent of heavy (CD_3)-labelling detected in tRNA species during growth is

only due to modifications introduced during growth in labelled medium, and can be used to assess turnover of tRNAs generated prior to medium exchange.

CH_3/CD_3 ratios were quantified by comparison of the integrals of light and heavy modifications by LC-MS analysis. Analysis of CH_3/CD_3 ratios was performed for the modifications m^1A, Am, m^7G, Gm, m^1G, m^2G, $m^2{}_2G$, m^5C and Cm. In labelled medium, the proportion of light modification content was on average about twice as abundant as the heavy modification. Upon exchanging of the medium, the CH_3/CD_3 ratio increased over the first five days until it stabilized at a ratio of about 70-80 between day 5 and day 7. Most modifications followed a similar pattern, with the exception of Am which on average showed lower ratios.

Having quantified the CH_3/CD_3 ratios for the aforementioned modifications, the variation in ratios over time was evaluated using two different models. In the first model, expected ratios over time were estimated based purely on changes in optical density. In the second, more complex model, the variation in tRNA content per OD unit (see Figure 11) was also taken into account. In both cases, measured ratios of all modifications (except for Am) were averaged and their average ratio at time 1 h was used to construct the models. Comparison was then done between the average measured ratios over the 7 days and the expected ones. Am was analyzed separately. Both models rely on the assumption that the heavy-labelled tRNA content present at time point 1 h is not degraded over time and that it can therefore be used to calculate the expected ratios assuming that there is a linear increase in modification content with increasing OD.

In the first model, we assume that tRNA content per OD unit is constant over time, and that the CH_3/CD_3 modification ratio can be estimated exclusively from the increase in optical density, i.e. the increasing CH_3/CD_3 ratio reflects the increasing dilution in labelled tRNA modification during cell division (with tRNA content proportionally increasing with OD). Therefore, starting from an OD_{600} of 1.32 (time 1 h) with an average CH_3/CD_3 ratio of 2.24 ± 0.12, we can assume that about one third of the tRNA modifications are labelled. This number is assumed to remain constant over the 7 days, and therefore the expected CH_3/CD_3 ratio can simply be estimated by calculating the proportion of non-labelled tRNAs from the measured OD_{600} at each time point (for details, see Table 1).

Figure 16A shows the comparison of the average measured CH_3/CD_3 ratios (for all modifications but Am) with respect to the expected ratios calculated based on optical density. As it can be seen, expected ratios accurately predicted measured values up to time point 24 h. However, starting from day 2 the measured CH_3/CD_3 ratios are higher than expected, suggesting that there is a greater proportion of light (CH_3)-modification than expected. This can result from two different scenarios: it can either be due to an increase in the number of modifications introduced starting from day 2, meaning that tRNA samples are more extensively modified with non-labelled SAM, or it could arise from turnover of labelled tRNA, which would lead to a gradual loss of heavy modifications and therefore to higher

CH_3/CD_3 ratios. Interestingly, Am follows the opposite trend: as shown in Figure 16B, starting from time point 20 h the average measured CH_3/CD_3 ratio for Am is lower than the ratio expected based on optical density, implying that this modification is introduced to a lesser extent than expected during growth into stationary phase. This would again support the hypothesis that Am might be introduced preferentially for low modified tRNA samples (first 24 h), later becoming less essential upon increasing extent of tRNA modification (day 2–7).

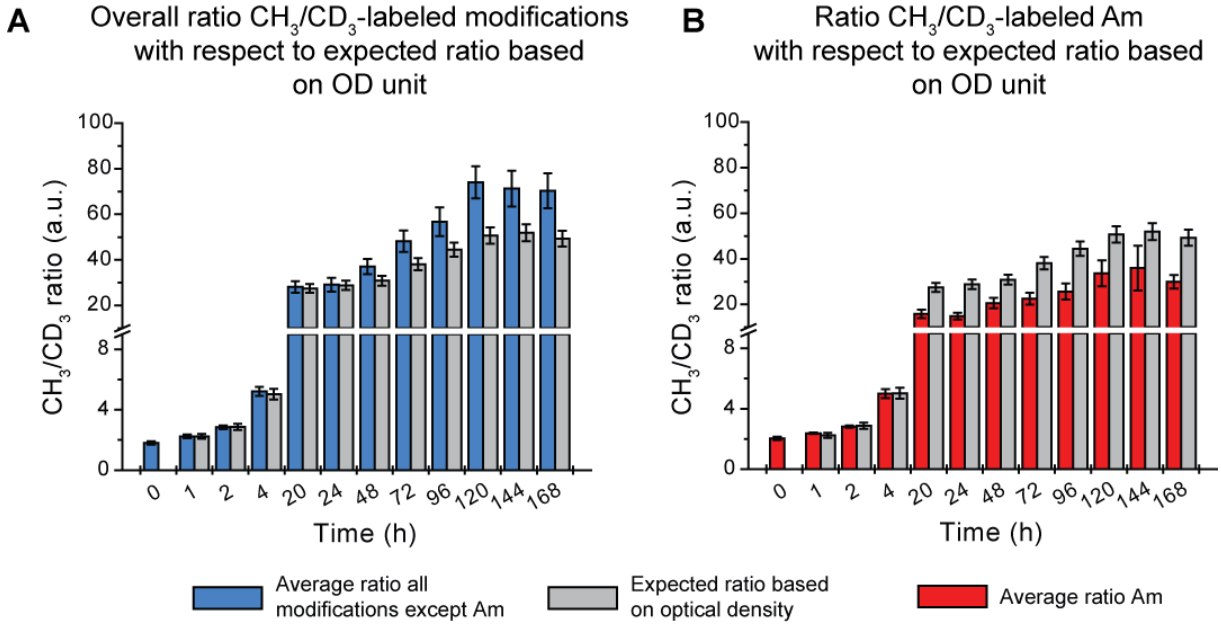


Figure 16 Comparison of measured and expected CH_3/CD_3 ratios based on optical density. A) Comparison of measured and expected ratios for all modifications except Am. Measured ratios were averaged for m^1A, m^1G, m^2G, m^7G, Gm, $m^2{}_2G$ (CD_3-, CH_3-labelled only), m^5C and Cm. Expected ratios were calculated assuming that light (CD_3)-modification content increased linearly with optical density, starting from time point 0 h, with an initial CH_3/CD_3 ratio of 2.24 and OD_{600} = 1.32. B) Comparison of average measured Am ratios and expected ratios. Expected ratios were calculated as in A).(see Table 7)

In the second model, expected CH_3/CD_3 ratios are calculated based on varying tRNA content per OD unit as well as optical density. This model accounts for the fact that the tRNA pool per cell is not constant over time. In fact, given that the tRNA per OD unit gradually decrease after the first 24 h (see Figure 11), a model assuming a constant tRNA content per OD unit (first model) would overestimate the amount of non-labelled tRNA produced during days 2 to 7.

Once again, the model is constructed from knowledge of the initial OD_{600} (1.32), CH_3/CD_3 ratio (2.24 ± 0.12) and normalized tRNA content per OD unit (68.4 % ± 3.1 %). The labelled tRNA pool is assumed to remain constant over the 7 days (i.e. no turnover), and the expected ratios are calculated as described above (see also Table 1). Figure 17A shows the comparison of the measured and expected CH_3/CD_3 ratios for all modifications except Am. As in the previous model, the expected ratios closely model the measured ones up to time point 24 h, after which the expected ratios largely underestimate the real CH_3/CD_3 values. As in the previous case, this might arise from increased modification extent starting from day 2, or it could be due to degradation of labelled tRNA. Interestingly, the expected ratio at time point 4 h overestimates the CH_3/CD_3 ratio. This, knowing that at this time point we

observe the largest proportion of tRNA content per OD unit (Figure 11B), could be explained in terms of an overestimation of the extent of tRNA modification at 4 h. In fact, as discussed in Section 4.2.2, Figure 13, the extent of modification at this time point reaches a minimum, implying that the tRNA sample is less modified. The model used to calculate the CH_3/CD_3 ratios shown in Figure 17 assumes instead that the proportion of tRNA modification is uniform at all time points, and is therefore not surprising that it might overestimate the ratio at time point 4 h.

Once again, Am shows a different pattern, with ratios which are generally fitting with the modelled ratios, suggesting that the Am content in tRNA is not increasing over time, but that it is rather more closely correlating with the increase in tRNA during growth.

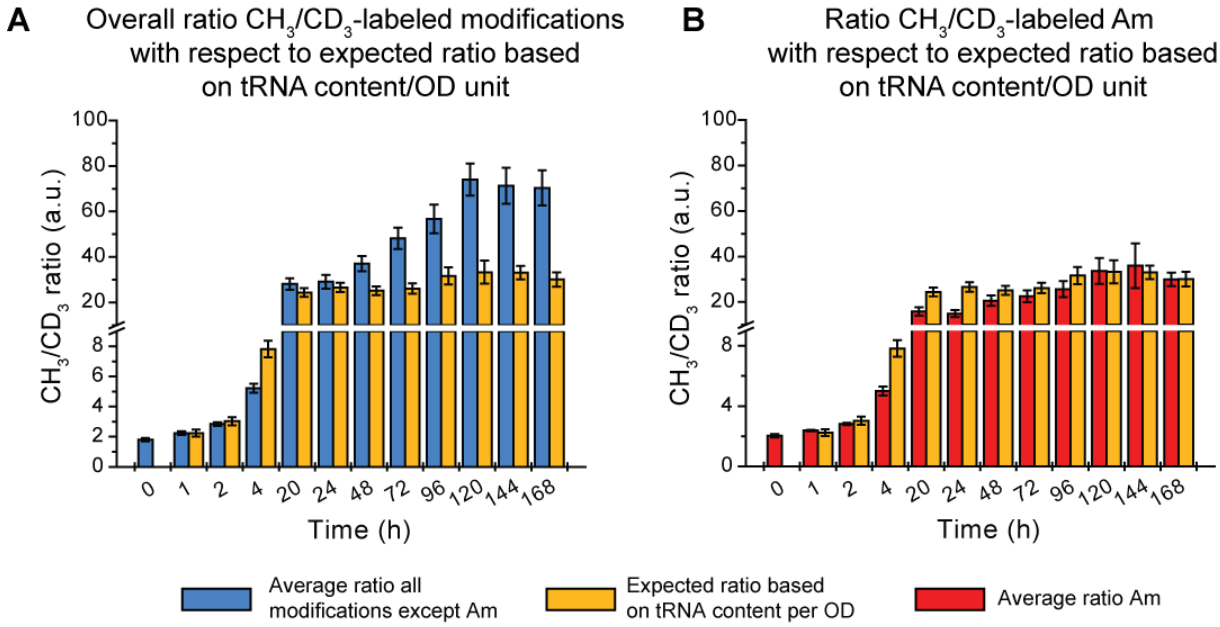


Figure 17 Comparison of measured and expected CH_3/CD_3 ratios based on tRNA content per OD unit. A) Comparison of measured and expected ratios for all modifications except Am. Measured ratios were averaged for m^1A, m^1G, m^2G, m^7G, Gm, $m^2{}_2G$ (CD_3-, CH_3-labelled only), m^5C and Cm. Expected ratios were calculated assuming that light (CH_3)-modification content increased linearly with tRNA content, starting from time point 0 h, with an initial CH_3/CD_3 ratio of 2.24, OD_{600} = 1.32 and tRNA content per OD unit of 68.38 %. B) Comparison of average measured Am ratios and expected ratios. Expected ratios were calculated as in A). (see Table 8)

In summary, evaluation of the CH_3/CD_3 ratios over time suggests that for all modifications but Am there is a greater proportion of light (CH_3)-modification than expected starting from day 2. This might either be due to the production of more highly modified tRNAs starting from this time point, or it might result from degradation of labelled tRNA, yielding a reduced heavy (CD_3)-modification content and higher CH_3/CD_3 ratios. Once again, Am shows a different trend, with ratios largely fitting with a linear increase in modification content following optical density and variation in tRNA content, again suggesting a different role of this modification compared to other modified nucleosides investigated in this study.

4.2.4 Evaluation of CH_3- and CD_3-modification content over time

We next calculated the content of heavy and light nucleoside for the modifications for which quantification by UV absorbance could be obtained in order to evaluate the variation in CD_3-labelled modification over time. Based on a simple model which only accounts for variations in cell number, the proportion of (CH_3)- and (CD_3)-modification for each point was estimated from the experimental CH_3/CD_3 ratio and the modification content quantified by UV absorbance, and subsequently multiplied by the optical density to account for cell growth over time (see Table 1). Data for m^1A, m^1G, m^2G, m^5C and Am are shown in Figure 18.

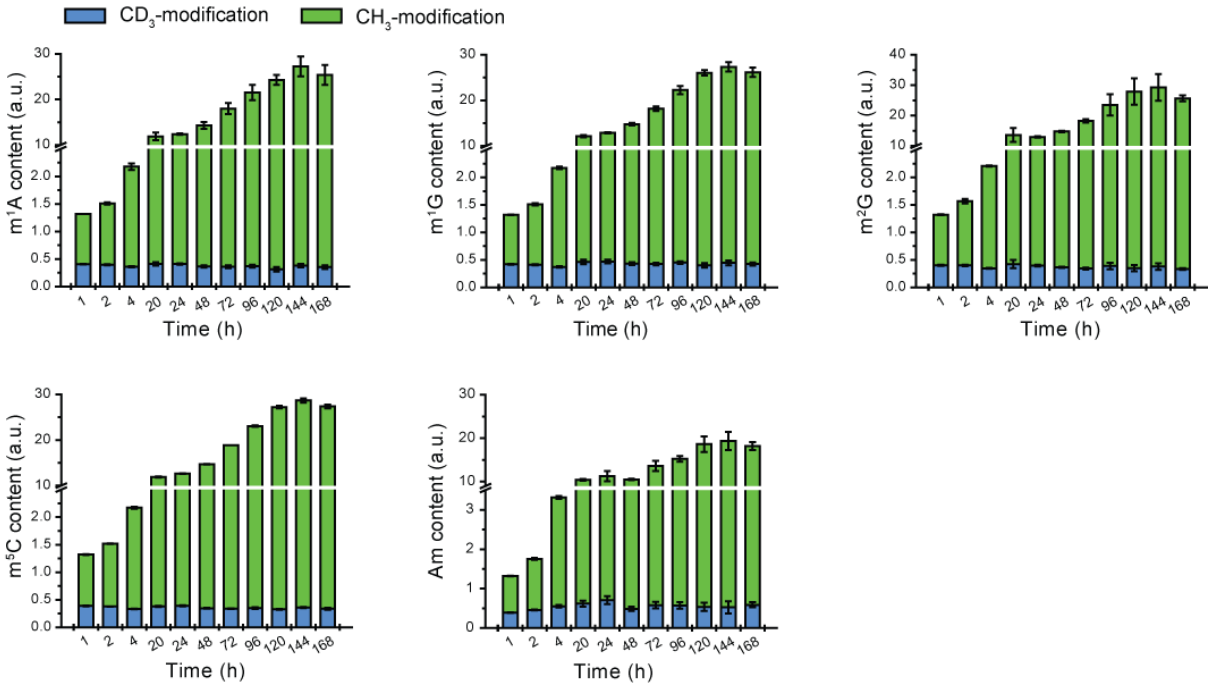


Figure 18 Evaluation of (CD_3)- and (CH_3)-modification content over time. Heavy and light modification content for m^1A, m^1G, m^2G, m^5C and Am was evaluated based on average ratios and normalized modification content as estimated from UV absorbance, and finally multiplied by the optical density to account for increasing cell amount (see Table 9).

Surprisingly, as it is evident from Figure 18, the (CD_3)-modification content of all modified nucleosides analyzed remains largely constant over the seven days. If turnover of labelled tRNA was taking place, we would expect the signal for (CD_3)-modification to gradually decrease, while this is not observed, implying that the labelled tRNA generated in the initial medium is present until the end of the experiment. This suggests that, based on a simplified model which only accounts for increasing cell number, the modified (labelled) tRNA samples are subject to very slow or no turnover under these experimental conditions.

The heavy and light modification content was also assessed using a more complex model which additionally accounts for the varying tRNA content per OD unit (see Table 1). As an example, comparison of the two different approaches in the quantification of the (CD_3)- and (CH_3)-content for m^1A is shown in Figure 19.

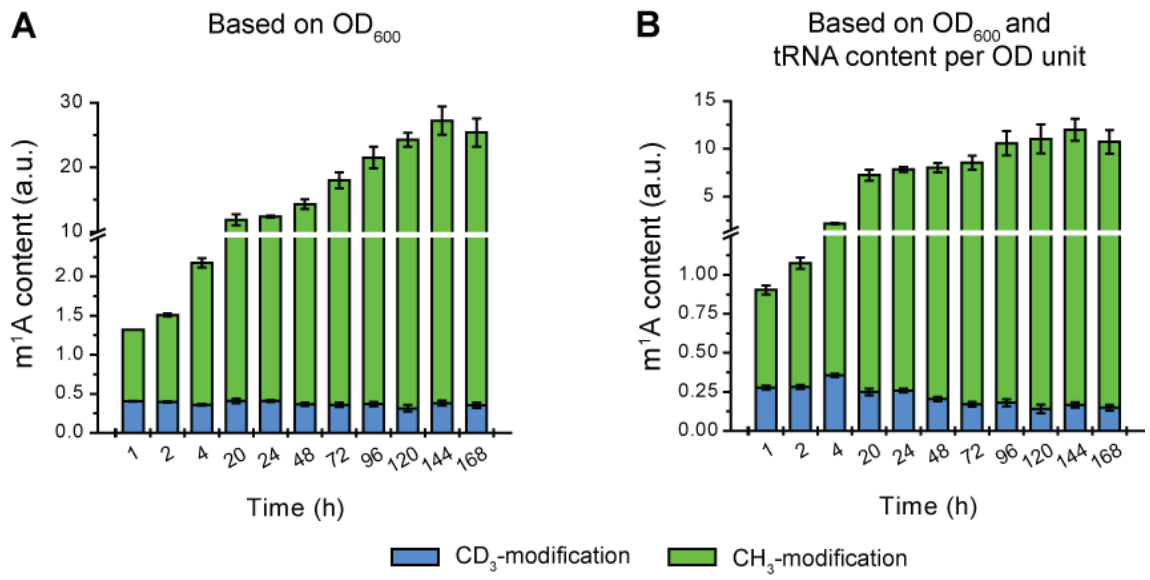


Figure 19 Comparison of (CD_3)- and (CH_3)-content evaluation based on A) OD and B) OD and tRNA content per OD unit. Heavy and light modification content for m^1A was evaluated based on average ratios and normalized modification content as estimated from UV absorbance. The two different models estimate the absolute (CH_3)- and (CD_3)-modification content over time by taking into account the optical density (plot A) versus optical density and variable tRNA content (plot B). Note that modification content is depicted in a.u. and units are not comparable between the different models (see Table 10).

As evident from Figure 19B, a model which accounts for variations in tRNA content suggests that there might be indeed some extent of turnover of labelled tRNA modifications. However, the current experimental set up does not allow to decouple the turnover of modified tRNA nucleosides from overall tRNA turnover (including un-/hypomodified tRNAs). This issue could be addressed by double labelling of the tRNA modifications and of the tRNA molecule, therefore allowing to accurately account for tRNA turnover with respect to modification content.

In conclusion, evaluation of the overall (CH_3)- and (CD_3)-modification content over time, using a simplified model which only accounts for changes in cell number, suggests that the turnover of modified (labelled) tRNA is slow or absent under these experimental conditions.

4.3 Discussion

The observed increase in tRNA modification reported in this chapter and in previous studies done by Dr. Antje Hienzsch,[138] raises a few interesting questions. How is the increase in modifications afforded by yeast cells? Are modifications the result of an enhanced modification of the tRNA population, or are they rather the result of passive enrichment of modified tRNAs? What is the benefit of a more modified tRNA pool?

tRNA molecules are known to be very stable,[145] and previous studies have reported half lives of 50 h in chicken muscle,[146] 44 h in *Euglena gracilis*[147] and 3 to 5 days in *Physcomitrella patens*,[148] while, to the best of our knowledge, no data is available concerning the turnover rates of tRNA in *S. cerevisiae*. One key finding presented in this chapter is the fact that modifications introduced in

tRNA molecules during exponential growth in labelled medium are detected throughout the entire duration of the study. Absence of a recycling pathway has allowed to use detection of modified tRNA nucleosides in urine as a reliable measure of tRNA turnover in higher eukaryotes[149]. Given that no recycling pathway is known for yeast, we deduce that labelled tRNA is being degraded only very slowly, and that, more generally, turnover of (modified) tRNA is rather slow in yeast cultures entering stationary phase. It should be noted, however, that a model which accounts for changes in tRNA content over time does suggest some turnover of modified tRNA nucleosides (Section 4.2.4), and this aspect should be further investigated by double-labelling of the tRNA molecule and tRNA modifications to highlight whether the respective turnover rates are coupled or not.

In fact, one possible reason for the low turnover rates observed in this study might be that, contrary to the experimental approach used for evaluation of the turnover in *P. patens* or in *E. gracilis*,[147-148] which relied on labelling of uracil or adenine, our method allows for the selective monitoring of the tRNA modified nucleosides. The longer turnover rates might in this case reflect the enhanced stability of highly modified tRNA molecules compared to unmodified tRNA (which could show higher turnover rates). In fact, tRNA modifications are known to play a crucial role in stabilization of tRNA molecules,[4a, 4b] and lack of specific tRNA modifications was shown to lead to rapid degradation.[4a]

The observed stability of modified tRNA suggests two possible scenarios: on the one hand, we can postulate that there is no tRNA turnover taking place and that tRNA modifications are gradually being introduced in tRNAs; on the other hand there might be some extent of turnover or degradation, but this might be selective for unmodified or hypomodified tRNA, leading to a gradual enrichment of the modified tRNA pool.

From analysis of the CH_3/CD_3 ratios presented in section 4.2.3, our model suggests that starting from day 1 there is not only an increase in modification content (which could arise purely from enrichment of the modified tRNA population), but that, additionally, there is a higher ratio of light to heavy modification content than expected. Given that we observe little or no turnover of labelled modifications (using a simplified model), the higher CH_3/CD_3 ratio observed might reflect a higher degree of modification content which is actively introduced in tRNA molecules starting from day 1. On the contrary, if modified tRNA molecules would simply accumulate due to passive enrichment, the CH_3/CD_3 ratio should increase linearly with cell density and tRNA content (as modelled in section 4.2.3). Therefore, we suggest that tRNA molecules after the diauxic shift might become more extensively modified and that the observed increase in modification content is not (exclusively) resulting from selective degradation of an unmodified tRNA subset.

Active regulation of modification content might be linked to different aspects. On the one hand, increased modification might simply result from the long half-life of the tRNA molecules; on the other hand, it could be due to a variation in catalytic activity of tRNA modifying enzymes during entry into

stationary phase or to selective modification of previously hypomodified tRNA subsets. Interestingly, modulation of the enzymatic activity could be further coupled to tRNA transport in and out of the nucleus, as discussed later in this section. One aspect which cannot be resolved using this approach and that might be interesting to investigate in the future, is whether modifications are exclusively introduced in newly produced tRNA molecules, or whether previously generated tRNA molecules (e.g. labelled tRNAs) might also be further modified over time.

However, from our study, we cannot exclude that tRNA degradation might play some role in enrichment of the tRNA population. In fact, investigation of the tRNA turnover was based only on methylated RNA nucleosides, excluding methylated uridines, and might therefore reflect the turnover rates of only a subset of tRNAs. In fact, although $tRNA^{Gly(UCC)}$ and $tRNA^{Gly(CCC)}$ are the only yeast tRNAs known to contain none of the methylated nucleosides investigated in this study,[1b, 150] we cannot exclude that some tRNAs might be hypomodified at some point during cell growth. Hypomodified tRNAs might then be either further modified, leading to an increase in (unlabelled) modification content, or could on the contrary be selectively degraded, contributing to a higher modification content by passive enrichment without affecting the levels of labelled modification. In order to accurately establish the contribution of tRNA degradation to the modification increase, our study should be complemented by studies which independently assess the extent of tRNA turnover. This would allow to unambiguously determine the extent to which tRNA is further modified as opposed to simply enriched by selective tRNA degradation.

Concerning the possible degradation pathways, tRNAs are known to be targeted by three different turnover mechanisms. Two of these pathways, the nuclear surveillance pathway involving the TRAMP complex, and the rapid tRNA decay (RTD) pathway, are known to generally degrade hypomodified or incorrectly processed tRNAs.[4a, 34a] However, the third pathway involves mature tRNAs, which undergo endonucleolytic cleavage in response to a variety of stresses.[42a, 43] In this case, Rny1, a member of the RNase T2 family, was shown to cleave tRNAs at the anticodon position in response to a variety of stress conditions, including growth into stationary phase.[151] However, previous reports have pointed out that only a small portion of tRNAs undergo endonucleolytic cleavage and that full length tRNA levels do not decline significantly.[42a, 43] Therefore, we can deduce that this is unlikely to be a major pathway for selective enrichment of the modified tRNA pool during entry into stationary phase, and that (hypomodified) tRNAs are more likely to be processed by alternative degradation pathways. However, one interesting implication of tRNA endonucleolytic cleavage in response to stress is the fact that, although yet unknown in yeast, there might be a RNA repair pathway similar to that observed in bacteria.[152] More specifically, Chan *et al*. suggested that tRNA repair is more efficient if coupled to 2´-O-methylation at the repair site (in the anticodon loop), and that this modification stabilizes tRNA molecules against further cleavage. Hen1, which is responsible for the repair in bacteria when complexed with Pnkp, is also found in eukaryotes and it is known to perform

the same chemical reaction although on different RNA substrates.[153] One interesting possibility is that the observed increase in Gm detailed in section 4.2.2 might be the result of a repair pathway aimed at stabilizing tRNA molecules against endonucleolytic cleavage during entry of yeast cell into quiescence.

Another interesting aspect of growth into stationary phase is the fact that yeast cells have been shown to accumulate cytoplasmic tRNAs in the nucleus during glucose starvation, and transport was shown to be reversible once nutrients become again available.[154] The exact role of retrograde tRNA transport is still unresolved but one of the main hypothesis put forward is that tRNAs are imported into the nucleus for storage or sequestration of the tRNA molecules from the translational apparatus during starvation.[40] Furthermore, it has been suggested that transport into the nucleus might serve as a proofreading mechanism, allowing the cell to eliminate defective and possibly undermodified tRNA.[40] Therefore, sequestration of tRNAs during growth of the cells into stationary phase by import into the nucleus could be a possible mechanism by which cells contribute to a selective enrichment of mature, fully modified tRNAs. This could either be linked to a regulated increase in modification of tRNA molecules to ensure tRNA retrograde transport into the nucleus, or, on the contrary, could be coupled to degradation of hypomodified tRNAs, leading in both cases to a more modified tRNA population.

tRNA modifications as a tool to modulate efficient protein translation during stationary phase

Yeast cells adapt their growth conditions depending on the nutrient availability in the surrounding environment. When grown to saturation in rich medium, the cells are known to enter quiescence, a resting state where cells stop proliferating and where they are able to survive for long periods of time under harsh conditions, while at the same time being able to rapidly proliferate again as soon as nutrients become available.[155] Entry into quiescence is governed by a network of signalling pathways which lead to a coordinated down-regulation of the overall transcription and translation machinery. Ribosomal biosynthesis, which during exponential growth accounts for about 60 % of the total transcription and for 50 % of Pol II transcription, is strongly down-regulated in response to the decrease nutrient availability,[156] overall transcription rates are about three to five times lower than in log-phase, and overall protein biosynthesis in stationary phase is reduced to about 0.3 % of that observed during exponential growth.[155a] Nevertheless, despite the reduced expression of a large subset of genes, there is also a group of genes, particularly those involved in stress response,[157] whose up-regulated expression is essential to ensure correct entry into quiescence and for proper life-span regulation.[155]

tRNA modifications have been shown to enhance translational efficiency by modulating codon-anticodon interactions, as well as ensuring structural stabilization of the tRNA molecule and favouring cognate tRNA selection at the ribosome A site.[11a, 11b, 18a] Furthermore, as detailed in chapter

3,[158] a higher tRNA modification content was found to correlated with an increase protein translation efficiency. Therefore, in the case of yeast entering stationary phase, one possible explanation for an increase in the tRNA modification content might be the need to ensure optimal translation of essential genes in a state where the translational apparatus is relatively inert.[157a, 159] In particular, previous studies have highlighted a role of tRNA modifications in modulation of the expression of genes involved in stress-response. In a recent publication, Chan *et al*. have shown that deletion of tRNA methyltransferase Trm4, responsible for generation of m^5C in tRNA, confers hypersensitivity to oxidative stress.[45] In fact, the increase in m^5C at the anticodon position of $tRNA^{Leu(CAA)}$ observed in yeast cells exposed to H_2O_2 was found to enhance the translation of stress-response proteins which are enriched in TTG codons.[45] This would fit with the observed increase in m^5C content detailed in this study, which is also in agreement with the observed increase in $m^2{}_2G$ described in the original publication investigating the effect of H_2O_2 exposure on tRNA modifications.[44a] In another study, Trm9 was found to modulate stress response upon exposure to methyl methanesulfonate (MMS) by generation of the modified uridine mcm^5U and mcm^5s^2U. These are found at the wobble position of $tRNA^{Arg(UCU)}$ and $tRNA^{Glu\ (UUC)}$, and an increase in modification was found to be essential to enhance translation of stress-response proteins where specific arginine and glutamic acid codons were overrepresented.[46]

Codon bias, which describes the different frequency with which synonymous codons are found in coding DNA, is known to vary strongly between species. However, variation in codon usage is also known within the same organism depending on the gene and it has been implicated in gene expression regulation.[48, 160] In particular, it has been noted that functionally related genes tend to display similar codon bias patterns, and that variations in codon usage are exploited to selectively modulate expression of certain genes during cell cycle development. Codon usage is generally correlated with tRNA abundance (and to some extent with tRNA gene copy number);[161] yet there is increasing evidence, as detailed in the previous paragraph, that tRNA modifications might contribute in the regulation of the composition of the proteome.[45-46] In this way, selective modification of the anticodon loop might change the subset of ´preferred´ codons by modulating codon-anticodon interactions, therefore favouring translation of normally ´less preferred´ codons. In this case, the increase in tRNA modification observed in this study might constitute a mechanism to modulated selective expression of a subset of genes required for survival during stationary phase by favouring translation in the presence of different codon usage patterns. Alternatively, modulation of protein synthesis might be affected by altered aminoacylation. In fact, it has been reported that oxidative stress leads to an increase misacylation of methionine, resulting in altered protein composition synthesis as a protective measure against oxidative damage.[44b] In relation to the current study, one possibility is that increased tRNA modifications might influence misacylation and lead to similar alterations in protein composition as part of a coordinated stress response mechanism, as previously suggested.[44a]

Alternatively, the higher modification content of tRNA molecules might play a role in exiting from stationary phase, where quiescent cells rapidly re-start transcription and translation in the presence of nutrients.[155a, 162] In line with this argument, findings by Radonjic *et al*. have revealed that, despite the global transcription repression during quiescence, the transcription machinery is present in an inactive form, where RNA Pol II is poised upstream of hundreds of genes ready to rapidly initiate transcription as soon as required.[163] Entry into stationary phase was also found to lead to an increase in EGP-bodies, a set of cytoplasmic granules which are responsible for mRNA storage during periods of translational inactivity.[164] These bodies are suggested to selectively preserve mRNA molecules which can then be rapidly translated upon new cell proliferation, allowing for more rapid changes in protein concentrations independently of the transcription machinery.[165] In line with the argument that tRNA modifications allow for more efficient translation and might additionally favour specific patterns of codon usage, storage of highly modified tRNAs might therefore allow for more effective translational activity upon exit from stationary phase. Furthermore, the enhanced stability afforded by the increased modification content might ensure stabilization of tRNA molecules from degradation or endonucleolytic cleavage.[4a, 4b]

The following table summarizes the aforementioned implications which might result from an increase in tRNA modification content.

Roles of tRNA modifications		**Implications of increase in tRNA modification levels during diauxic shift**
Known to enhance translational efficiency, modulate codon-anticodon interactions, reduce frameshift and misincorporation[11a, 11b, 18a]	→	Ensure optimal translation of essential genes required during quiescence (e.g. genes involved in stress response and life-span regulation)
tRNA modification contribute to stabilization of tRNA molecules[4a, 4b]	→	Stabilization and storage of tRNA during quiescence
Loss of Trm4 leads to hypersensitivity to H_2O_2;[44a, 45] loss of Trm9 leads to hypersensitivity to MMS[46]	→	Ensure optimal translation of genes involved in stress response
Specific modifications enhance translation of subset of genes enriched for a specific codon[45-46]	→	Contribute to modulation of codon usage by modulation of wobble capacity of specific modifications
Oxidative stress results in increased methionine misacylation as protection against oxidative damage[44b]	→	Might contribute to alteration in aminoacylation
Transcription machinery is present in poised state during quiescence;[163] mRNA molecules are stored in EGP-bodies[164]	→	Ensure effective translation upon exit from quiescence

Summary Table: Overview of possible implications of increase tRNA modification content during growth into stationary phase.

In conclusion, this study confirms that yeast tRNAs become increasingly modified during entry into stationary phase. Using a simplified approach which relies on labelling of the modified tRNA nucleosides at the early phase of exponential growth, we were able to show that modified tRNAs have low turnover and that tRNA molecules are increasingly modified after diauxic shift. On the one hand, this suggests that modification content might increase over time due to the long permanence of tRNAs in the cell. On the other hand, we propose that, if there is some extent of tRNA turnover taking place, that it is mostly targeting hypomodified tRNAs, therefore leading to an enrichment of the modified tRNA pool. Further studies assessing the extent of tRNA turnover with respect to turnover of tRNA modifications will contribute to further elucidate the nature of the observed increase in tRNA modification levels.

4.4 Materials and Methods

4.4.1 Growth and handling of yeast

Yeast *S. cerevisiae* (strain 70449) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen, GmbH (DSMZ). Glycerol stocks were stored at -80 °C. Yeast was grown on YPD-Agar plates at 30 °C for 3 days and subsequently stored at 4 °C. For the labelling experiment, yeast cells were grown for two generations in labelled complete minimal medium (yeast nitrogen base – AA/AS, ammonium sulfate, 2 % dextrose and amino acid powder mix containing [*methyl*-D_3]-L-methionine)[166] and finally inoculated in 1 L of labelled medium. Cells were allowed to grow at 30 °C shaking vigorously (200 rpm in baffled flasks) to an optical density OD_{600} = 1, at which point they were briefly pelleted and resuspended in YPD. Cells were allowed to grow at 30 °C shaking for the following 7 days, and samples were collected at time points 0 h (prior to medium exchange) and 1, 2, 4, 20, 24, 48, 72, 96, 120, 144 and 168 h after resuspension in YPD.

For sample collection, 25–50 ml of yeast culture were removed, optical density was measured, and cells were pelleted at 4 °C, washed once in tRNA Extraction Buffer (10 mM $Mg(OAc)_2$, 50 mM NaOAc, 150 mM NaCl, pH 4.5) and stored at -80 °C after freezing in liquid nitrogen.

4.4.2 tRNA extraction

tRNA extraction was performed as described previously with minor adaptations.[142] Briefly, cells were resuspended in 5 mL of tRNA Extraction Buffer and mixed with 5 mL 80 % aq. phenol. The suspension was vortexed vigorously and incubated at 65 °C for 1 h, vortexing every 10 minutes. The two phases were separated by centrifugation (1 h, 4 °C, 4000 rpm), and the aqueous phase was again incubated with 5 mL of 80 % aq. phenol for 15 min at 65 °C. After centrifugation (30 min, 4 °C, 4000 rpm), the aqueous phase was extracted twice with 2 mL of $CHCl_3$, shaken vigorously and centrifuged to separated the two phases (10 min, 4000 rpm, 4 °C). The aqueous phase was mixed with 20 % KOAc, pH 4.5 (0.1 vol) and with 12 M LiCl solution (0.2 vol). Samples were incubated for 4 h on ice to precipitate genomic DNA and long RNAs and finally centrifuged (30 min, 4 °C, 18000 rpm).

The supernatant was mixed with absolute EtOH (3 vol) and precipitated overnight at -20 °C. Finally, tRNA samples were precipitated by centrifugation (1 h, 4 °C, 12000 rpm), the supernatant discarded, and pellets were resuspended in 3 mL of DEAE Buffer I (100 mM Tris HCl, pH 7.5, 10 mM $MgCl_2$).

4.4.3 tRNA purification by ion-exchange chromatography

Samples resuspended in DEAE Buffer I were further purified by anion exchange chromatography. All purification steps were performed at 4 °C. Crude tRNA samples were loaded on a *DEAE Sepharose Fast Flow* column (5 mL, *GE Healthcare*) using a *ÄKTA purifier*, and tRNA samples were eluted with a gradient from 0 % to 40 % DEAE Buffer II (100 mM Tris HCl, pH 7.5, 10 mM $MgCl_2$, 1 M NaCl). Samples eluted between 0.2–0.4 M NaCl, and showed a characteristic absorbance ration of 2:1 for absorbance at 254 nm with respect to 280 nm. Fractions were combined and precipitated by addition of absolute EtOH (3 vol) and incubation overnight at -20 °C. Purified tRNA was collected by centrifucation (1 h, 4 °C, 12000 rpm), pellets were air-dried and resuspended in 1 mL of ddH_2O. Concentrations of the tRNA samples were evaluated by UV absorbance at 260 nm using a *Nanodrop UV-spectrometer*.

4.4.4 Enzymatic digestion of tRNA samples

Prior to enzymatic digestion, tRNA samples were denatured at 99 °C for 3 min and rapidly cooled on ice to prevent refolding. For each sample, 12 µg of tRNA (in 100 µL of ddH_2O) were digested. After denaturation, samples were mixed with 10 µL of Digest Buffer A (300 mM ammonium acetate, 100 mM $CaCl_2$, 1 mM $ZnSO_4$, pH 5.7) and digested with nuclease S1 (4 µL, 80 units, *Aspergillus oryzae*, *Sigma Aldrich*) for 3 h at 37 °C. Subsequently, samples were mixed with Digest Buffer B (12 µL, 500 mM Tris HCl, 1 mM EDTA, pH 8.0) and digested with Antarctic phosphatase (2 µL, 10 units, *New England Biolabs*) and snake venom phosphodiesterdase I (2 µL, 0.2 units, *Crotalus adamanteus venom, USB Corporation*) for further 3 h at 37 °C. After enzymatic digest, labelled isotopes standards Ar, i^6A and t^6A were added to the mixture.[142] Samples were centrifuged (15 min, 13000 rpm), concentrated and analyzed by LC-MS. For each sample, three independent digestions were performed.

4.4.5 HPLC-ESI-MS

Digested samples were analyzed using a *Dionex Ultimate 3000 HPLC* system coupled to a *Thermo Finnigan LTQ Orbitrap XL* as detailed before.[142, 158] The flow rate was set to 0.15 mL/min and the column was kept at a temperature of 30 °C. Buffer A (2 mM $HCOONH_4$ in H_2O (pH 5.5)) and Buffer B (2 mM $HCOONH_4$ in H_2O/MeCN 20/80 (pH 5.5)) were used for sample elution with the following gradient: 0 → 55 min, 0 % → 8 % Buffer B; 55 → 100 min, 8 % → 60 % Buffer B; 100 → 102 min, 60 % → 100 % Buffer B; 102 → 120 min, 100 % Buffer B; 120 → 125 min, 100 → 0 % Buffer B; 125 → 135 min, 0 % Buffer B. Elution was monitored by UV absorbance at 260 nm. The eluent was injected into the ion source without prior splitting and ions were scanned using a positive polarity mode with a range of *m/z* 200-1000 with resolution of 30,000. The mass spectrometer was set with the

following parameters: sheath gas flow rate, 16 arb; auxiliary gas flow rate, 10 arb; sweep gas flow rate, 4 arb; spray voltage, 5.0 kV; capillary temperature, 200 °C; capillary voltage, 25 V; tube lens, 60 V.

4.4.6 Quantification of modified nucleosides and evaluation of light to heavy modification ratio

Quantification of the modified nucleosides was performed using three different approaches depending on the nature of the modification. Nucleosides i^6A, t^6A and Ar were quantified by means of labelled isotope standards as previously described using a *Qualbrowser* program.[142, 158] Quantification of m^1A, m^1G, m^2G, m^5C, Am and $m^2{}_2G$ was based on absorbance at 260 nm. Calibration curves were generated by HPLC-ESI-MS using synthetic standards and integration of the corresponding UV absorbance signal (see Figure 20). The corresponding linear fits (obtained using *ORIGIN*®) were used to determine relative changes in modification content. Each technical replicate was analyzed in series for samples from time point 0 h to 168 h on three different days, and relative changes were calculated within each individual series. Quantification of nucleoside A was done by means of UV absorbance using the calibration curve previously measured by Dr. Veronika Reiter.[167] Modifications Gm, Cm and m^7G were quantified by comparison of the sums of the areas of labelled and unlabelled nucleosides (CH_3- and CD_3-, determined using *Qualbrowser*) with those of m^1G (for Gm) and m^5C (for Cm and m^7G). The relative change was then normalized with respect to the relative changes of m^1G and m^5C as calculated from UV absorbance normalized to the absorbance of A (see Table 1). Irrespective of the quantification method, modification content was in all cases normalized to the A nucleoside content which was estimated from UV absorbance. For the evaluation of the CH_3- and CD_3-modification proportion in each sample, ratios of the CH_3- to CD_3-modification content were estimated by comparison of the areas of the labelled and unlabelled nucleosides using *Qualbrowser*. Formulae used in the analysis of CH_3/CD_3 ratios are listed in Table 1.

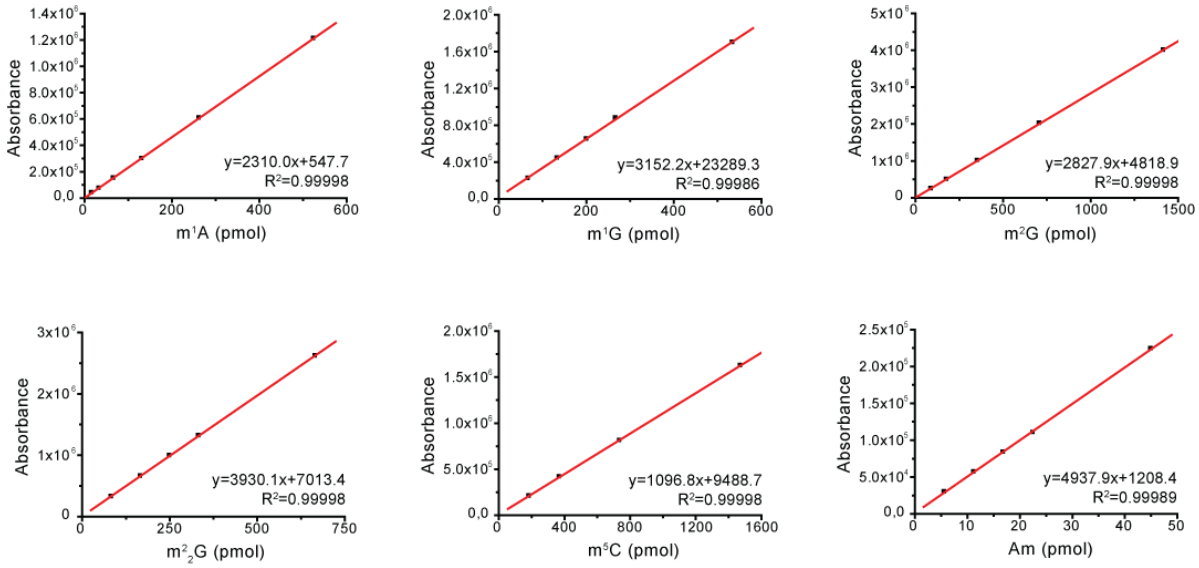


Figure 20 Calibration curves for m^1A, m^1G, m^2G, $m^2{}_2G$, m^5C and Am based on absorbance at 260 nm. Linear fitting was performed using *ORIGIN*®.

Section	Parameter to calculate	Formulae
4.2.2	Quantification of Gm with respect to m^1G; similar procedure for Cm and m^7G with respect to m^5C	= [MS(CH_3-Gm + CD_3-Gm) / MS(CH_3-m^1G + CD_3-m^1G)] * [UV(m^1G) / UV(A)]
4.2.3	Expected CH_3/CD_3 ratio based on OD_{600}	Expected CD_3-content (1 h) = OD(1 h) / [CH_3/CD_3 (1 h, average experimental ratios) + 1] Expected CH_3/CD_3 ratio (x h) = [OD(x h) – CD_3-content(1 h)] / (CD_3-content(1 h))
4.2.3	Expected CH_3/CD_3 ratio based on OD_{600} and tRNA content	Expected CD_3-content (1 h) = tRNA(1 h) / [CH_3/CD_3 (1 h, average experimental ratios) + 1] Expected CH_3/CD_3 ratio (x h) = [tRNA(x h) – CD_3-content(1 h)] / (CD_3-content(1 h))
4.2.4	Evaluation CH_3- and CD_3-modification content based on OD_{600} (e.g. m^1A)	CD_3-m^1A content (x h) = [m^1A(UV, x h, n) / [CH_3/CD_3-m^1A(x h) + 1]] * OD(x h) x CH_3-m^1A content (x h) = [m^1A(UV, x h, n) - m^1A(UV, x h, n) / [CH_3/CD_3-m^1A(x h) + 1]] * OD(x h)
4.2.4	Evaluation CH_3- and CD_3-modification content based on OD_{600} and tRNA content (e.g. m^1A)	CD_3-m^1A content (x h) = [m^1A(UV, x h, n) / [CH_3/CD_3-m^1A(x h) + 1]] * OD(x h) * tRNA(x h, %) CH_3-m^1A content (x h) = [m^1A(UV, x h, n) - m^1A(UV, x h, n) / [CH_3/CD_3-m^1A(x h) + 1]] * OD(x h) * tRNA(x h, %)

Table 1 Summary calculations. Table lists formulae used for evaluating CH_3/CD_3 ratios and CH_3- and CD_3-modification content as detailed in Sections 4.2.3 and 4.2.4. Parameters listed in the table: MS(x) = integral ion count; UV(a) = quantification of a based on UV absorbance; tRNA(x h) = tRNA content at time x; m^1A(UV, x h, n) = m^1A content at time x based on UV absorbance, normalized to time 1 h; CH_3/CD_3-m^1A(x h) = CH_3/CD_3 ratio for m^1A at time x.

4.5 Quantification data

4.5.1 tRNA content per OD unit during growth of *S. cerevisiae*

Time (h)	OD_{600}	tRNA content per OD unit (%)	St.Dev. for tRNA content	St.Dev. for tRNA content %
0	0.92	62.58	1.31	2.09
1	1.32	68.38	3.13	4.57
2	1.58	71.18	2.80	3.94
4	2.46	100.00	0.00	0.00
20	11.61	61.04	2.29	3.75
24	12.18	63.27	1.97	3.12
48	13.00	56.09	1.93	3.44
72	15.96	47.45	2.43	5.11
96	18.56	49.12	4.60	9.36
120	21.09	45.38	5.97	13.15
144	21.60	44.04	2.42	5.49
168	20.52	42.27	3.26	7.72

Table 2 Optical density and tRNA content per OD unit over time (see Figure 11). tRNA content was evaluated by two independent measurements using the *Nanodrop UV-spectrometer* prior and after ion-exchange purification, and measurement of the maximal absorbance at 260 nm during chromatographic purification of the tRNA samples. tRNA content was normalized to the highest value (time point 4 h).

4.5.2 Quantification of modified nucleosides

Time (h)	Normalized m^1G content	St.Dev.	St.Dev. %
0	0.98	0.001	0.1
1	1.00	0.000	0.0
2	0.96	0.013	1.4
4	0.88	0.008	0.9
20	1.04	0.022	2.1
24	1.06	0.010	0.9
48	1.14	0.020	1.8
72	1.14	0.031	2.7
96	1.20	0.048	4.0
120	1.23	0.028	2.3
144	1.27	0.048	3.8
168	1.28	0.050	4.0

Table 3 Normalized content of m^1G (see Figure 12). Relative change in m^1G modification content. Quantification was performed by UV absorbance at 260 nm with respect to content of A nucleoside and normalized to time point 1 h. Values are averages of three independent digests and measurements.

Time (h)	m^1A change (%)	St.Dev.	m^1G change (%)	St.Dev.	m^2G change (%)	St.Dev.	m^5C change (%)	St.Dev.
0	2.1	8.4	-2.0	0.1	5.0	9.4	-1.4	0.1
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	-4.4	1.5	-4.4	1.3	-0.8	2.5	-3.8	0.5
4	-11.5	2.6	-11.7	0.8	-10.4	0.2	-11.7	0.7
20	2.3	7.3	4.1	2.2	17.1	19.9	2.5	1.1
24	1.7	1.3	5.7	1.0	6.3	2.2	3.9	0.4

Time (h)	i^6A change (%)	St.Dev.	Ar change (%)	St.Dev.	t^6A change (%)	St.Dev.	Am change (%)	St.Dev.
0	4.9	6.2	0.8	4.3	-27.4	2.1	-2.2	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	2.7	4.8	-2.1	4.9	4.9	1.7	10.9	1.3
4	-0.6	1.4	-12.3	5.0	-7.0	1.2	34.8	0.6
20	3.6	7.6	5.9	9.6	-8.4	3.7	-10.2	1.5
24	3.9	4.5	3.7	7.7	-13.6	2.8	-7.4	9.4

Table 4 Percental change in modification content during the first 24 h (see Figure 13). Quantification of absolute modification content per A nucleoside was performed by UV absorbance (m^1A, m^1G, m^2G, m^5C, Am) or by means of isotope labelled standards (i^6A, t^6A, Ar). Values are averages of three independent digests and measurements, and were normalized to time point 1 h.

Time (h)	m^1A change (%)	St.Dev.	St.Dev. %	i^6A change (%)	St.Dev.	St.Dev. %	Ar change (%)	St.Dev.	St.Dev. %
first 24h	0	5,52		0	6,89		0	10,76	
48	12.19	1.87	15.36	5.79	2.74	47.31	28.47	2.69	9.44
72	14.92	3.81	25.51	0.62	5.19	831.94	39.88	0.54	1.35
96	18.19	4.52	24.87	3.90	2.87	73.65	57.93	4.62	7.98
120	19.63	3.22	16.43	7.58	4.05	53.46	79.19	12.62	15.94
144	28.47	5.02	17.63	8.62	3.54	40.99	75.63	12.43	16.43
168	26.07	5.62	21.56	7.58	2.64	34.87	71.65	12.09	16.87

Time (h)	m^1G change (%)	St.Dev.	St.Dev. %	m^2G change (%)	St.Dev.	St.Dev. %	$m^2{}_2G$ change (%)	St.Dev.	St.Dev. %
first 24h	0.0	6.0		0.0	8.9		0.0	0.3	
48	15.1	1.4	9.4	10.6	4.1	38.6	4.3	1.6	35.9
72	15.5	2.3	14.6	11.6	3.7	31.8	4.7	1.1	22.5
96	21.6	3.7	17.1	23.1	10.7	46.3	5.5	0.1	2.5
120	25.2	1.9	7.7	28.5	11.9	41.6	8.2	1.0	11.7
144	28.4	3.7	13.0	31.8	11.4	35.9	10.9	0.5	4.7
168	29.3	3.9	13.4	24.9	3.1	12.6	11.4	0.3	2.9

Time (h)	t^6A change (%)	St.Dev.	St.Dev. %	m^5C change (%)	St.Dev.	St.Dev. %	Am change (%)	St.Dev.	St.Dev. %
first 24h	0.0	8.1		0.0	5.4		0.0	15.3	
48	-8.7	3.3	-38.4	15.0	0.4	2.4	-22.9	0.3	-1.1
72	-0.8	2.4	-297.7	20.4	0.5	2.5	-18.6	4.3	-22.9
96	-5.0	4.2	-84.2	26.6	0.5	1.9	-21.5	1.6	-7.6
120	14.5	3.3	22.9	31.5	0.8	2.5	-15.8	5.1	-32.1
144	68.4	5.9	8.6	35.3	1.3	3.8	-14.5	5.8	-40.4
168	97.9	11.2	11.4	35.9	1.3	3.5	-15.4	2.3	-14.8

Table 5 Percental change in modification content with respect to day 1 (see Figure 14). Modification content during the first 24 h was averaged and used as reference for evaluation of the percental change starting from day 2. Quantification was performed by UV absorbance (m^1A, m^1G, m^2G, m^5C, Am) or by means of isotope labelled standards (i^6A, t^6A, Ar) and normalized to content of A nucleoside. Values are averages of three independent digests and measurements.

Time (h)	Gm change (%) based on m^1G	St.Dev.	Cm change (%) based on m^5C	St.Dev.	m^7G change (%) based on m^5C	St.Dev.
0	-0.3	1.4	-8.4	2.4	-6.2	5.0
1	0.0	0.0	0.0	0.0	0.0	0.0
2	3.3	1.2	-9.7	7.4	-7.1	4.1
4	-6.4	1.3	-12.9	4.2	-15.7	3.4
20	6.6	8.1	-4.0	4.9	0.0	5.4
24	2.4	11.2	-3.2	1.5	-7.0	3.5
48	18.3	8.6	3.9	5.2	4.1	3.0
72	30.5	6.9	8.5	5.4	8.4	6.3
96	43.8	10.5	7.0	8.1	8.1	4.8
120	55.8	7.3	22.3	2.2	17.7	3.1
144	54.6	8.1	13.9	10.1	21.8	6.6
168	52.3	9.5	16.6	8.9	18.1	5.0

Table 6 Percental change of Gm, Cm and m^7G (see Figure 15). Quantification of Gm, Cm and $m^7G$was based on comparison of the total areas of labelled and unlabelled ion count signals with respect to that of m^1G (for Gm) and m^5C (for Cm and m^7G). Values were then normalized based on m^1G and m^5C quantification by UV absorbance and total A nucleoside content. Values are averages of three independent digests and measurements.

4.5.3 Evaluation of CH_3/CD_3 ratio during entry into stationary phase

Time (h)	Average CH_3/CD_3 all mod.	St.Dev.	St.Dev. %	Expected CH_3/CD_3 (OD)	St.Dev.	St.Dev. %	Average CH_3/CD_3 Am	St.Dev.	St.Dev. %
0	1.80	0.11	6.24	-	-	-	2.04	0.11	5.22
1	2.24	0.12	5.25	2.24	0.16	7.07	2.38	0.05	2.11
2	2.84	0.12	4.12	2.87	0.20	7.07	2.82	0.08	2.81
4	5.22	0.29	5.50	5.03	0.36	7.07	5.00	0.30	5.98
20	28.13	2.48	8.82	27.46	1.94	7.07	15.85	1.88	11.88
24	29.12	3.00	10.30	28.85	2.04	7.07	14.91	1.51	10.12
48	37.12	3.33	8.96	30.86	2.18	7.07	20.60	2.31	11.22
72	48.21	4.70	9.74	38.12	2.70	7.07	22.52	2.57	11.41
96	56.74	6.29	11.09	44.49	3.15	7.07	25.64	3.53	13.77
120	74.04	7.04	9.51	50.69	3.58	7.07	33.66	5.70	16.95
144	71.28	7.86	11.03	51.94	3.67	7.07	35.96	9.85	27.39
168	70.34	7.70	10.95	49.29	3.49	7.07	29.94	2.97	9.92

Table 7 Comparison of measured and expected CH_3/CD_3-ratio based on optical density (see Figure 16). Average ratios for all modifications were obtained by averaging values for m^1A, m^1G, m^2G, m^7G, Gm, $m^2{}_2G$ (CD_3-, CH_3-labelled only), m^5C and Cm. Expected ratios were estimated on the assumption that CH_3-modification content increased linearly with OD_{600} and were calculated starting from time point 0 h, with an initial CH_3/CD_3 ratio of 2.24 and OD_{600} = 1.32. Values are averages of three independent digests and measurements.

Time (h)	Average CH_3/CD_3 all mod.	St.Dev.	St.Dev. %	Expected CH_3/CD_3 (OD and tRNA)	St.Dev.	St.Dev. %	Average CH_3/CD_3 Am	St.Dev.	St.Dev. %
0	1.80	0.11	6.24	-	-	-	2.04	0.11	5.22
1	2.24	0.12	5.25	2.24	0.23	10.10	2.38	0.05	2.11
2	2.84	0.12	4.12	3.03	0.27	9.01	2.82	0.08	2.81
4	5.22	0.29	5.50	7.82	0.55	7.02	5.00	0.30	5.98
20	28.13	2.48	8.82	24.40	1.95	7.99	15.85	1.88	11.88
24	29.12	3.00	10.30	26.62	2.05	7.68	14.91	1.51	10.12
48	37.12	3.33	8.96	25.14	1.97	7.84	20.60	2.31	11.22
72	48.21	4.70	9.74	26.14	2.29	8.76	22.52	2.57	11.41
96	56.74	6.29	11.09	31.68	3.77	11.91	25.64	3.53	13.77
120	74.04	7.04	9.51	33.31	5.07	15.23	33.66	5.70	16.95
144	71.28	7.86	11.03	33.09	2.97	8.98	35.96	9.85	27.39
168	70.34	7.70	10.95	30.09	3.19	10.59	29.94	2.97	9.92

Table 8 Comparison of measured and expected CH_3/CD_3-ratio based on optical density and tRNA content per OD_{600} unit (see Figure 17). Average ratios for all modifications were obtained by averaging values for m^1A, m^1G, m^2G, m^7G, Gm, $m^2{}_2G$ (CD_3-, CH_3-labelled only), m^5C and Cm. Expected ratios were estimated on the assumption that CH_3-modification content increased linearly with OD_{600} and with tRNA content, and were calculated starting from time point 0 h, with an initial CH_3/d_3 ratio of 2.24 and OD_{600} = 1.32 and tRNA content per OD unit of 68.38 %. Values are averages of three independent digests and measurements.

4.5.4 Evaluation of CH_3- and CD_3-modification content

Time (h)	Average CD_3-m^1A	St.Dev.	Average CH_3-m^1A	St.Dev.	Average CD_3-m^1G	St.Dev.	Average CH_3-m^1G	St.Dev.
1	0.405	0.004	0.915	0.00	0.420	0.006	0.900	0.006
2	0.394	0.010	1.116	0.02	0.410	0.010	1.100	0.023
4	0.357	0.011	1.820	0.06	0.372	0.014	1.800	0.024
20	0.408	0.031	11.465	0.85	0.464	0.042	11.626	0.255
24	0.409	0.012	11.973	0.16	0.471	0.036	12.401	0.125
48	0.364	0.021	13.943	0.75	0.431	0.026	14.326	0.265
72	0.359	0.026	17.637	1.24	0.425	0.025	17.761	0.500
96	0.369	0.028	21.158	1.66	0.452	0.023	21.811	0.888
120	0.312	0.043	23.978	1.08	0.403	0.041	25.640	0.588
144	0.379	0.033	26.856	2.18	0.446	0.044	26.908	1.035
168	0.352	0.033	25.040	2.18	0.426	0.028	25.745	1.036

Time (h)	Average CD_3-m^2G	St.Dev.	Average CH_3-m^2G	St.Dev.	Average CD_3-m^5C	St.Dev.	Average CH_3-m^5C	St.Dev.
1	0.401	0.008	0.919	0.008	0.388	0.006	0.932	0.006
2	0.399	0.014	1.169	0.042	0.378	0.002	1.141	0.007
4	0.346	0.008	1.860	0.010	0.332	0.005	1.839	0.018
20	0.423	0.075	13.168	2.308	0.380	0.008	11.521	0.125
24	0.394	0.016	12.548	0.263	0.390	0.010	12.262	0.049
48	0.365	0.009	14.371	0.254	0.345	0.009	14.329	0.049
72	0.344	0.018	17.911	0.555	0.339	0.004	18.531	0.030
96	0.389	0.058	23.111	3.518	0.348	0.017	22.721	0.175
120	0.349	0.055	27.551	4.383	0.327	0.007	26.899	0.272
144	0.379	0.061	28.912	4.381	0.359	0.009	28.334	0.450
168	0.336	0.016	25.320	0.976	0.336	0.016	27.041	0.402

Time (h)	Average CD_3-Am	St.Dev.	Average CH_3-Am	St.Dev.
1	0.391	0.008	0.929	0.008
2	0.458	0.014	1.294	0.025
4	0.552	0.033	2.763	0.037
20	0.619	0.074	9.812	0.188
24	0.709	0.102	10.564	1.148
48	0.487	0.055	10.023	0.178
72	0.579	0.083	13.046	1.168
96	0.573	0.082	14.694	0.634
120	0.538	0.105	18.093	1.798

144	0.524	0.154	18.855	2.099
168	0.589	0.065	17.623	0.905

Table 9 Evaluation of CD_3- and CH_3- modification content over time (see Figure 18). Evaluation of heavy and light modification content for m^1A, m^1G, m^2G, m^5C and Am was based on average ratios and normalized modification content, finally multiplied by the optical density to account for increasing cell amount. Values are averages of three independent digests and measurements.

Average CH_3- and CD_3-modification content based on OD

Time (h)	Average CD_3-m^1A	St.Dev.	St.Dev. %	Average CH_3-m^1A	St.Dev.	St.Dev. %
1	0.405	0.004	0.9	0.915	0.00	0.39
2	0.394	0.010	2.4	1.116	0.02	2.23
4	0.357	0.011	2.9	1.820	0.06	3.50
20	0.408	0.031	7.6	11.465	0.85	7.38
24	0.409	0.012	2.9	11.973	0.16	1.35
48	0.364	0.021	5.8	13.943	0.75	5.39
72	0.359	0.026	7.1	17.637	1.24	7.02
96	0.369	0.028	7.7	21.158	1.66	7.86
120	0.312	0.043	13.9	23.978	1.08	4.49
144	0.379	0.033	8.6	26.856	2.18	8.13
168	0.352	0.033	9.4	25.040	2.18	8.69

Average CH_3- and CD_3-modification content based on OD and tRNA content

Time (h)	Average CD_3-m^1A	St.Dev.	St.Dev. %	Average CH_3-m^1A	St.Dev.	St.Dev. %
1	0.277	0.013	4.7	0.626	0.029	4.6
2	0.281	0.013	4.6	0.794	0.036	4.5
4	0.357	0.011	2.9	1.820	0.064	3.5
20	0.249	0.021	8.5	6.998	0.579	8.3
24	0.259	0.011	4.3	7.575	0.257	3.4
48	0.204	0.014	6.7	7.821	0.500	6.4
72	0.170	0.015	8.8	8.368	0.727	8.7
96	0.181	0.022	12.1	10.392	1.270	12.2
120	0.141	0.027	19.1	10.882	1.512	13.9
144	0.167	0.017	10.2	11.826	1.160	9.8
168	0.149	0.018	12.1	10.584	1.230	11.6

Table 10 Comparison of CD_3- and CH_3- content evaluation based on OD and tRNA content per OD unit (see Figure 19). Evaluation of the absolute CH_3- and CD_3-modification content over time by accounting for the changes in optical density OD versus optical density and variable tRNA content. Units are not comparable between the different models. Values are averages of three independent digests and measurements.

5 Investigation of the structure-function relationships for the bifunctional tRNA-modifying enzyme MnmC in complex with substrate tRNA

5.1 Introduction

tRNA molecules are the most highly modified RNA species. Modified nucleosides, which can range from simple methylations to hypermodified nucleosides, are found at various positions and play essential roles in structural stabilization[4b] as well as modulation of codon-anticodon interaction.[11b] Post-transcriptional modifications are particularly frequent at position 34 of tRNA molecules, the so called wobble position in the anticodon, where 48 % of nucleosides are modified in *E.coli* tRNAs. In *E.coli*, U34 is either found as a 5-hydroxyuridine derivative (xo^5U34) or as a 5-methyl(-2-thio)uridine derivative ($xm^5(s^2)U34$) (see Modomics database[1b, 150]). These two sets of modified uridines are responsible for different effects upon codon-anticodon pairing: the xo^5U derivatives expand the wobble capacity to ensure efficient recognition of all four bases A, U, C and G opposite position 34, while xm^5U modifications limit pairing to only purine-ending codons.[14, 15b] One such modification is 5-methylaminomethyl-2-thiouridine (mnm^5s^2U34), found at position 34 of *E.coli* $tRNA^{Glu}$ and $tRNA^{Lys}$ and the 5-methylaminomethyluridine (mnm^5U34) of *E.coli* $tRNA^{Arg}$.

The biosynthetic pathway of $mnm^5(s^2)U34$ is rather complex and it involves a variety of enzymes (see Figure 21A). Thiolation at position 2 of uridine, which can occur independently of modification at position 5,[168] is catalyzed by MnmA and by the cysteine desulfurase IscS.[169] Initial modification at position 5 is performed by MnmE and MnmG (which form a $\alpha_2\beta_2$ heterotetramer), yielding the intermediate 5-carboxymethylaminomethyl(-2-thio)uridine ($cmnm^5(s^2)U34$).[170] Subsequent C-N bond cleavage and methylation of $cmnm^5(s^2)U34$ to yield $mnm^5(s^2)U34$ is performed by the bifunctional enzyme MnmC.[171] MnmC is composed of two domains: the C-terminal domain (MnmC1) is responsible for the initial FAD-dependent cleavage reaction on $cmnm^5(s^2)U34$ to yield the 5-aminomethyluridine intermediate ($nm^5(s^2)U34$), while the N-terminal domain (MnmC2) performs the methylation reaction in the presence of cofactor *S*-adenosyl L-methionine (SAM) to give $mnm^5(s^2)U34$.

MnmC is only found in bacteria (predominantly γ-proteobacteria), and knock-out of this protein leads to reduced growth rates,[172] suggesting that incomplete modification of U34 results in impaired efficiency of protein translation. Furthermore, biochemical kinetic assays revealed that the catalytic machinery of this protein is probably tuned to yield only fully modified $mnm^5(s^2)U34$, given the tighter binding at MnmC2.[172] The crystal structure of *E.coli* MnmC with bound FAD and SAM cofactors was solved only recently.[173] Yet questions remain regarding the selective mode of RNA processing and the extent of cooperativity of the two modification reactions taking place at MnmC1 and MnmC2.

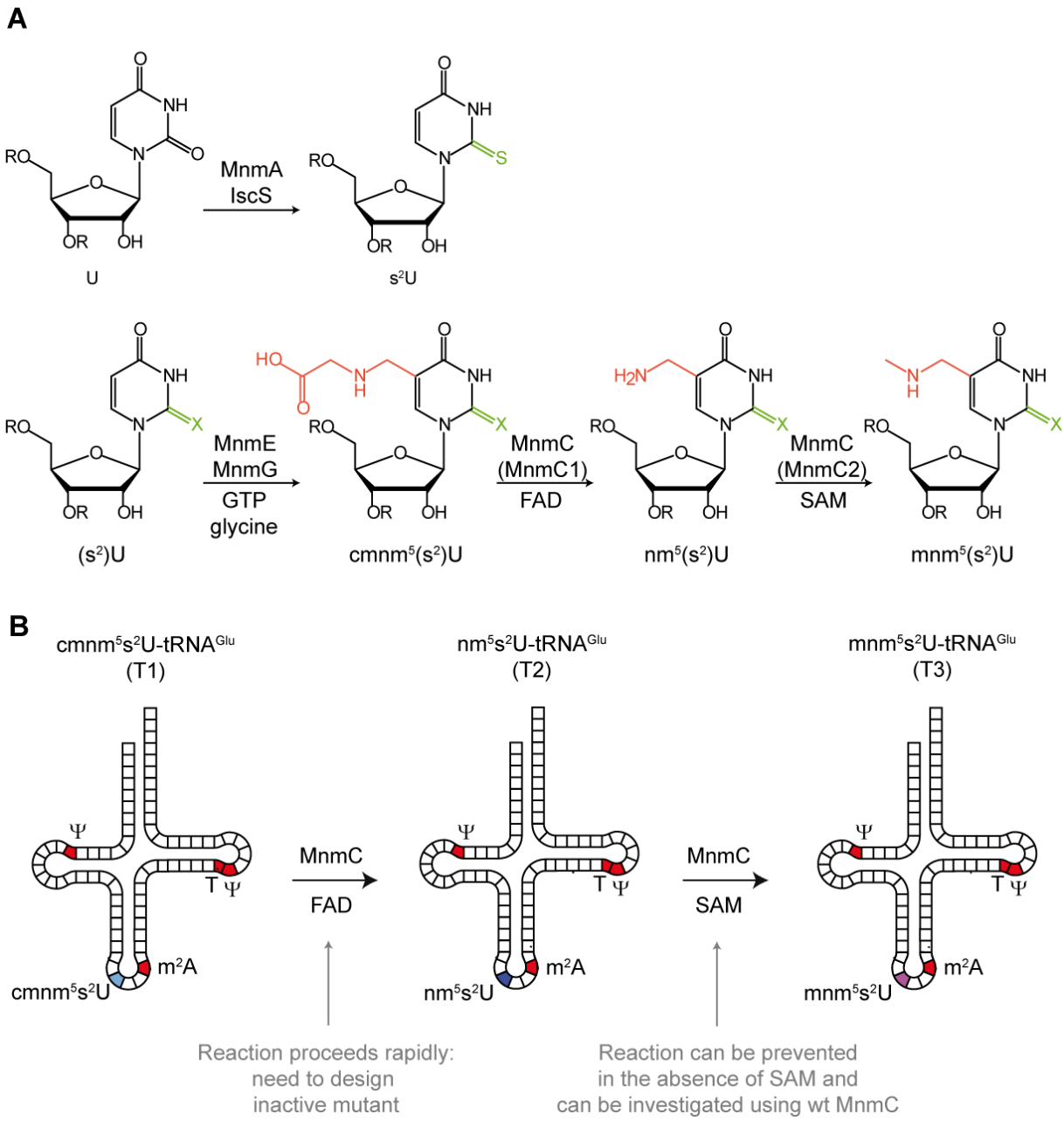


Figure 21 A) Biosynthetic pathway of mnm^5(s^2)U and B) Summary of biosynthetic steps under investigation. In section A) R = RNA, X = O (for U) or S (for s^{2}U). Modification at position 2 is shown in green, while at position 5 in red. B) Description of selectively undermodified tRNA^{Glu} (T1, T2) and fully modified tRNA^{Glu} T3, alongside the biosynthetic steps investigated with wild type (wt) and mutant MnmC.

The main aim of the project was to obtain structural information of the MnmC-tRNA complex, in order to elucidate the biochemical and structural features which ensure selective modification of the uridine at position 34 of *E.coli* tRNA. Our study proceeded along two main lines of investigation: on the one hand characterization of the MnmC-T1 complex, which required prior design of mutant proteins for inhibition of the modification reaction, and on the other hand of the MnmC-T2 complex, where the reaction could be easily prevented in the absence of cofactor SAM (see Figure 21B). Biochemical and structural studies were carried out using selectively undermodified *E.coli* tRNA^{Glu} which allowed for the investigation of the binding and activity of wild type and mutant MnmC in the

presence of substrate cmnm^{5}s^{2}U- and nm^{5}s^{2}U-containing tRNA^{Glu} (T1 and T2, see Figure 21B). The final aim was to obtain detailed snapshots of the sequential modification reactions.

5.2 Results and discussion

5.2.1 Purification of MnmC

Purification of endogenous MnmC from *Escherichia coli* was initially described by Taya and Nishimura in 1973,[174] followed by the subsequent purification of recombinant MnmC from *E. coli* published by Hagervall *et al*. in 1987.[171a] The expression and purification protocol described in this work is an adaptation of a later procedure published by Bujnicki *et al*..[171b] Two different expression vectors were used for recombinant expression of MnmC, pET-28a and pET-30a, both providing an N-terminal poly-histidine tag. The MnmC gene sequence was cloned in the expression vector pET-28a between the NdeI and BamHI restriction sites, yielding a protein with overall molecular weight of 76.6 kDa. Cloning of this construct for wild type MnmC was originally performed by Dr. David Pearson. In the case of pET-30a, the construct was generated by PCR amplification of MnmC using primers that carried the KpnI site before the start codon and EcoRI site after the stop codon. The forward primer was designed to introduce a new sequence coding for an enterokinase cleavage site that would yield, after cleavage of the N-terminal tag, a final MnmC construct with no additional residues at its N-terminus (74.5 kDa), a feature that is ideal for crystallization studies.

Both expression plasmids were transformed in *E.coli* strain BL21 and protein overexpression was performed as detailed in section 5.4.3. MnmC was purified by affinity chromatography using a *HiScreen IMAC FF* column charged with Ni^{2+}, followed by ion-exchange chromatography on a MonoQ column and by a final purification step using a *HiTrap Heparin HP* column (see Figure 22). For MnmC constructs cloned in pET-30a, an additional tag-cleavage step was performed overnight prior to purification using a Heparin resin. Chromatographic purification of MnmC was monitored at the UV wavelengths 260 nm, 280 nm and 360 nm (this last one used to monitor the characteristic absorption of the FAD cofactor), followed by analysis by SDS-PAGE.

The purified protein was concentrated (see section 5.4.3), aliquoted and stored at -80 °C. A yield of about 10 mg of pure protein was obtained per liter of *E.coli* culture.

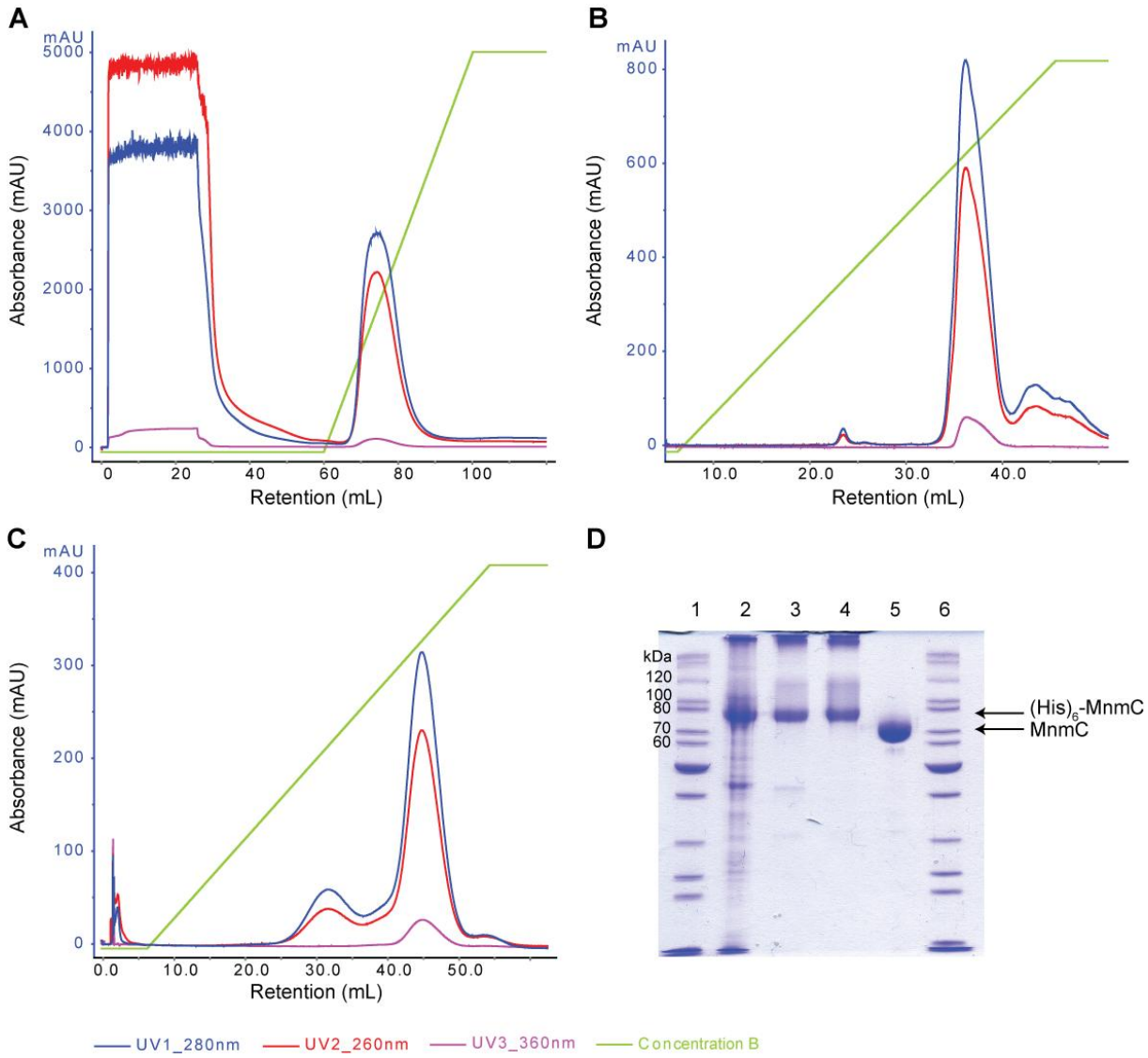


Figure 22 Purification of MnmC (pET-30a construct). Chromatograms display UV absorbance at 280 nm (blue trace), 260 nm (red trace) and 360 nm (pink trace). The green trace indicates the gradient of the elution buffer. A) Affinity purification on *HiScreen IMAC FF* column. B) Ion-exchange chromatography on MonoQ. C) Purification on *HiTrap Heparin HP*. In all cases, elution fractions displaying absorption at 360 nm (characteristic of the FAD cofactor) were pooled for further purification. D) SDS-PAGE of sequential MnmC purification steps, lane 1: protein marker, lane 2: cell lysate, lane 3: pooled fractions after affinity purification, lane 4: pooled fractions after ion-exchange, lane 5: pure protein after enterokinase tag-cleavage and purification on Heparin column, lane 6: protein marker.

5.2.2 Purification of undermodified *E.coli* $tRNA^{Glu}$

Cloning, expression and purification of undermodified $tRNA^{Glu}$ was first established by Dr. David Pearson, who also derived the MnmC knockout ΔMnmC T7 express *E.coli* strain.[172] The knockout strain was used for expression of undermodified $tRNA^{Glu}$ containing $cmnm^5s^2U$ at position 34 (T1, see Figure 21B). This protocol enabled us to isolate selectively undermodified $tRNA^{Glu}$ which carries the characteristic ψ, T and m^2A modifications, alongside the $cmnm^5s^2U34$, therefore allowing to investigate binding and activity of MnmC on substrate tRNA with higher accuracy compared to *in vitro* transcribed tRNA molecules or crude tRNA samples from other species (see Figure 21B).[172]

T1 was overexpressed in the knock-out *E.coli* strain and total tRNA was extracted and precipitated as described in section 5.4.4. The tRNA pellet (about 20 mg per liter of *E.coli* culture) was resuspended in tRNA Buffer A to an approximate concentration of 10 mg/mL, heated to 70 °C for complete deacylation and subsequently purified by anion exchange HPLC chromatography. The first purification of T1 was performed at pH 5.0, the second at pH 8.0, in both cases eluting with a gradient 0–500 mM NaCl (see Figure 23). In order to obtain nm^{5}s^{2}U-containing tRNA^{Glu} (T2), purified T1 was reacted with MnmC (in the absence of SAM) for 1 h at 37 °C and subjected to further chromatographic purification at pH 8.0 as in the previous step. Purification and complete conversion of T1 to T2 prior to final purification was monitored by analytical HPLC. Purified tRNA^{Glu} T1 or T2 (0.5 mg per liter of *E.coli* culture) was concentrated in tRNA Buffer A and stored at -80 °C.

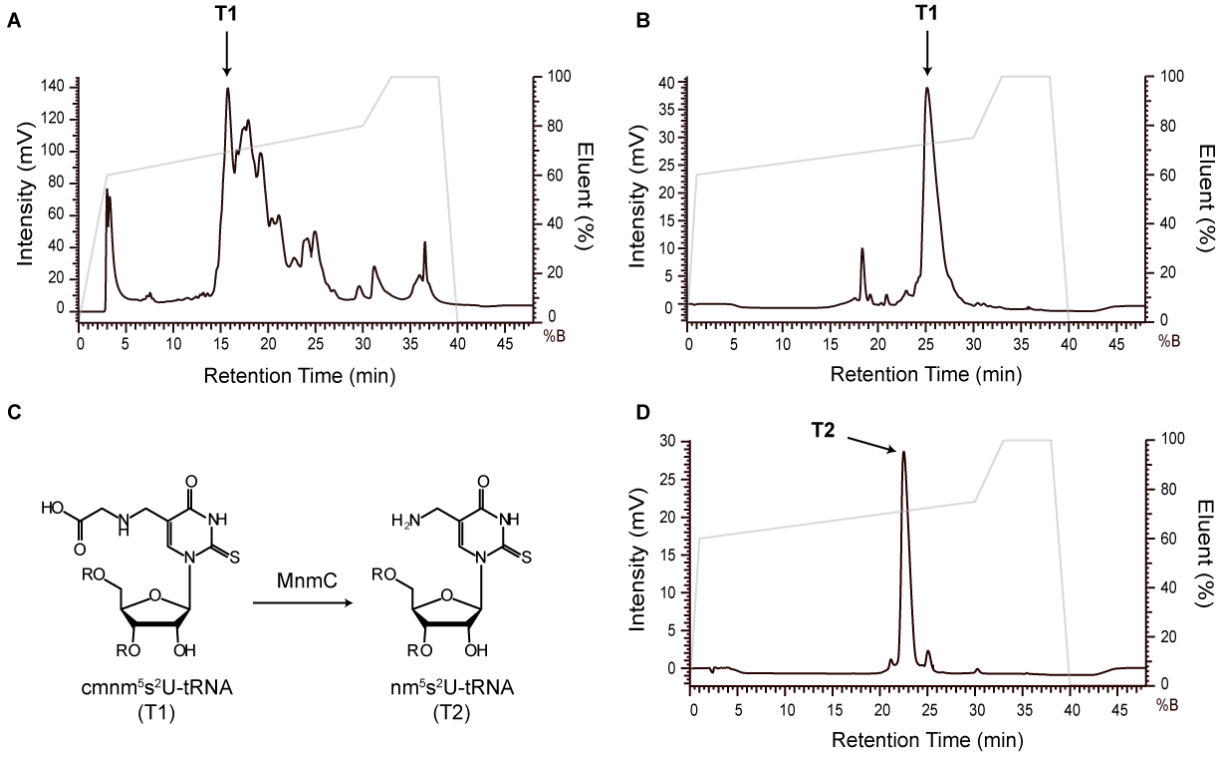


Figure 23 HPLC purification of selectively undermodified *E.coli* tRNA^{Glu} T1 and T2. In all cases purification was performed by ion-exchange chromatography. A) Purification of crude *E.coli* tRNA (tRNA Buffer B, pH 5.0). B) Second purification step to isolate T1 (tRNA Buffer A, pH 8.0). C) Reaction of T1 to T2 in the presence of MnmC in MnmC Reaction Buffer (R = RNA). D) Final purification of T2 (tRNA Buffer A, pH 8.0). Note the characteristic shift in retention time between T1 (about 25 min) and T2 (about 22.5 min)

5.2.3 Characterization of the FAD-binding domain and design of inactive mutants for crystallization studies

The C-terminal domain of MnmC (residues 255-668) is responsible for the FAD-dependent C-N bond cleavage of cmnm^{5}s^{2}U to the intermediate nm^{5}s^{2}U. This domain is characterized by a protein fold typical of the glutathione reductase 2 (GR_2) family of FAD-binding proteins, and it is closely related to the D-amino acid oxidases (DAAO) superfamily,[175] in particular to the glycine oxidase ThiO from *Bacillus subtilis* (PDB 1NG3), as well as to monomeric sarcosine oxidase (PDB 3M13).[173a]

Initial structural modelling studies of MnmC based on protein fold-recognition,[175a] largely confirmed by later determination of the crystal structure of *E.coli* MnmC in complex with FAD (see Figure 24A and B, PDB 3AWI, PDB 3PS9),[173] suggested that residues G271, G272 and G273 are involved in binding of the adenosine moiety of FAD, N306 interacts with the flavin, while residues R567, S617 and R618 are responsible for binding of the substrate base $cmnm^5s^2U$ and its conversion to nm^5s^2U.[175a] The mechanism, suggested to largely resemble that of ThiO, is thought to proceed via a hydride transfer to the N5 of the flavin, facilitated by the nitrogen lone pair of the $cmnm^5$ moiety.[171b, 176] A superposition model of MnmC and ThiO[173a] reveals that R567 might be involved in hydrogen bonding to the carboxylate moiety of $cmnm^5U$, a feature also observed in ThiO where Arg302 coordinates the carboxylate oxygen atoms of N-acetylglycine and ensures correct orientation of the Cα atom with respect to the flavin moiety.[176] From the same superposition model, R618 is suggested to be involved in binding of the tRNA anticodon loop.[173a]

Mutational analysis of MnmC confirmed that R567A and R618A lead to partial inactivation of the C-N bond cleavage reaction, while no mutagenesis data support the role of S617.[175a] However, enzymatic activity assays of these mutants were only performed with total tRNA from *Bacillus subtilis* (due to the natural occurrence of $cmnm^5s^2U$-containing tRNA in this organism and absence of nm^5s^2U or mnm^5s^2U) and no accurate data is yet available concerning their activity on substrate *E.coli* $tRNA^{Glu}$.

5.2.3.1 Cloning, purification and enzymatic activity assay of MnmC mutants R567A, S617A and R618A

In order to design a MnmC mutant which would be suitable for structural characterization of the MnmC-T1 complex, we decided to investigate the extent of inactivation of the R567A, R618A and S617A mutants on conversion of *E.coli* T1 to T2. Protein mutants were generated for both pET-28a and pET-30a constructs using the *Agilent Quick-change Lightning Site-directed Mutagenesis kit*. Given the large plasmid, a different mutagenesis approach was used in order to avoid the potential problems which could arise from amplification of large targets. The method relies on a two-primer, two-stage PCR protocol which initially generates a megaprimer containing the desired mutation, followed by full plasmid amplification.[177] Different mutagenic primers were used for the three desired mutants, while a single, non-mutagenic primer (not complementary to the first one) was used for megaprimer formation for all three constructs. After sequence verification, mutated plasmids were transformed in a BL21 *E.coli* expression strain and mutant proteins expression and purification was performed as detailed in section 5.2.1.

Purified mutant proteins were subsequently tested for enzymatic activity on *E.coli* ΔMnmC mutant-derived crude tRNA to obtain a first estimate of the extent of inactivation. The assay was performed by incubating 20 µg of protein with 20 µg of crude tRNA contaning undermodified $tRNA^{Glu}$ T1 as detailed in section 5.4.5, and samples were subsequently subjected to analytical ion-exchange

chromatography. Reaction progress was estimated by analysis of the peaks corresponding to T1 and T2.

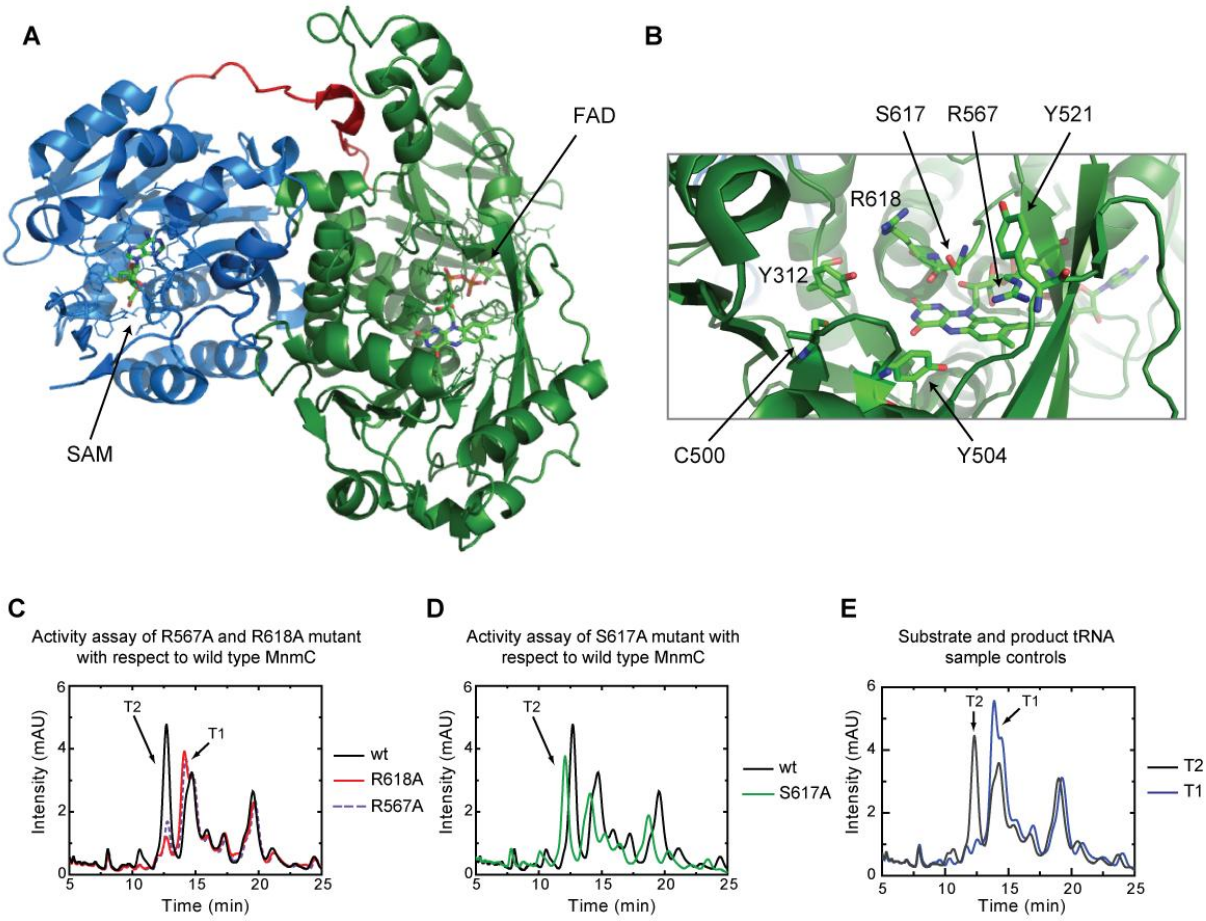


Figure 24 Activity assay of MnmC mutants R567A, R618A and S617A. A) Crystal structure of wild type *E.coli* MnmC bound to cofactors FAD and SAM (PDB 3ps9[173b]). B) Zoom into FAD active site (PDB 3ps9) highlighting residues which were mutated in the course of this study. Panels C-E display chromatograms monitoring the activity of wild type and mutant MnmC R567A, R618A and S617A on crude *E.coli* tRNA containing T1. C) Comparison of wild type, R567A and R618A MnmC. Wild type MnmC leads to conversion of T1 to T2 (large peak at 12.5 min). This conversion is largely suppressed in the case of the two mutant proteins. D) Comparison of wild type and S617A MnmC. Conversion of T1 to T2 is comparable for the two proteins, suggesting that the S617A mutation does not lead to inactivation (Note the slight shift in retention time, arising from small changes in salt concentration upon replenishing of elution buffer. Nevertheless, the two chromatographic profiles look very similar, indicating comparable conversion of T1 to T2). E) Control chromatogram displaying the two unreacted crude *E.coli* tRNA samples, one containing T1 and the other T2, for comparison of the expected profile and retention times for the two species.

Shown in Figure 24 are the chromatograms of the reaction assays for the different mutant proteins. For comparison, unreacted crude tRNA (contaning T1) is presented in panel E alongside tRNA contaning T2, highlighting the expected difference in retention time between T1 (about 14 min) and T2 (12.5 min), which allows to monitor the extent of reaction between the two differently modified tRNA species. As shown in Panel C, R567A and R618A mutant proteins display a large extent of inactivation of the conversion from T1 to T2, as evident from the small peak at retention time 12.5 min (T2) compared to the wild type enzyme and the correspondingly higher peak for T1 observed at 14 min, confirming that both residues are essential for catalysis as previously shown by Roovers

et al..[175a] On the contrary, mutant S617A (in Panel D) shows a chromatogram largely resembling that of the wild type enzyme, therefore suggesting that residue S617 is not essential for catalytic activity and that it might only play a minor role in binding of the target nucleobase in the enzyme active site.

For each mutant protein a second set of reactivity tests was performed to verify that mutagenesis did not compromise the catalytic activity of the N-terminal methyltransferase domain. In all cases, conversion of T2 (nm^5s^2U-$tRNA^{Glu}$) to T3 (mnm^5s^2U-$tRNA^{Glu}$) in the presence of SAM was comparable to that of the wild type enzyme, in line with previous data published by Roovers *et al*. which highlighted the high degree of independence of the two domains in their catalytic activity.[175a]

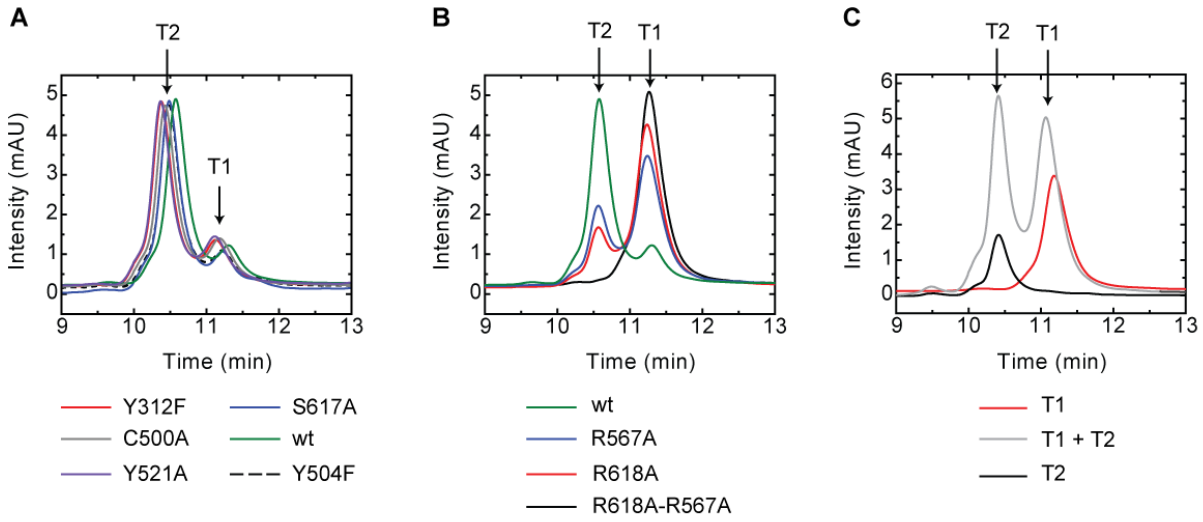
The results shown in Figure 24 therefore confirm that the alanine mutations of residues R567 and R618 can yield largely inactive proteins. In order to verify whether inactivation could be brought to completion upon lowering of the reaction temperature, a second activity test was performed at 4 °C at concentrations comparable to those used for crystallization (200 µM MnmC, 240 µM T1). Incubation overnight followed by analysis by HPLC showed that both R567A and R618A mutants retain partial activity at 4 °C, making these mutants unsuitable for structural characterization since the residual catalytic activity would not allow to obtain structural information on substrate binding to the C-terminal domain.

5.2.3.2 Cloning, purification and enzymatic activity assay of MnmC mutants Y312F, C500A, Y504F, Y521A and double mutant R567A/R618A

Given that the first MnmC mutants proved unsuitable for crystallization studies, a second set of mutants was generated in an attempt to fully inactivate the protein and to further characterize residues in the active site which could be involved in substrate binding and catalysis. The following mutants were therefore generated: the double mutant R567A/R618A, to verify whether combining two inactivating mutations would lead to full suppression of the catalytic activity; Y504F, since Y504 is thought to coordinate the flavin ring through hydrogen bonding to its side chain;[173a] Y312F, given that Y312 is placed adjacent to residue R618 and might be involved in coordination of the substrate in the active site, as previously observed for DAAO;[178] C500A and Y521A, to investigate their potential role in substrate binding in the active site pocket.

As described in Section 5.2.3.1, generation of the new mutants was performed using the Megaprimer PCR-based protocol, where the desired mutation was inserted by using different forward mutagenic primers for each individual mutant, together with a common reverse primer. The mutants for this second set were generated using the pET-30a plasmid construct, mutagenesis was verified by sequencing, and the plasmids were transformed in *E.coli* BL21 for subsequent protein expression and purification.

Activity assays for the newly designed MnmC mutants were performed using purified T1. Reactions were set up and analyzed by HPLC as described in section 5.2.3.1. The resulting chromatograms are shown in Figure 25.



D

	R567A/R618A	R618A	R567A	wt
Activity (%)	~ 0	10	16	100

Figure 25 Activity assay of new set of MnmC mutants. Mutants Y312F, C500A, Y504F, Y521A, R567A/R618A were tested for conversion of T1 to T2. Previously generated mutants R567A, R618A and S617A (see Section 5.2.3.1) were also used in this assay for comparison. Conversion of T1 to T2 was monitored by ion-exchange chromatography. A) Comparison of activity of wild type MnmC and mutants Y312F, C500A, Y521A, S617A and Y504F. In all cases, mutant proteins display the same activity as the wild type. B) Comparison of wild type MnmC and mutants R567A, R618A and the double mutant R567A/R618A. The double mutation leads to complete inactivation of the FAD-dependent reaction. C) Control chromatograms with unreacted T1, T2 and coinjection of the two species for comparison of the characteristic retention times. D) Summary of the relative activities of wild type and mutant MnmC analyzed in Panel B.

As evident from Figure 25, Panel A, the majority of the newly designed mutants (Y312F, C500A, Y504F, Y521A) displayed the same chromatographic profile as the wild type enzyme, suggesting that conversion of T1 (retention time about 11 min) to T2 (retention time 10.5 min) is not affected by the mutation (note that S617A from the previous set of mutants was also tested in the same assay for comparison). On the contrary, comparison of the double mutant R567A/R618A with the wild type enzyme and with the single mutations R567A and R618A, shown in panel B, revealed that removal of both arginines in the active site leads to complete inactivation of the deacetylation reaction (a comparison of the relative activity of mutants shown in Figure 25B is summarized in Panel D). Panel C displays the control samples containing unreacted T1, T2 as well as the co-injection, for comparison of the retention times. We deduce that the key players in the catalytic conversion of $cmnm^5s^2U$ to nm^5s^2U are R567 and R618, where R567 is suggested to coordinate the carboxylate group in the $cmnm^5$-modification and R618 is thought to be involved in binding of the target nucleobase.[173a] The complete inactivation of the FAD-dependent reaction in R567A/R618A makes this mutant suitable for further crystallization studies aiming at the investigation of MnmC in complex with T1.

5.2.4 Circular dichroism studies of MnmC mutants

In order to verify that the MnmC mutants display the same conformation as the wild type enzyme, circular dichroism studies were performed to compare the secondary and tertiary structures of the purified proteins. Circular dichroism is defined as the differential absorption of left- and right-handed circularly polarized light, and it is an excellent tool for studying the conformation and folding of proteins. More specifically, the far UV absorption (240 nm and below) arises from the peptide bond and provides information concerning the secondary structure, while spectra in the near UV region are mostly resulting from aromatic amino acids and organic cofactors (such as e.g. flavins) and inform about the tertiary structure of the protein under investigation. Wild type MnmC, as well as R567A, R618A and R567A/R618A were diluted to 1.6 µM (far UV) and 20 µM (near UV) in MnmC CD buffer and far and near spectra were measured using a 1 mm or 1 cm pathlength, respectively. The resulting spectra are shown in Figure 26.

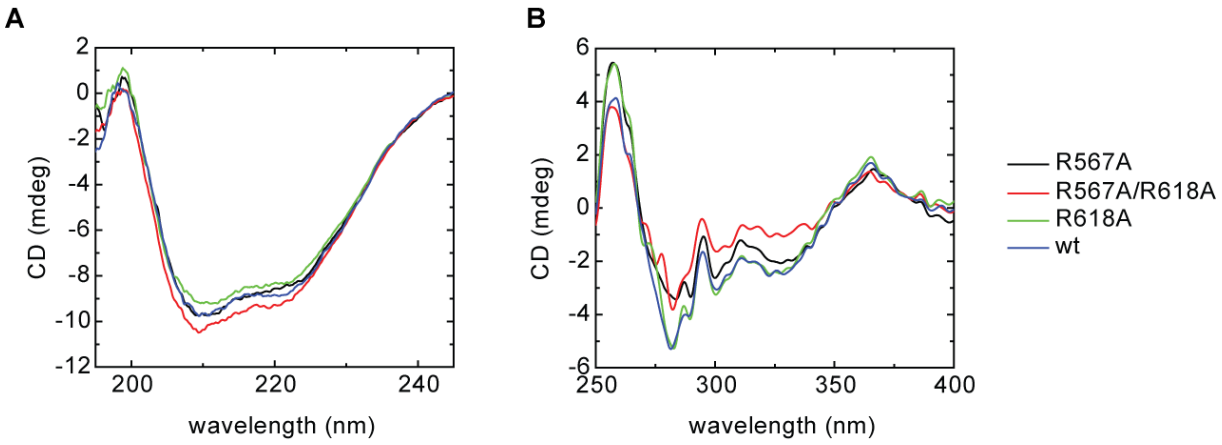


Figure 26 CD spectra of wild type and mutant MnmC. A) Far and B) near UV spectra of the mutant and wild type proteins show that there is no change in the secondary and tertiary structures upon mutation of the arginine residues in the active site.

From Figure 26 it is evident that both the far (Panel A) and near (Panel B) UV spectra, corresponding to secondary and tertiary structure, respectively, display largely similar features in the wild type and mutant proteins, suggesting that loss of function in the arginine mutants is not due to unfolding of the MnmC C-terminal domain. Furthermore, the similar conformational features of the double mutant MnmC compared to the wild type confirm that this mutant is also a suitable candidate for structural characterization of the MnmC-T1 complex.

5.2.5 Spectroscopic properties of MnmC and comparison with mutants

The absorption spectrum of MnmC shows the characteristic features of a flavoprotein (Figure 27), with maximum absorption at 357 nm and 430 nm corresponding to the flavin cofactor, as originally described by Bujnicki *et al.*, who also identified the flavin derivative as FAD.[171b] Spectra of mutant MnmC proteins displayed similar features, suggesting that mutations did not lead to loss of the cofactor.

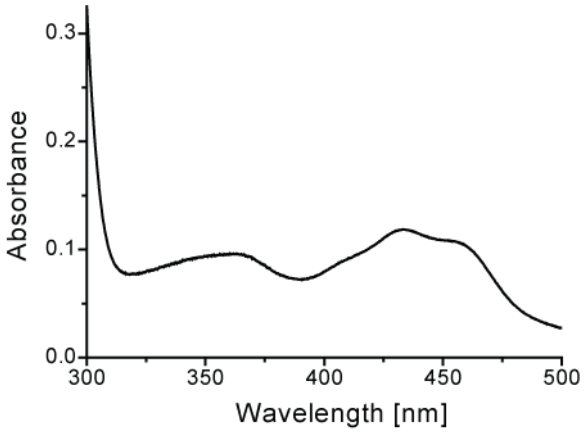


Figure 27 Absorption spectrum of wild type MnmC.

5.2.6 Binding studies with T1 and T2

In order to further characterize the protein-tRNA complex as well as to optimize the crystallization conditions, binding studies were performed with MnmC mutant and wild type proteins in complex with T1 and T2, respectively. Complexation was monitored using gel electrophoresis mobility shift assays (EMSA), where the free and bound tRNA could be monitored by SYBR Green II staining, and optimization of the experimental conditions was performed together with Dr. David Pearson. In order to derive the dissociation constant K_d, the tRNA concentration was optimized to ensure that free protein, free tRNA and complex would be present at equilibrium. As shown in Figure 28, different protein concentrations were incubated with a constant tRNA amount. In the case of MnmC R567A/R618A protein (Figure 28A) concentrations ranged from 0.1 µM to 20 µM with T1 set constant at 0.5 µM, while 1 µM T2 (Figure 28B) was incubated with wild type MnmC ranging from 0.37 µM to 15.6 µM.

In both cases, the dissociation constant was calculated by nonlinear least-squares fit to Equation 1 (see Section 5.4.8)[179] which yields values for K_d assuming that there is a degree of receptor depletion. The K_d (T1) and the K_d (T2) are respectively 1.35 µM and 1.25 µM (see Figure 28). This is to some extent surprising given that previous kinetic studies suggest that the affinity of MnmC for T2 should be about ten-fold higher than for T1.[172] However, the different binding affinity observed by gel shift assay might be due to the fact that complexation could be affected by the gel matrix, therefore disfavouring protein-RNA binding at lower concentrations and yielding an underestimation of the binding affinity for T2 (expected to be around 100 nM).

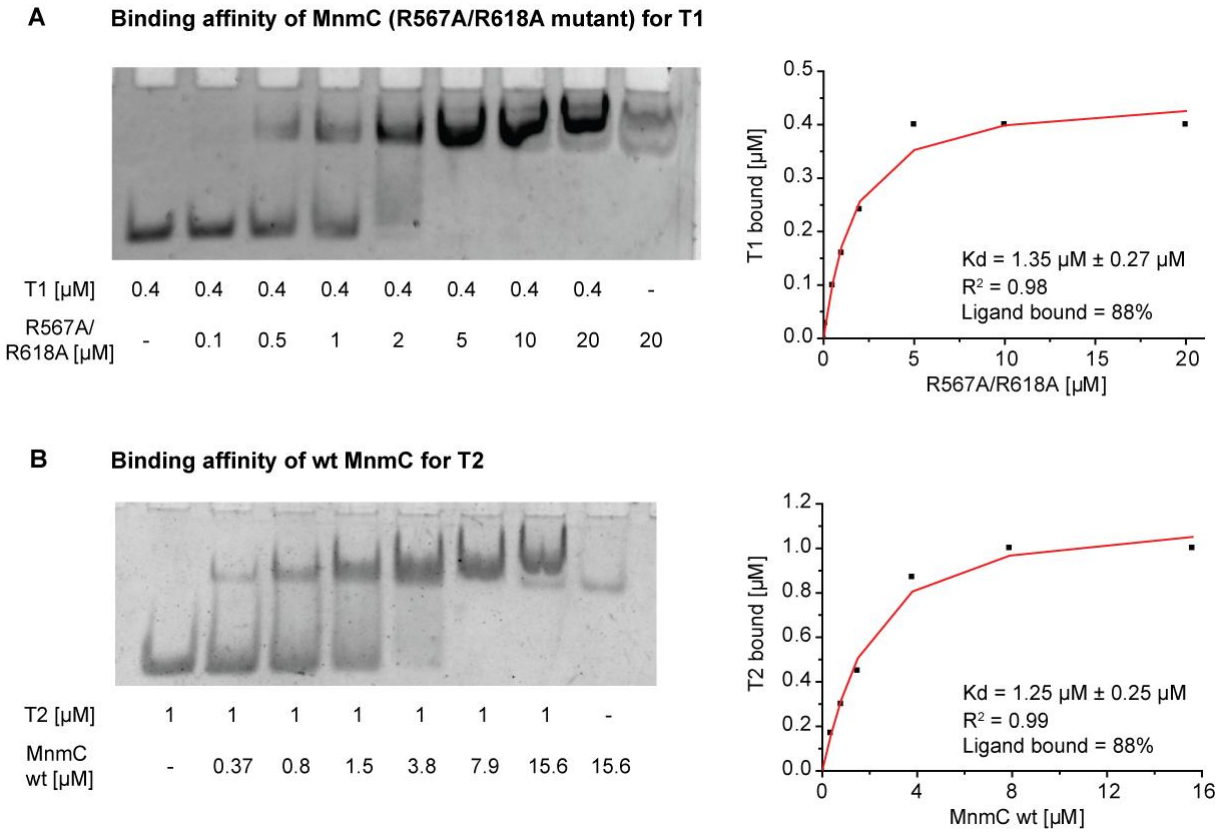


Figure 28 Binding studies of R567A/R618A and wild type MnmC with T1 and T2, respectively. A) EMSA binding study for R567A/R618A bound to T1 and corresponding binding isotherm plot used to calculate the K_d(T1). B) EMSA binding for wild type MnmC bound to T2 and corresponding binding isotherm plot.

Most importantly, these binding studies confirm that incubation of MnmC with substrate tRNA at concentrations at least 10-fold above the obtained K_d result in full complexation, a feature which is essential for crystallization, and that could be used for later optimization of the crystallization conditions.

Binding studies for MnmC in complex with T2 were also performed in the presence of Aza-SAM to investigate whether the AdoMet homologue could affect the binding affinity, but no significant effect was observed.

Furthermore, in order to optimize the crystallization conditions, a stoichiometric titration of wild type MnmC with T2 was performed in order to determine the stoichiometry of the protein-tRNA complex. In fact, T2 should be bound at the N-terminus (where it is the preferred substrate) but could also bind at the C-terminus (where it is the product of the C-N bond cleavage reaction). For this purpose, T2 at concentrations at least 10-fold above the estimated K_d (in this case [T2] = 20 µM) was incubated with increasing concentrations of MnmC (up to 4-fold the tRNA concentration, 0–80 µM). The expected pattern is a linear increase in the complex signal with increasing protein concentration, which reaches a plateau in correspondence of the protein equivalents yielding full tRNA complexation. However, as shown in Figure 29, stoichiometric curves present a few problems for the MnmC-tRNA complex: first of all, the large amounts of tRNA required for stoichiometric titrations result in signal cancellation

effects which make it impossible to quantify the tRNA vs complex signal intensity; second, the MnmC intrinsic fluorescence further distorts the signal intensity for the MnmC-tRNA complex; third, at these concentrations tRNA binding seems to be strongly affected by the gel matrix, possibly because of the small sample volumes used and the subsequently larger dilution effects resulting upon loading onto the gel (which could explain the observed smearing). However, from the stoichiometric binding curve presented in Figure 29 we can at least exclude a 2:1 tRNA to protein complex. More accurate evaluation of the stoichiometry of the complex could be investigated in the future by alternative methods such as, for example, capillary electrophoresis.

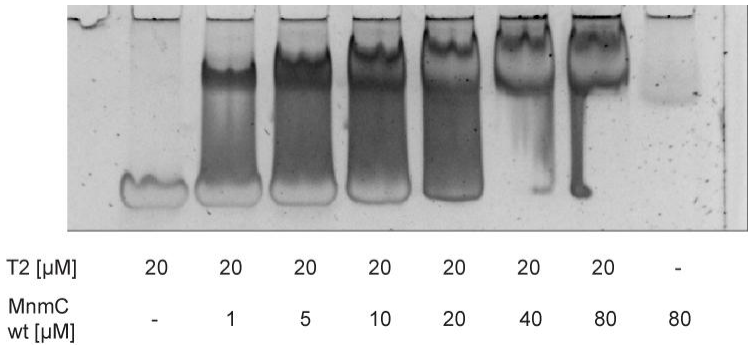


Figure 29 Stoichiometric binding study with wild type MnmC and T2. Note the distorted fluorescence signal arising from large tRNA amounts. Smearing of the signal is most likely due to dilution effects upon loading of the sample on the gel.

5.2.7 Preliminary crystallization screening of the MnmC-T2 complex

One of the principal aims of the project was to investigate the structure-function relationships for the MnmC-tRNA complex to elucidate the specific features which ensure selective modification at the U34 position of *E.coli* tRNAs Glu, Lys and Arg. The structure of MnmC from *E.coli* was recently solved by two groups (PDB 3awi,[173a] PDB 3ps9[173b]), but no crystal structure of the protein-tRNA complex is available yet.

Given that the conversion reaction from T2 to T3 could easily be trapped in the absence of cofactor SAM, the majority of the crystallization tests were performed using wild type MnmC in complex with T2. Preliminary stability tests were performed incubating MnmC in crystallization buffer (see Section 5.4.3) at 4 °C and 18 °C over three days and samples were analyzed by SDS-PAGE. In both cases, no protein degradation was observed. Similarly, MnmC was incubated at 4 °C and at 18 °C over three days and its tertiary structure evaluated by circular dichroism to verify whether the protein was unfolding during prolonged incubation in crystallization buffer. In both cases the tertiary structure was comparable to a control sample, suggesting that no protein unfolding was taking place.

Crystallization screening was performed under a wide variety of conditions. High throughput screening using sitting drop was performed at the Max Planck Institute for Biochemistry (Conti Lab, Martinsried) as well as at the HT Crystallization Facility at the University of Zürich. Commercially available kits from *Qiagen* and *Hampton* were used, as well as in-house designed crystallization

screenings. In addition, manual screening were performed in a 24-well hanging drop-format using crystallization conditions previously published for MnmC (PDB 3ps9, conditions taken from PDB prior to publication).[173b]

Crystallization drops were generally set up using a 1:1 mixture of complex to reservoir, with a 1:1.2 molar ratio of MnmC:T2 in crystallization buffer supplemented with 16 µM FAD and, occasionally, 100 µM Aza-SAM. The complex was set up on ice prior to HT crystallization, and screening was performed both at 4 °C and at 18 °C. No crystals could be obtained so far.

5.3 Conclusion

Investigation focused on the two sequential biosynthetic steps that yield the modified mnm^{5}s^{2}U34 in *E.coli* tRNA^{Glu}. Mutations of MnmC residues located in the FAD active site revealed that the main residues involved in catalysis are the two arginines R567 and R618, responsible for coordination of the cmnm^5-moiety and binding of the substrate nucleobase,[175a] while other residues might only play a minor role in coordination of the anticodon loop. Binding of the wild type and mutant MnmC to substrate tRNA species T1 and T2 lies in the low µM range, largely in agreement with previous kinetic studies,[172] and the protein is likely to bind the tRNA substrates in a 1:1 complex. Furthermore, the fully inactive mutant R567A/R618A is a suitable candidate for crystallization screening in complex with substrate tRNA T1, given that it retains a similar secondary and tertiary structure to that of wild type MnmC as well as binding of the FAD cofactor. Further optimization of the crystallization conditions is required for both protein-tRNA complexes, and will provide insight into the selective recognition and modification of tRNA^{Glu} at the wobble position by MnmC.

5.4 Materials and methods

5.4.1 Cloning of pET-28a and pET-30a MnmC

Cloning of wild type *E.coli* MnmC in pET-28a (*Novagen*) was performed by Dr. David Pearson according to Bujnicki *et al*..[171b] Cloning of wild type MnmC in pET-30a (*Novagen*) was performed by PCR amplification of the MnmC gene from the pET-28a construct using the following primers:

5´-GGATTAGCGAGGTACCGACGACGACGACAAGATGAAACACTACTCCATACAACCTGCC AAC- 3´

5´-GCTACGGTCTGAATTCTCATTACCCCGCCTTAACCGCTTTACCCTTCAA- 3´

This primer pair yields a PCR product containing the KpnI and EcoRI restriction sites for cloning into pET-30a, as well as a newly designed Enterokinase cleavage site that, after incubation with the protease, results in complete removal of non-MnmC residues at the N-terminus of the purified protein. The PCR amplified gene was cloned in pET-30a between the KpnI and EcoRI restriction sites and correct amplification and cloning was verified by DNA sequencing. A similar cloning protocol was

also used to generate R567A and R618A pET-30a constructs using mutant pET-28a constructs (generated as detailed in the next section) as templates.

5.4.2 Mutagenesis of pET-28a and pET-30a

Mutations of the pET-28a MnmC construct were generated by initial megaprimer generation (introducing the desired mutation), followed by full plasmid amplification.[177] Mutants R567A, S617A and R618A were generated from the wild-type MnmC-pET28a plasmid using the respective primers listed in Table 11, in combination with MnmCmid primer, yielding megaprimers of about 1200 bp (for R567A) and 1350 bp (for S617A and R618A).

Primer	Template plasmid	Sequence
MnmC R567A – rev	pET-28a MnmC (wt)	GATCGCGGGTGGCACAAGCCACACCGCAGCGCGCC (Roovers 2008)
MnmC R618A – rev	pET-28a MnmC (wt)	GGCAGAACACAAACCGGCAGAACCTAAAGCCGC (Roovers 2008)
MnmC S617A – rev	pET-28a MnmC (wt)	CAGAACACAAACCGCGAGCACCTAAAGCCGCAAAC
MnmCmid		AGTAGATGCCTGGTTTCTGGACGG

Table 11 Forward non-mutagenic (MnmCmid) and reverse primers carrying the R618A, R567A and S617A mutations used for mutagenesis of pET-28a MnmC.

Mutagenesis by PCR amplification was performed using the *Agilent Quick-change Lightning Site-directed Mutagenesis Kit* (*Agilent*) with the PCR program detailed in Table 12, followed by plasmid digestion using DpnI (*New England Biolabs*) and transformation in *XL-10 Gold Ultracompetent cells* (*Stratagene*). Correct mutagenesis was verified by DNA sequencing.

Temperature (°C)	Time (min)	Number of cycles
95	3	
95	0.5	3 x
55	1	
72	1	
95	0.5	20 x
68	8	
68	15	
4	hold	

Table 12 PCR program cycles for megaprimer-based mutagenesis.

A second set of mutants was generated by a similar protocol as detailed in the previous paragraph. Individual primers containing the desired mutation (Y312F, C500A, Y504F, Y521A, R567A, see Table 13) were used in combination with MnmCmid to generate megaprimers of pET-28a MnmC constructs (wild type or R618A) to yield single mutants and the double mutants R567A/R618A pET-28a construct. Correct mutagenesis was verified by sequencing. The generated pET-28a mutant constructs were subsequently digested with HindIII and NcoI restriction endonucleases and cloned in pET-30a MnmC constructs between the same sites, therefore yielding mutated pET-30a constructs Y312F, C500A, Y504F, Y521A and R567A/R618A.

Primer	**Template plasmid**	**Sequence**
MnmC Y312F-rev	pET-28a MnmC (wt)	GTTTGCTTAATAACGGAAACAGCGCCCCCTGGCG
MnmC C500A – rev	pET-28a MnmC (wt)	GAGATAACCGTCATAGGCCAGCACCTGCTTCAGC
MnmC Y504F – rev	pET-28a MnmC (wt)	CGGATTTTGTGGCGTGAGAAAACCGTCATAGCACAG
MnmC Y521A – rev	pET-28a MnmC (wt)	CGCTGCCGCGATGAGCACTGGCACCAATACAATG
MnmC R567A – rev	pET-28a MnmC R618A	GATCGCGGGTGGCACAAGCCACACCGCAGCGCGCC (Roovers 2008)
MnmCmid primer		AGTAGATGCCTGGTTTCTGGACGG

Table 13 Forward non-mutagenic (MnmCmid) and reverse primers carrying the C500A, Y504, Y312F, Y521 and R567A mutations used for mutagenesis of pET-28a MnmC. Plasmids used as template for mutagenesis are listed in the section "Plasmid".

5.4.3 Protein expression and purification

The protocol for protein expression and purification was adapted from Bujnicki *et al*..[171b] Plasmids containing wild type or mutant MnmC were transformed in *E.coli* BL21 cells (*New England Biolabs*). Cells were grown in LB medium supplemented with 15 µg/mL kanamycin, at 37 °C, until a cell density OD_{600} = 0.7 was reached. Protein expression was induced by addition of IPTG (1 mM final concentration) and cells were grown for further 3 h at 37 °C and finally harvested.

Cells were resuspended in cold MnmC Buffer A (50 mM Tris pH 8.0, 10 % glycerol, 10 mM $MgCl_2$, 200 mM KCl, 5 mM imidazole, 3 mM β-mercaptoethanol) supplemented with a cocktail of protease inhibitors (*Complete EDTA-free, Roche*) and lysed using a French press. The lysate was cleared by centrifugation using a *Sorvall SS-34 rotor* (18000 rpm, 4 °C, 30 min) and applied onto a *HiScreen IMAC FF* column (*GE Healthcare*) at 4 °C. The protein was eluted by applying a gradient to MnmC

Buffer B (MnmC Buffer A + 0.5 M imidazole) and the eluted fractions pooled, concentrated using an *Amicon Ultra centrifuge filter* (*Millipore*, 30,000 MWCO), and buffer-exchanged to MnmC Buffer C (50 mM Tris pH 8.0, 10 % glycerol, 10 mM $MgCl_2$, 3 mM β-mercaptoethanol). The protein sample was then loaded onto a 1 mL *MonoQ* ion-exchange column (*GE Healthcare*) equilibrated with MnmC Buffer C. Protein was eluted by applying a linear gradient to MnmC Buffer D (MnmC Buffer C + 0.5 M KCl), concentrated using an *Amicon Ultra centrifuge filter* (*Millipore*) and buffer exchanged to MnmC Buffer E (MnmC Buffer C + 50 mM KCl). Tag-cleavage of pET-30a constructs was performed by incubating the sample in MnmC Buffer E supplemented with 7.5 mM $CaCl_2$ in the presence of Enterokinase light chain protease (*New England Biolabs*). The sample was incubated overnight at 4 °C, and complete cleavage was monitored by SDS-PAGE. Finally, the sample was applied onto a 1 mL *HiTrap Heparin HP* (*GE Healthcare*) and eluted by applying a gradient to MnmC Buffer D. Purified protein was stored in crystallization buffer MnmC Buffer F (50 mM Tris pH 8.0, 10 % glycerol, 1 mM $MgCl_2$, 100 mM KCl, 3 mM β-mercaptoethanol) at -80 °C.

5.4.4 tRNA expression and purification

Plasmid $tRNA^{Glu}$-pSGAT2 was transformed in ΔMnmC T7 express *E.coli*.[172] Transformed cells were grown in LB medium (with 100 µg/mL carbenicillin) at 37 °C until an OD_{600} of 0.5. tRNA expression was induced with 1 mM IPTG for 6 h at 30 °C. Subsequent tRNA extraction steps were performed at 4 °C as described previously.[142] Harvested cells were resuspended in 15 mL of tRNA extraction buffer (0.01 M $Mg(OAc)_2$, 0.05 M NaOAc, 0.15 M NaCl, pH 4.5) and mixed with an equal volume of 80 % aq. phenol. The suspension was shaken vigorously for 1 h at 4 °C and centrifuged for 30 min at 3220 g. The aqueous layer was once again extracted with 15 mL of 80 % aq. phenol, followed by centrifugation (20 min, 3,220 g). The aqueous phase was extracted twice with chloroform (5 mL) and separated by centrifugation (10 min, 3,220 g). DNA and long RNAs were precipitated by addition to the aqueous phase of 0.1 vol of 20 % KOAc, pH 4.5 and LiCl to a final 2.0 M concentration, followed by incubation for 4 h at 4 °C and centrifugation (20 min, 38,724 g). The supernatant was combined with 3 vol of absolute EtOH and incubated overnight at -20 °C. The tRNA was pelleted by centrifugation (60 min, 24,336 g) and resuspended in tRNA Buffer A (100 mM Tris pH 8.0, 50 mM $MgCl_2$). Subsequently, tRNA samples were subjected to anion exchange HPLC purification using a *DNAPac PA100 22 x 250 mm* column (*Dionex*) on a *Merck Hitachi Lachrom* system. Purification of $cmnm^5s^2U$-tRNA (T1) was performed once with tRNA Buffer B (100 mM NaOAc, pH 5.0, 50 mM $MgCl_2$), followed by purification in tRNA Buffer A, in both cases eluting the tRNA samples by applying a gradient to 500 mM NaCl (T1 and T2 eluted at about 200–250 mM NaCl). Purity and correct isolation of T1 was verified by incubation with purified MnmC and analysis by analytical HPLC on a *DNAPac PA100 4 x 250 mm* column (*Dionex*). Isolation of T2 required prior incubation of purified T1 with MnmC (1 h, 37 °C) in MnmC Reaction Buffer (50 mM Tris pH 8.0, 20 mM NH_4Cl), followed by purification by anion exchange chromatography in tRNA Buffer A (elution with gradient to 500 mM NaCl). After each purification step, tRNA fractions were pooled, concentrated and buffer

exchanged using an *Amicon Ultra centrifugal filter* (10,000 MWCO, *Millipore*). Purified T1 and T2 were stored in tRNA Buffer A at -80 °C.

5.4.5 Mutant activity assay

Mutant MnmC proteins were incubated either with crude *E.coli* tRNA containing undermodified $tRNA^{Glu}$ or with purified T1 as detailed in sections 5.2.3.1 and 5.2.3.2. Assays were performed incubating wild type or mutant MnmC with substrate tRNA in MnmC Reaction buffer at 37 °C. For initial screening of R567A, S617A and R618A mutant activity, 20 µg of protein were incubated with 20 µg of crude tRNA containing undermodified T1 (after ethanol precipitation, resuspended in tRNA Buffer A, see section 5.2.2) in a total 10 µL reaction volume, and incubated at 37 °C for 30 minutes. In the case of conversion of T2 to T3, the reaction with T2-containing crude *E.coli* tRNA was supplemented with 500 µM SAM. Activity tests of mutants described in section 5.2.3.2 were set up with 130 nM protein and 130 nM T1 in 60 µL total reaction volume. Reaction was incubated at 37 °C for 30 min. Reaction samples were quenched by addition of tRNA Buffer B (at least one vol) and analyzed on a *DNAPac PA100 4 x 250 mm* column (*Dionex*) using a *Merck Hitachi Lachrom* system. Elution was achieved by applying a gradient of 175–225 mM NaCl. Reaction progress was estimated by analysis of the peaks corresponding to T1 and T2, each displaying a characteristic retention time.

5.4.6 Circular dichroism

Circular dichroism studies were performed using a *Jasco Spectropolarimeter J-810*. Protein samples were diluted in MnmC CD Buffer (20 mM Tris pH 8.0, 1 mM $MgCl_2$, 5 % glycerol) to the concentrations detailed in Section 5.2.4, and spectra were recorded in triplica at 20 nm/min at 20 °C, using 1 mm and 1 cm pathlength for far and near UV, respectively.

5.4.7 UV-Vis Spectroscopy

Spectra were measured using a *Jasco V-650 Spectrophotometer*. Protein samples were diluted in MnmC UV Buffer (50 mM Tris pH 8.0, 10 mM $MgCl_2$) at a concentration of 2 mg/mL. Spectra were measured in the range of 300–500 nm using a 1 cm pathlength, at 20 °C.

5.4.8 Electrophoretic mobility shift assay for the protein-tRNA complex

Wild type and mutant MnmC protein stocks were prepared by serial dilutions in Binding buffer (50 mM Tris pH 8.0, 1 mM $MgCl_2$, 20 mM NH_4Cl, 1 mM β-mercaptoethanol, 50 mM KCl, 10 % glycerol) and incubated with the appropriate amount of T1 or T2 in a final 10 µL (for equilibrium binding curves) and 4 µL (for stoichiometric curves). The complex was allowed to reach equilibrium for 1 hour at 4 °C and separated through a 8 % native polyacrylamide gel (8 % polyacrylamide, 2.5 % glycerol, 0.5 x TAE buffer) for 50 min at constant 100 V in 0.25 x TAE buffer at 4 °C. A 5 µL loading dye sample was loaded separately to monitor progress. The gel was subsequently incubated in 0.25 x TAE buffer containing *SYBR Green II* (*Sigma*, 1:10,000 dilution) for 10 min. The tRNA and tRNA-MnmC bands were visualized using a *Fuji LAS 3000* fluorescence imager, and band quantification was performed using *AIDA* software, and fitting to Equation 1[179] was performed using

ORIGIN®. Samples containing T1 in complex with MnmC R567A/R618A were prepared in duplica, one aliquot loaded on a native polyacrylamide gel as detailed above, and one analyzed by analytical HPLC to confirm the absence of conversion of T1 to T2.

Equation 1[179]

$$B = \frac{(L_{tot}+P_{tot}+K_{dis})-\sqrt{(L_{tot}+P_{tot}+K_{dis})^2-(4\times L_{tot}\times P_{tot})}}{2}$$

where L_{tot} = total ligand concentration (µM)
P_{tot} = total protein concentration (µM)
K_{dis} = dissociation constant
B = ligand bound (µM)

5.4.9 Protein crystallization screening

For co-crystallization, the MnmC-T2 complex was set up in a 1:1.2 molar ratio in crystallization buffer (50 mM Tris pH 8.0, 10 % glycerol, 1 mM $MgCl_2$, 3 mM β-mercaptoethanol, 20 mM NH_4Cl, 100 mM KCl) supplemented with 16 µM FAD and, depending on the screening, 100 µM Aza-SAM. Protein:tRNA concentrations were screened ranging from 10:4 mg/mL, to 30:12 mg/mL. The complex was set up on ice with freshly purified protein.

High throughput screening was performed using a *Phoenix nanodispenser robot* (*Art Robbins Instruments*, drop size: 100 nL complex + 100 nL precipitant) in the research group of E. Conti (Max Planck Institute, Martinsried). The following screening kits were used: *Qiagen JCSG+ kit*, *Qiagen MPDs kit*, *Qiagen PEGs kit*, *Hampton research Index kit*, all commercially available, as well as additional screening kits designed by the Conti group (*AJ1*, *AJ2*, *CP-PEGS-Salt screen* kits). Plates were incubated either at 4 °C or at 18 °C.

An additional robot screening was performed at the HT Crystallization facility at the University of Zürich using in-house designed crystallization screening kits. Plates were incubated at 20 °C.

Finally, manual screenings were performed using a 24-well hanging drop-format with 1.2–2.2 M tri-ammonium citrate, pH 7.0, 0.1–0.8 % ethyl acetate, at 20 °C (conditions used for PDB 3ps9).[173b]

In all cases, no crystals could be obtained. In the best cases, drops showed fine crystalline precipitate or small spherulites.

6 Investigation of putative pathways of cytosine demethylation involving C-C bond cleavage or replication-coupled dilution

6.1 Introduction

In eukaryotes, methylcytosine (mC) is thought to be a rather stable DNA modification and it plays an important role in a variety of processes such as modulation of gene expression, genomic imprinting and X-chromosome inactivation.[180] However, methylation has also been observed to have a dynamic nature: on the one hand, *de novo* methylation taking place during development plays a crucial role in loss of pluripotency and in cellular specification;[181] on the other hand, methylcytosine has also been reported to undergo depletion in both genome-wide and gene-specific demethylation events,[122b] where loss of the methylation pattern is thought to take place *via* active or passive mechanisms.[122b] For example, a rapid wave of genome-wide demethylation takes place in the male pronucleus after fertilization,[182] resulting from genome-wide oxidation of mC,[114] while gene-specific demethylation has been mainly observed in somatic cells,[122b] such as activated T lymphocytes[183] and neurons.[127b]

While passive demethylation, where the methylation pattern is lost by dilution during replication, is broadly accepted, active demethylation remains a subject of controversy.[122] Nevertheless, discovery of the oxidized cytosine derivatives 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxycytosine (caC) has led to increasing interest in these modified nucleosides as possible intermediates in the demethylation process.[98b, 100] Up to date, two main pathways of active demethylation have been proposed (see Figure 30). Demethylation might involve DNA glycosylases coupled to base excision repair (BER) while an alternative mechanism postulates direct C-C bond cleavage at the C5 position, yielding an unmodified cytosine without excision of the nucleobase.

Two different BER-coupled demethylation mechanisms have been reported. On the one hand, thymine DNA glycosylase (TDG) was shown to efficiently excise fC and caC,[100a] and these results were confirmed by various groups.[126, 184] The second pathway is suggested to involve the activation-induced cytidine deaminase (AID)/apolipo-protein B mRNA editing enzyme, catalytic polypeptide (APOBEC) family of enzymes.[127b, 128] These are thought to deaminate mC and hmC to thymine and 5-hydroxymethyluracil (hmU), respectively, followed by removal by the glycosylases TDG or single-strand-selective monofunctional uracil DNA glycosylase 1 (SMUG1) and by processing *via* the BER machinery.[125, 127b] However the involvement of deamination in the process of active demethylation remains controversial.[129] In order to elucidate the nature of hmU generation and its relation to deamination of hmU, a more detailed study was undertaken in the course of this thesis and is discussed in Chapter 8.

The second mechanism of active demethylation is thought to involve oxidation of methylcytosine to hmC, fC and ultimately caC, followed by decarboxylation via C-C bond cleavage (Figure 30),[132] or to proceed directly from hmC or fC via a retro-Aldol-type chemistry. A similar decarboxylation reaction is known to occur in the thymidine salvage pathway, where a iso-orotate decarboxylase converts iso-orotate (previously generated by a thymine 7-hydroxylase) to uracil,[122b, 131] but no such protein has

yet been found to perform this reaction on cytosine derivatives. In a report published by Schiesser *et al.*, incubation of [$^{15}N_2$]-caC-containing strands in mESC cell lysate followed by DNA re-isolation and HPLC-ESI-MS analysis revealed the presence of [$^{15}N_2$]-cytosine, suggesting that decarboxylation via C-C bond cleavage had taken place.[132] However, there is yet no evidence available as to which protein or complex is responsible for this putative pathway of active demethylation.

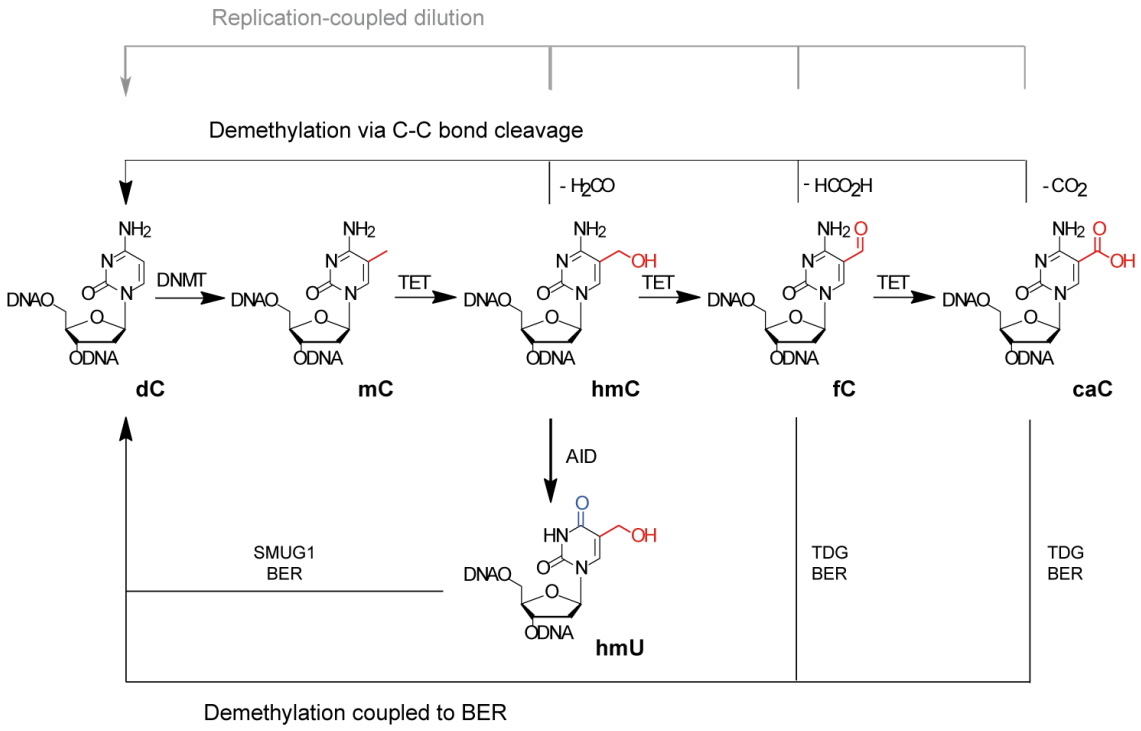


Figure 30 Pathways of active demethylation. Demethylation via C-C bond cleavage proceeds via elimination of formaldehyde from hmC, formic acid from fC and carbon dioxide from caC. An alternative mechanism involves removal of fC, caC and hmU (after deamination) by glycosylases TDG and SMUG1, followed by base excision repair (BER). Modifications at the C5 position are highlighted in red, while deamination is shown in blue.

6.2 Results and discussion

The investigation reported in this chapter aimed at the elucidation of pathways of active demethylation via C-C bond cleavage. Studies were performed using selectively labelled substrates incubated with nuclear extracts or with recombinant proteins. Additionally, recombinant mouse Dnmt1 was purified and used for *in vitro* assays aimed at assessing the interaction and enzymatic activity of this methyltransferase in the presence of oxidized cytosine derivatives.

6.2.1 Substrates

In order to investigate the possible pathways of active demethylation via direct C-C bond cleavage (Figure 30), selectively labelled substrates were designed to enable precise reaction monitoring. Nucleosides 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxycytosine (caC) were synthesized containing a double ^{15}N-labelled nucleobase (see Figure 31). In the case of hmC two additional deuterium atoms were introduced at the 5-hydroxymethyl group. The labelled nucleosides were either generated as phosphoramidite reagents and subsequently incorporated in short

oligonucleotide strands,[132] or, alternatively, they were synthesized as the corresponding triphosphates and used for generation of longer double-stranded DNA segments using a Polymerase Chain Reaction (PCR).[185] For further details concerning the later method, see Chapter 7. In both cases, DNA substrates contained a biotin tag to facilitate isolation of the DNA strands. Synthesis of the labelled triphosphates and of the short, labelled oligonucleotides was performed by Stefan Schiesser and Barbara Steigenberger.

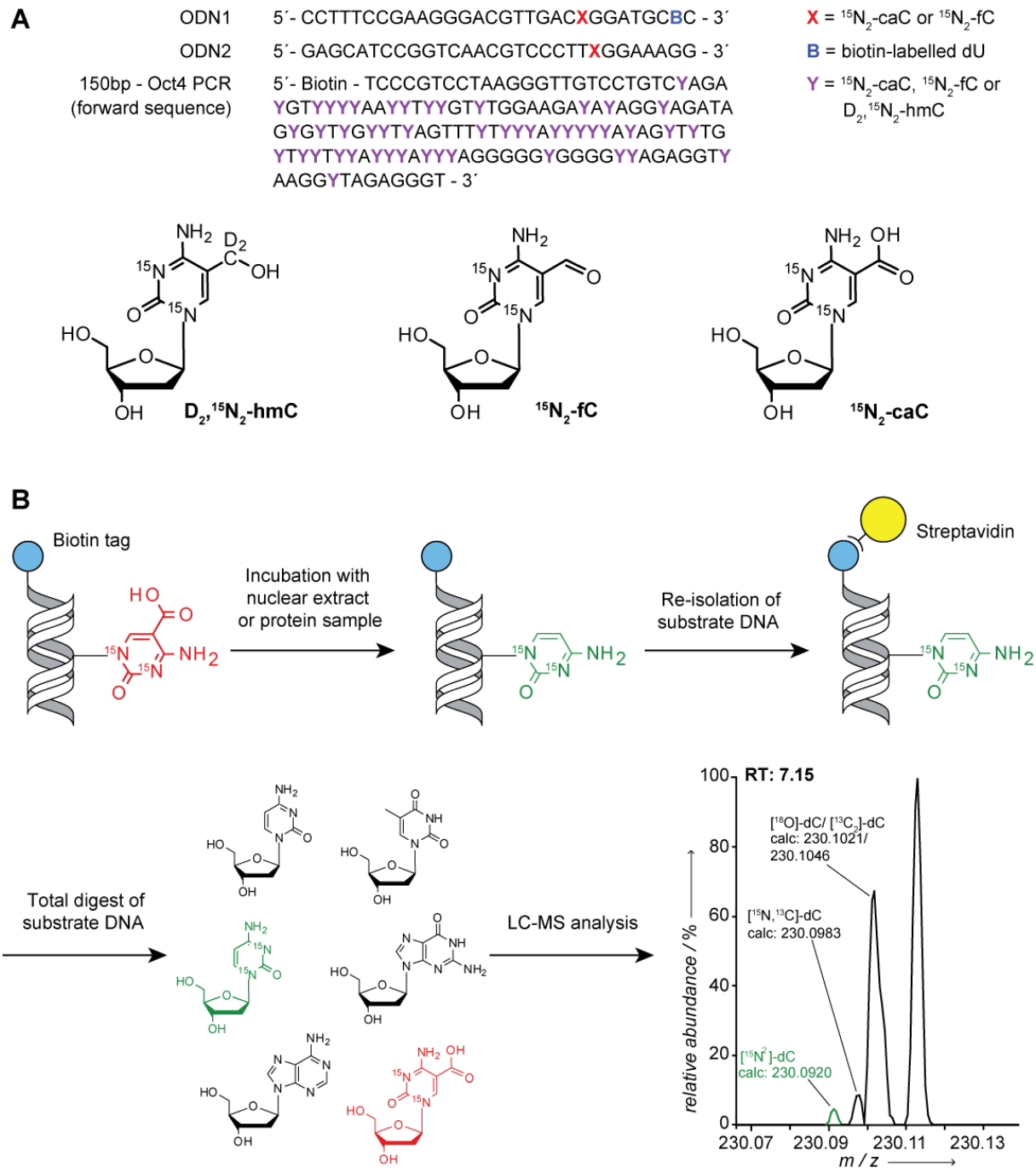


Figure 31 Substrate DNA and schematic representation of experimental set up. A) Sequences for ODN1 and ODN2, both derived from the Ecat1 promoter,[132] alongside the forward sequence of the 150bp – Oct4 PCR fragment.[185] X and Y represent labelled oxidized cytosine derivatives, shown in the lower part of section A). B) Schematic representation of the experimental set up, taking [$^{15}N_2$]-caC-containing oligonucleotides as an example.

The two approaches used for oligonucleotide synthesis present different advantages and disadvantages: solid phase synthesis of labelled oligonucleotides allows for precise sequence control of the site of incorporation of the labelled nucleoside, but it is more labour intensive and, especially in the case of fC, it can be problematic if multiple fC-containing sites are required. On the contrary, synthesis of longer strands by PCR allows for a more rapid generation of the desired substrates containing multiple labelled nucleotides, but there is no control over the site of incorporation.

The sequences adopted for our study were taken from the Ecat1 and Oct4 promoter sequences.[132, 185] The Ecat1 promoter (sequence shown in Figure 31), was chosen because this promoter is known to undergo dehydroxymethylation during stem cell differentiation,[106] while the pluripotency-associated Oct4 promoter is characterized by low levels of methylation in undifferentiated cells, followed by increasing methylation during differentiation.[186]

The labelled substrates were used in *in vitro* experiments by incubation of the labelled DNA with either cell nuclear extracts or specifically enriched protein samples. The DNA was subsequently re-isolated, digested to single nucleosides and analyzed by high resolution HPLC-ESI-MS. Successful detection of [$^{15}N_2$]-labelled cytosine or methylcytosine in the analyte would confirm that the starting material (labelled-hmC, fC or caC) has been converted into cytosine or methylcytosine by direct C-C bond cleavage. Part of the experiments detailed in sections 6.2.2 and 6.2.3 were performed together with Benjamin Hackner.

6.2.2 *In vitro* experiments using cell lysate samples

To further investigate the possible pathways of active demethylation, various experiments were performed using cell nuclear extract samples based on the approach described by Schiesser *et al.*.[132] In particular, this investigation extended the study presented in that publication to further elucidate which of the oxidized cytosine bases (hmC, fC or caC) is the most likely substrate for C-C bond cleavage, as well as to pinpoint specific cell types or conditions which could favour this pathway.

In vitro incubation of labelled oligonucleotides was generally performed with undifferentiated mouse Embryonic Stem Cells (mESC), while HEK-293T cells served as negative control samples. Depending on the experiment, incubation was either performed with nuclear extracts or with intact nuclei, using substrates containing labelled hmC, fC or caC (see Section 6.4.1). Test incubations were also performed with mESC at different stages of differentiation (0, 4, 8 and 16 h) to investigate whether the previously observed peaking and decline of fC and caC during this time period would correlate with an enhanced decarboxylation activity (Section 6.4.1). However, neither [$^{15}N_2$]-C nor [$^{15}N_2$]-mC could be detected in any of the performed experiments, suggesting that there had been no removal of the C5-functionality (hydroxymethyl-, formyl- or carboxyl-group) of the labelled substrate nucleoside. Problems of reproducibility of the decarboxylation reaction in stem cell nuclear extracts had also been previously reported by the author of the original study, Benjamin Hackner,[132] and might be linked

either to a very sensitive decarboxylation protein/complex, or, alternatively, they might suggest that the previously reported decarboxylation is the result of an enzyme-independent reaction.[132]

6.2.3 *In vitro* experiments using protein-enriched samples

Experiments were also carried out using partially purified protein samples. GFP-tagged proteins were expressed in HEK-293T cells, purified on GBP-coated beads,[187] and incubated with substrate DNA as described in section 6.4.2. Incubation experiments focused on the methyltransferases Dnmt1, Dnmt3a and Dnmt3b, and on the Tet proteins.

The Dnmt enzymes were chosen based on the results reported by Schiesser *et al*.[132] which showed that rapid decarboxylation takes place following saturation of the C5-C6 bond of caC. This mechanism mirrors the methylation mechanism adopted by the Dnmt proteins (see Chapter 1) and suggests that the methyltransferases might be putative candidates in this pathway of active demethylation. In support of this hypothesis, Dnmt3a and Dnmt3b were suggested to act as redox-dependent dehydroxymethylases[133] and the same group further reported that all three Dnmts could perform mC demethylation in a Ca^{2+} ion- and redox state-dependent manner, though they could not exclude that these reactions might be linked to impurities in the protein samples.[133]

On the other hand, the Tet proteins were found to oxidize mC to hmC, fC and caC,[98b, 100a, 100b] suggesting that there might be a pathway of active demethylation involving further oxidation of mC, and various studies have reported a Tet1-dependent, gene-specific regulation of the mC levels.[188]

Incubation experiments were performed using all the Dnmt and Tet proteins individually isolated, or Tet1 in combination with the different Dnmts. Experiments investigated reactivity in the presence of labelled hmC, fC and caC. However, neither [$^{15}N_2$]-C nor the (re-)methylated [$^{15}N_2$]-mC could be detected by LC-MS analysis. This suggests that, under the condition tested, Dnmt and Tet proteins are not involved in C-C bond cleavage at the C5 position of the modified cytosine, or that demethylation activity might require some additional proteins or cofactor.

6.2.4 Dnmt1

Given that *in vitro* experiments using protein pull-down samples on beads did not yield any insight into the decarboxylation activity of the Dnmt enzymes, we next proceeded to purification of recombinant mouse Dnmt1 to test whether the purified enzyme could show enhanced activity on our substrate DNA. With the purified methyltransferase, a variety of studies were performed to investigate its interaction with the oxidized cytosine bases 5-formylcytosine and 5-carboxycytosine. Methylation activity and covalent trapping assays were performed together with Benjamin Hackner, while decarboxylation studies were carried out together with Stefan Schiesser.

6.2.4.1 Purification of recombinant Dnmt1

In order to investigate whether Dnmt1 could act as a potential decarboxylase, recombinant mouse Dnmt1 was overexpressed and purified. The mDnmt1(731–1602) construct was kindly provided by

Dr. Jikui Song, and overexpression and purification was performed following the previously published protocol, with only minor variations.[94] Briefly, the plasmid harbouring the mDnmt1 construct was transformed in *E.coli* strain Rosetta 2 (DE3) cells, and protein overexpression was performed as detailed in section 6.4.5. mDnmt1 was initially purified by affinity chromatography on a *HisTrap HP* column, followed by fractionation on a Heparin column and by final purification by size exclusion chromatography (see Figure 32). Fractions with more than 90 % purity were concentrated and stored at -80 °C. Note that in this study, removal of the His_6-SUMO tag by ULP1 cleavage detailed in the original publication was not performed.[94]

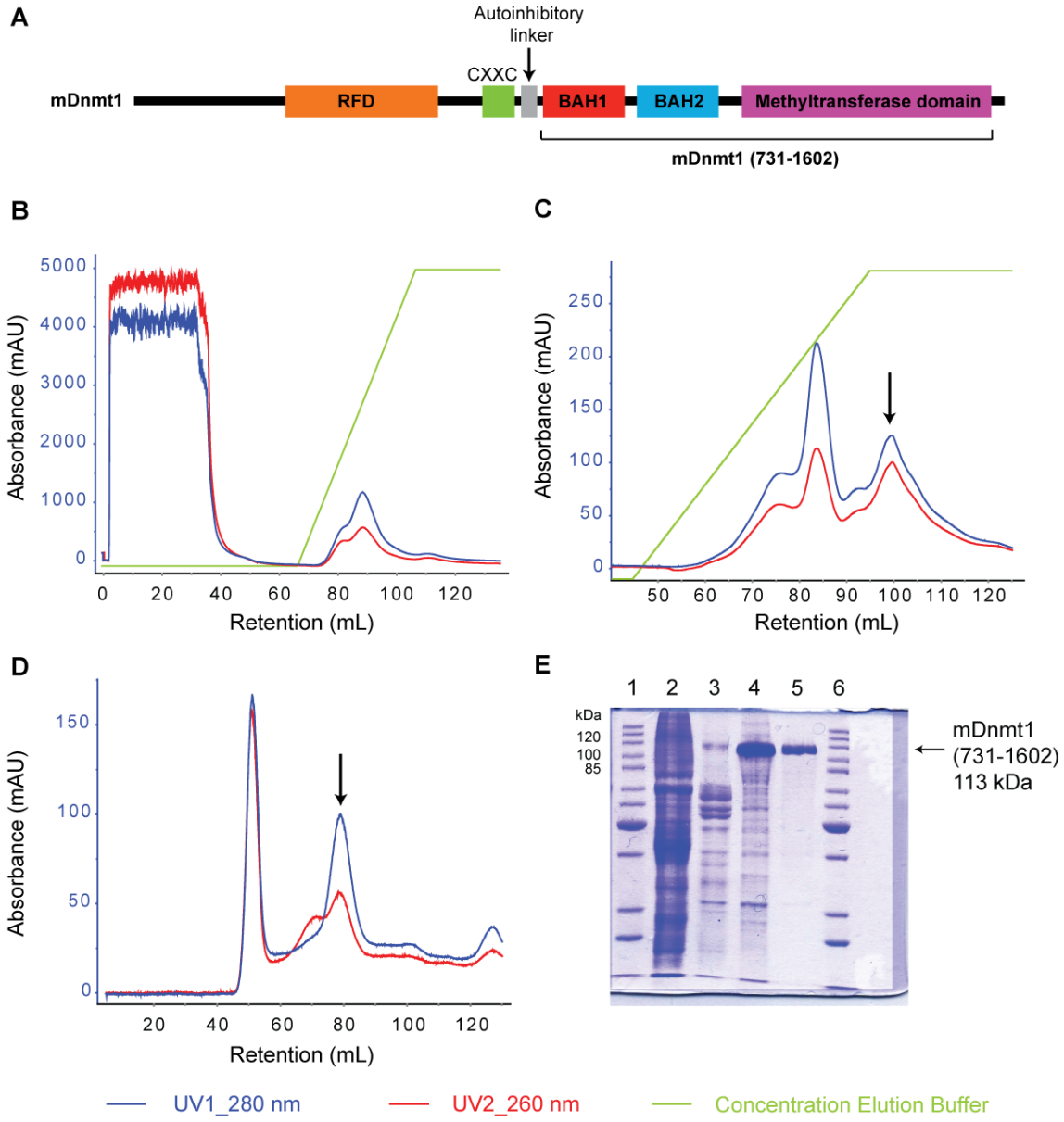


Figure 32 mDnmt1 structural overview and summary of purification of mDnmt1(731–1602). A) Summary of domain structure of mDnmt1 and of mDnmt1(731–1602) fragment. B) Affinity purification of HisTrap column. C) Affinity purification of HiTrap Heparin column. D) Final purification step by size exclusion chromatography. In graphs C) and D) peak corresponding to mDnmt1(731–1602) is indicated with an arrow. Chromatographic purification was monitored by absorbance at 260 nm (red trace) and 280 nm (blue trace). E) SDS-PAGE gel of various purification steps. 1: protein marker; 2: cell lysate; 3: fractions pooled after HisTrap column; 4: fractions pooled after HiTrap Heparin column; 5: pure mDnmt1(731–1602) after size exclusion chromatography; 6: protein marker.

The purified mDnmt1(731-1602) construct, which includes the two bromo-adjacent homology (BAH) domains as well as the methyltransferase domain (see Figure 32), was reported to retain catalytic activity with both hemimethylated and non-methylated substrates.[189] Additionally, the truncated mouse protein shares 85 % sequence homology with the human equivalent and can therefore provide useful insights in the enzymatic features of hDnmt1.[94]

6.2.4.2 Methylation assay

One of the first issues addressed with the recombinant mDnmt1(731-1602) was whether it could methylate cytosines opposite fC or caC in a CpG context. In fact, given that these oxidized cytosine derivatives are suggested to be intermediates in a pathway of demethylation (or dehydroxymethylation), it would be interesting to understand their potential for maintenance of the methylation pattern during DNA replication. Previous reports have in fact shown that *in vitro* maintenance methylation opposite hmC is reduced, suggesting that hmC might be involved in replication-coupled dilution of mC,[123] but no information is currently available concerning the behaviour of fC and caC in this context.

Purified mDnmt1 was incubated with double stranded DNA bearing a single, hemimodified CpG site containing either mC, hmC, fC or caC (annealed ODN3 + ODN4, see Figure 33). Experiments were performed in the presence of SAM, and control experiments were done in the absence of the methyl-donor. As shown in Figure 33, mDnmt1 shows methylation activity opposite methylcytosine and hydroxymethylcytosine. Note that the experiment only assesses the ability of the protein to methylate opposite a modified cytosine-containing CpG site, no infomation can be deduced concerning the kinetics of methylation. The ability of mDnmt1 to methylate opposite hmC is in agreement with previous studies,[123] although Hashimoto *et al*. reported a reduced activity compared to methylation opposite mC. Nevertheless, the result might stem from the extended incubation times of the experiment, and, more importantly, Dmnt1 activity might be further enhanced by the absence of the CXXC domain and the autoinhibitory CXXC-BAH1 linker in this protein construct.[94] In fact, as described by Song *et al*., the CXXC and the autoinhibitory linker are essential in discrimination between unmethylated and hemimethylated DNA and might play a similar role in discrimination of hydroxymethylated CpG sites.[94]

Interestingly, there is almost no methylation activity observed opposite formylcytosine and carboxycytosine (Figure 33). This observation suggests that these oxidized cytosine derivatives do not contribute to maintenance of the methylation pattern upon DNA replication. These results therefore support a possible role for these bases in the removal of methylation by replication-coupled dilution, where lack of methylation opposite fC and caC results in a gradual erasure of the methylation pattern during cell division.

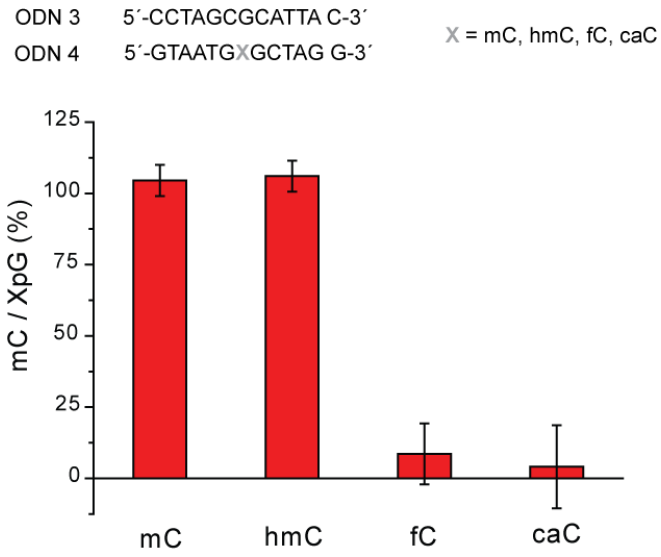


Figure 33 Methylation activity assay for recombinant mDnmt1(731-1602). ODN3 and ODN4 were annealed and used for methylation activity assay. Plot details the percental change of the mC content per XpG site, where X = mC, hmC, fC or caC (shown on x-axis) with respect to control reaction (for exact values, see Table **16**).

6.2.4.3 Interaction with oxidised cytosine derivatives

Additionally, we were interested in investigating whether mDnmt1 is able to generate a covalently bound intermediate with the oxidized cytosine derivatives. In fact, decarboxylation, as proposed by Schiesser *et al*., is thought to proceed via initial saturation of the C5–C6 bond by nucleophilic attack at the C6 position.[132]. However, failure to remove the carboxyl- or formyl- group at the C5 position in the presence of a saturated C5-C6 bond would result in trapping of the mDnmt1-DNA adduct in a manner similar to aza-SAM. This in itself could provide a pathway for demethylation, since trapping of Dnmt1 would result in a suicide complex which would then be processed by the nucleotide excision repair (NER) machinery and lead to replacement by cytosine.

Double-stranded, fluorescently-labelled oligonucleotides containing a hemi-modified CpG site with various cytosine derivatives (ODN5 + ODN6) were incubated in the presence of mDnmt1(731-1602) and subsequently analyzed by SDS-PAGE. Covalently bound oligonucleotides are expected to migrate at a molecular weight corresponding to about 115 kDa (i.e. the molecular weight of mDnmt1), while unbound DNA should migrate at a lower molecular weight.

As it can be noted in Figure 34, in all cases most of the fluorescently-labelled DNA migrates at the bottom of the gel. In the case of cytosine, mC and hmC, both in the presence or absence of *S*-adenosyl-methionine (SAM), a faint band is also observed at high molecular weight, suggesting some extent of covalent trapping. This might reflect a rapid equilibrium of C5–C6 bond saturation upon nucleophilic attack of the unmodified cytosine in the counterstrand prior to methylation, consistent with the results detailed in 6.2.4.2 showing methylation activity of mDnmt1 opposite these bases. Alternatively, it might result from erroneous attack at the C6 position of the modified nucleoside. In fact, since both strands are fluorescently-labelled, we cannot distinguish whether covalent trapping involves the modified strand or the unmodified counterstrand.

On the contrary, in the case of fC and caC no band is detected at higher molecular weight. This suggests that no nucleophilic attack is taking place at the C6 position of the modified nucleoside, and that most likely there is also no attack occurring at the C6 position of the unmodified CpG cytosine in the counterstrand. This is again in agreement with results reported in section 6.2.4.2, where no methylation activity is detected opposite fC and caC. However, from this study we cannot exclude that mDnmt1 might nevertheless interact with the fC and caC-containing CpG site. Presence of SAM did not show any effect on covalent trapping opposite these bases.

ODN 5 5′ - TGA GCA TCX GGT CAA CGT CCC TTC GGA A FITC - 3′
ODN 6 5′ - FITC TTC CGA AGG GAC GTT GAC CGG ATG CTC A - 3′
X = C, mC, hmC, fC, caC

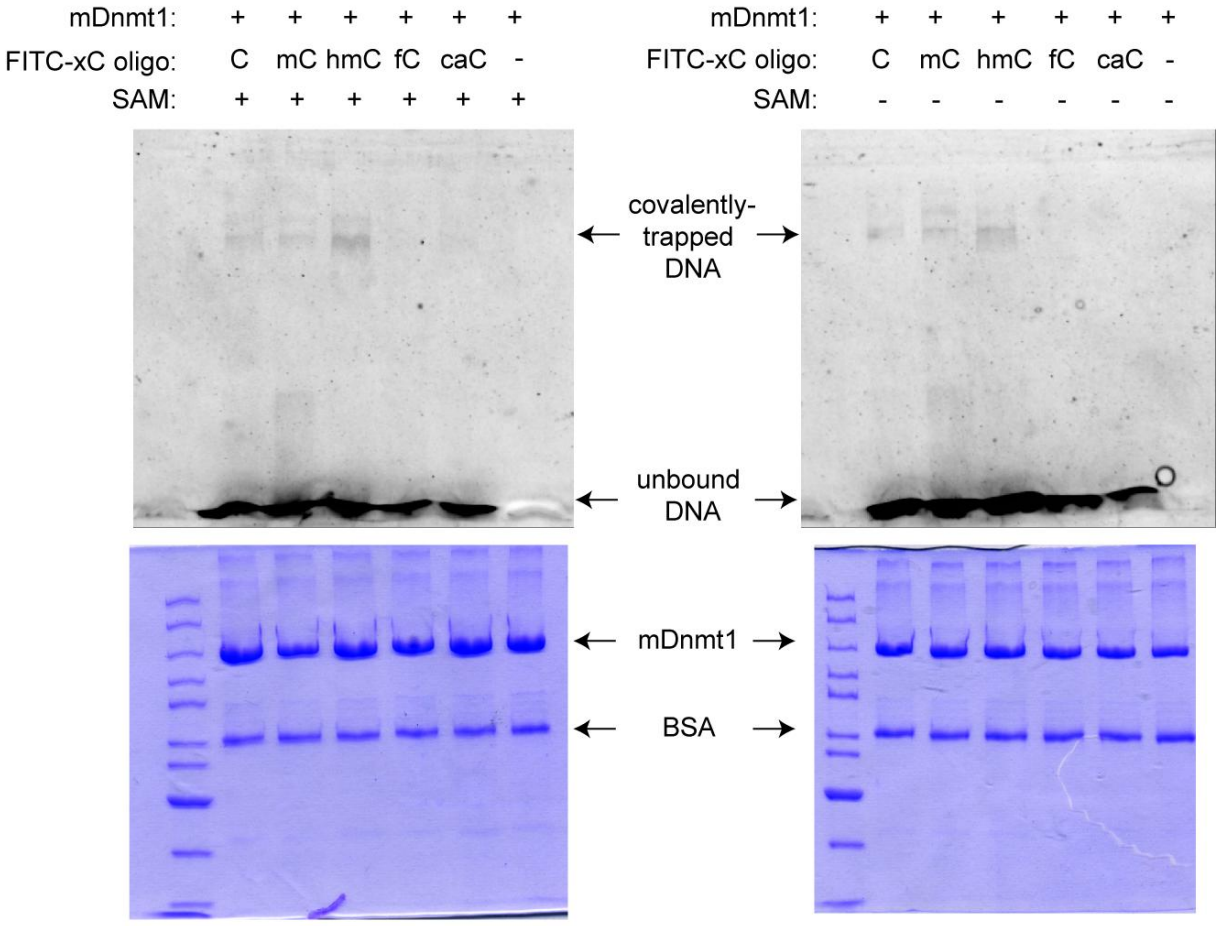


Figure 34 Investigation of covalent trapping of mDnmt1. mDnmt1 was incubated with double stranded DNA containing a XpG site (X = C, mC, hmC, fC, caC). Top panels = fluorescence detection of DNA on SDS gel; bottom panels = protein detection by coomassie blue staining of SDS gel.

In order to distinguish whether failure to detect a trapped intermediate for fC and caC is due to altered affinity of mDnmt1 or instead to lack of nucleophilic attack at the C6 position, binding studies of mDnmt1(731-1602) in the presence of fC or caC-containing double-stranded oligos were attempted by

Electrophoretic Mobility Shift Assay (EMSA). However, information concerning the binding affinity could not be obtained due to precipitation of the protein upon loading of the complex on the gel. Future studies could be performed by Fluorescence Polarization to avoid unwanted effects arising from interaction of the protein with the gel matrix.

6.2.4.4 Decarboxylation and deformylation assays

Purified mDnmt1 was also tested for any decarboxylation or deformylation activity. The purified methyltransferase was incubated with caC- or [$^{15}N_2$]-fC-containing oligonucleotides with or without SAM. The DNA was subsequently isolated and analyzed by LC-ESI-MS/MS or LC-ESI-MS, (for caC- or [$^{15}N_2$]-fC-containing oligonucleotides, respectively). In all cases, no conversion of caC or of [$^{15}N_2$]-fC to [$^{15}N_2$]-C or [$^{15}N_2$]-mC could be detected. This suggests that, if Dnmt1 is involved in the process of decarboxylation or deformylation, then the methyltransferase might require some additional protein or cofactor in order to perform this reaction. Previous studies have also reported that Dnmt1 is unable to act as a dehydroxymethylase, contrary to Dnmt3a and Dnmt3b.[133] Therefore future studies should address whether Dnmt1 can indeed act as a deformylase or decarboxylase in combination with additional proteins or cofactors, or whether its contribution to demethylation is only through replication-coupled dilution.

6.3 Conclusion

The study presented in this chapter investigated the putative mechanism of active demethylation *via* direct C-C bond cleavage of the oxidized cytosine derivatives hmC, fC and caC. Selectively labelled substrates were generated for monitoring of the demethylation products. Incubation experiments performed with different cell nuclear extracts and protein samples did not reveal any product of C-C bond cleavage-dependent demethylation, hinting that, under the experimental conditions tested, this process is not taking place. This is to some extent not surprising, given that we are expecting the decarboxylase enzyme to be very tightly regulated in order to ensure correct demethylation, possibly linked to specific time points of cell growth and differentiation and/or in a sequence-specific manner. On the other hand, investigations performed with recombinant mouse Dnmt1 revealed that maintenance of DNA methylation during replication does not occur opposite fC and caC, supporting a role for this modified nucleosides in removal of the methylation pattern via replication-coupled dilution.

6.4 Materials and methods

6.4.1 *In vitro* experiments using cell lysate samples

Preparation of the intact nuclei or nuclear lysate for *in vitro* studies was performed using an adapted Dignam protocol.[190] All the steps were performed on ice or at 4 °C. The cells used in this study (wild type mESC J1 or HEK-293T) were provided by Udo Müller, from research group of Prof. H. Leonhardt (Department of Biology, LMU). Cell pellets were resuspended in 5 vol. of Lysate Buffer A

(10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT) supplemented with *Complete-EDTA free* protease inhibitor cocktail (*Roche*). After 10 min incubation cells were gently pelleted (10 min, 2000 rpm) and resuspended in 2 vol. of Lysate Buffer A. Homogenization was performed using a douncer (20 strokes) and verified using trypan blue staining. Nuclei were pelleted (10 min, 2000 rpm) and could be used directly for incubation with labelled DNA substrates. For nuclear extract preparation, the pelleted nuclei were resuspended in 2.5 vol. Lysate Buffer C (20 mM HEPES, pH 7.9, 20 % glycerol, 1.5 mM $MgCl_2$, 420 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT, *Complete EDTA-free* protease inhibitor). The mixture was incubated for 30 min followed by centrifugation (15 min, 13000 rpm). The nuclear extract (supernatant) was subsequently dialyzed for 2 h against 1 L of Lysate Buffer D (20 mM HEPES, pH 7.9, 25 % glycerol, 100 mM NaCl, 0.2 mM EDTA, 0.5 mM DTT) and subsequently used for incubation with labelled DNA substrates.

Incubation with DNA substrates was performed in Lysate Buffer A (for intact nuclei) or in Lysate Buffer D (nuclear extracts). Samples were incubated at 37 °C for 1 h, shaking gently (300 rpm). Labelled substrates and amounts used for each experiment are listed in Table 14.

Experiment	Cell type	Lysate preparation	DNA substrate	DNA substrate amount/exp.
1	HEK-293T vs HEK-293T overexpressing Dnmt3b1	nuclear extract	ODN1 + ODN2, [$^{15}N_2$]-caC (ds)	0.3 nmol/exp
2	mESC vs HEK-293T	intact nuclei	ODN1 + ODN2, [$^{15}N_2$]-caC (ds), [$^{15}N_2$]-caC (mC), [$^{15}N_2$]-caC (C), [$^{15}N_2$]-fC (C)	1 nmol/exp
3	mESC (differentiation time points 0, 4, 8, 16 h)	intact nuclei	ODN1 + ODN2, [$^{15}N_2$]-caC (ds)	1 nmol/exp
4	HEK-293T	intact nuclei and nuclear extract	150bp-Oct4, [D_2,$^{15}N_2$]-hmC	4 pmol/exp
5	mESC	nuclear extracts, anaerobic conditions	150bp-Oct4, [D_2,$^{15}N_2$]-hmC	4 pmol/exp

Table 14 Summary of *in vitro* experiments using cell extracts. ODN1 and ODN2 refer to oligonucleotides described in Figure 31. 150bp-Oct4 refers to PCR product containing labelled substrates.[185] Labelled nucleosides are listed under DNA substrates, in brackets is shown whether both strands contain the same modified nucleoside (ds), or whether the opposite CpG site contains a methylcytosine (mC) or cytosine (C).

Substrate DNA was subsequently isolated by extraction using *Roti-Phenol/CHCl3/isoamylalcohol* (*Roth*), precipitated with EtOH (3 vol), and purified using Streptavidin coated magnetic beads (*Dynabeads M-270*, *Invitrogen*) following the manufacturer´s procedure. Removal of the DNA from the beads was performed by applying a temperature gradient (20–99 °C).[191] DNA samples were subsequently digested to nucleosides as detailed in section 6.4.3 and analyzed by HPLC-ESI-MS.

6.4.2 *In vitro* experiments using protein pull-down samples

Overexpression of GFP-tagged target proteins was performed in HEK-293T cells, followed by cell lysis and protein pull-down following previously published procedures.[187] Protein pull-down samples were prepared by Udo Müller from the research group of Prof. H. Leonhardt (Department of Biology, LMU). Depending on the experiment, reactions were carried out in Buffer Pull down A (50 mM HEPES, pH 8.0, 100 µM ammonium iron (II) sulfate, 1 mM α-ketoglutarate, 2 mM ascorbic acid, 2.5 mM DTT, 100 mM NaCl, 1.2 mM ATP),[118b] or in Buffer Pull down B (10 mM Tris-HCl, pH 7.4, 50 mM NaCl, 1.5 mM $MgCl_2$).[133] Substrate DNA was hybridized (or simply resuspended) in 2x incubation buffer in half the reaction volume and added to the protein pull down sample (previously washed in 1x incubation buffer). Incubation condition, DNA substrates and protein samples are listed in Table 15. For every experiment, GFP protein (purified on beads as the other proteins) was used as a negative control.

Experiment	Protein tested	DNA substrate	Conditions
1	Tet1 (full length vs catalytic domain)	ODN1 + ODN2, [$^{15}N_2$]-caC (ds), [$^{15}N_2$]-caC (C), [$^{15}N_2$]-fC (ds); ODN1, [$^{15}N_2$]-caC, [$^{15}N_2$]-fC; 200 pmol/exp.	Buffer Pull down A, reaction vol. 60 µL, 37 °C, 1 h
2	Tet1, Tet2, Tet3; Dnmt1, Dnmt3a, Dnmt3b; Tet1 + Dnmt1, Tet1 + Dnmt3a, Tet1 + Dnmt3b	ODN1 + ODN2, [$^{15}N_2$]-caC (ds), 100 pmol/exp.; 150bp-Oct4, [D_2, $^{15}N_2$]-hmC, 4 pmol/exp.	Buffer Pull down A, SAM (160 µM), reaction vol. 60 µL, 37 °C, 1 h
3	Dnmt1	ODN1 + ODN2, [$^{15}N_2$]-caC (ds), [$^{15}N_2$]-caC (mC), [$^{15}N_2$]-caC (C); 100 pmol/exp.	Buffer Pull down B, reaction vol. 100 µL, 37 °C, 16 h
4	Dnmt3a, Dnmt3b	ODN1 + ODN2, [$^{15}N_2$]-caC (ds), [$^{15}N_2$]-caC (mC); 100 pmol/exp.	Buffer Pull down B, reaction vol. 100 µL, 37 °C, 16 h

Table 15 Summary of *in vitro* experiments using protein pull-down samples. ODN1 and ODN2 are described in Figure 31, while 150bp-Oct4 refers to labelled PCR product.[185] Description of DNA substrates is as noted in Table 14.

After incubation, DNA was isolated and processed as described in Section 6.4.1. The only difference was that digestion was performed directly after precipitation with EtOH, without further purification on streptavidin coated magnetic beads.

6.4.3 Enzymatic digestion of DNA samples

DNA samples isolated after *in vitro* experiments were resuspended in 25 µL H_2O. The first digestion step was carried out in 100 µM $ZnSO_4$ with 42 units Nuclease S1 (*Aspergillus oryzae*, *Roche*) and 5 units antarctic phosphatase (*New England Biolabs*) and incubated at 37 °C for 3 h. The second digestion step was performed by addition of $[Na]_2$-EDTA (final concentration 100 µM) and 0.2 units snake venom phosphodiesterase I (*Crotalus adamanteus venom*, *UBS Corporation*) and incubation at 37 °C for 3 h.

6.4.4 HPLC-ESI-MS analysis

Digested samples were analyzed using a *Thermo Finnigan LTQ Orbitrap XL* coupled to a *Dionex Ultimate 3000* HPLC system. Flow was set to 0.15 mL/min and chromatographic separation was performed using an *Uptisphere 120-3HDO* column (*Interchim*) at 30 °C. Elution buffer were Buffer A (0.01 % formic acid in H_2O) and Buffer B (0.01 % formic acid in MeCN). The gradient used was the following: 0 → 12 min, 0 % → 1 % buffer B; 12 → 20 min, 1 % → 2 % buffer B; 20 → 30 min, 2 % → 10 % buffer B; 30 → 35 min, 10 % → 80 % buffer B; 35 → 45 min, 80 % buffer B; 45 → 50 min, 80 % → 0 % buffer B; 50 → 60 min, 0 % buffer B. Elution was monitored by absorbance at 260 nm. The eluted sample was directly injected into the ion source without prior splitting, and ions were scanned in positive mode using a scan range of *m/z* 100–500. Parameters used were the following: sheath gas flow rate, 16 arb; auxiliary gas flow rate, 11 arb; sweep gas flow rate, 4 arb; spray voltage, 5.0 kV; capillary temperature, 275 °C; capillary voltage, 35 V; tube lens, 65 V; resolution, 60000.[132]

6.4.5 Overexpression and purification of mDnmt1

The plasmid harbouring the mDnmt1(731–1602) fragment was kindly provided by Dr. Jikui Song.[94] The plasmid was transformed in *E.coli* strain Rosetta 2(DE3) (*Novagen*) and cells were grown in LB medium in the presence of kanamycin (15 µg/mL) and chloramphenicol (34 µg/mL) to an OD600 = 0.6. Protein overexpression was induced with isopropyl-β-D-thiogalactopyranoside (0.4 mM) in the presence of $ZnCl_2$ (0.1 mM final concentration), and cells were let grown overnight at 20 °C.

Cells were resuspended in cold Dnmt1 Buffer A (50 mM Tris HCl, pH 8.0, 1 M NaCl, 25 mM imidazole, approximately 20 mL per L of *E.coli* culture), and lysed using a French press. The suspension was supplemented with a cocktail of protease inhibitors (*Complete EDTA-free*, *Roche*), and the lysate was clarified using a *Sorvall SS-34 rotor* (18000 rpm, 4 °C, 1 h). The supernatant was loaded on a pre-equilibrated *HisTrap HP 5 ml* column (*GE Healthcare*) and the protein was eluted with Dnmt1 Buffer B (50 mM Tris HCl, pH 8.0, 1 M NaCl, 250 mM imidazole). Eluted protein

fractions were concentrated using an *Amicon Ultra* centrifuge filter (*Millipore*, 30,000 MWCO), the buffer was exchanged to Dnmt1 Buffer C (20 mM Tris HCl, pH 7.5, 50 mM NaCl, 5 mM DTT) and the protein sample was loaded on a *HiTrap Heparin 5 mL* column (*GE Healthcare*). Elution was performed by applying a gradient to Dnmt1 Buffer D (20 mM Tris HCl, pH 7.5, 1 M NaCl, 5 mM DTT), where mDnmt1 is expected to elute at about 0.4–0.5 M NaCl. Fractions showing a significant level of purity were pooled, concentrated using an *Amicon Ultra* centrifuge filter (*Millipore*, 30,000 MWCO), and buffer-exchanged to Dnmt1 Buffer E (20 mM Tris HCl, pH 7.5, 250 mM NaCl, 5 mM DTT) and purified by size exclusion chromatography on a *Superdex 16/60 HiLoad 200* (*GE Healthcare*). Fractions with > 90 % purity were pooled together, concentrated and stored at -80 °C. For all steps, chromatographic purification was monitored by UV absorbance at 260 nm and 280 nm and purity of the eluted protein fractions was verified by SDS-PAGE.

6.4.6 Methylation assay

Oligonucleotides (ODN3 + ODN4) bearing a single CpG site (modified or unmodified) where annealed by heating to 95 °C for 4 min followed by cooling to 4 °C over a period of 45 min. Methylation assay was performed in 1 x *Dnmt1 Reaction Buffer* (*New England Biolabs*), with double stranded DNA (1 µM), purified mDnmt1 (1 µM), Bovine Serum Albumin (0.1 mg/mL) and with or without SAM (160 µM). Experiments were performed in triplicate. The total reaction volume was 20 µL. Samples were incubated at 37 °C for 3 h, followed by extraction using *Roti-Phenol/CHCl3/isoamylalcohol* (*Roth*) and by desalting using *ZipTip* (*Millipore*). Purified oligonucleotides were subsequently digested to single nucleosides (as detailed in section 6.4.3) and analyzed by LC-ESI-MS/MS. Mass spectrometry analysis was performed by Toni Pfaffeneder. Data is presented in Table 16.

XpG	mC / XpG (%)	St.Dev. %
mC	104,52	5,50
hmC	106,07	5,43
fC	8,55	10,72
caC	4,08	14,58

Table 16 Percental change of mC upon incubation with mDnmt1. Table lists the percental change in mC content between positive and negative controls (with or without SAM) per XpG site (X = mC, hmC fC or caC). Experiments were performed in triplicate.

6.4.7 Covalent trapping of mDnmt1 in the presence of modified cytosine derivatives

Fluorescently-labelled oligonucleotides containing a single CpG site with C, mC, hmC, fC or caC were annealed to unmodified CpG-containing counterstrands (also fluorescently-labelled). Annealed oligos (3 µM) were incubated in 1 x *Dnmt1 Reaction Buffer* (*New England Biolabs*) together with purified mDnmt1 (1 µM) and Bovine Serum Albumin (BSA, 0.1 mg/mL) in a total volume of 20 µL. Incubation was performed in the absence or presence of SAM (160 µM). Samples were incubated at

37 °C for 1 h, and subsequently denatured by addition of 0.2 vol. of 5 x SDS Loading Buffer (62.5 mM Tris, 4 % (w/v) SDS, 20 % (w/v) Glycerol, 5 % (w/v) β-mercaptoethanol, pH 6.8) and heating to 90 °C for 5 min. Samples were subsequently electrophoresed under denaturing conditions through a 8 % polyacrylamide gel and fluorescence detection was performed using a *Fuji LAS 3000* (*Raytest*).

6.4.8 Decarboxylation and deformylation assays

Decarboxylation assays were performed in the presence of double- and single-stranded DNA. Oligonucleotides were annealed in different combinations to yield either fully carboxylated CpG sites (ODN7 + ODN8), hemi-carboxylated/hemi-methylated CpG sites (ODN7 + ODN9), or mismatch sites with a caCpG opposite a ApG site (ODN7 + ODN10). Single-stranded caC-containing oligonucleotides were also tested (ODN7). A positive control was performed with hemimethylated strands to verify the activity of purified mDnmt1 (ODN11 + ODN12). Oligonucleotides are listed in Table 17.

Name	Sequence	Oligo
ODN 7	5´-TTTTTTTcaCGTTTTTTTcaCGTTTTTTTcaCGTTTTTTT	strand
ODN 8	5´-AAAAAAAcaCGAAAAAAAcaCGAAAAAAAcaCGAAAAAAA	counter-strand
ODN 9	5´-AAAAAAAmCGAAAAAAAmCGAAAAAAAmCGAAAAAAA	counter-strand
ODN 10	5´-AAAAAAAAGAAAAAAAAGAAAAAAAAGAAAAAAA	counter-strand
ODN 11	5´-TTTTTTTmCGTTTTTTTmCGTTTTTTTmCGTTTTTTT	strand
ODN 12	5´-AAAAAAACGAAAAAAACGAAAAAAACGAAAAAAA	counter-strand

Table 17 Oligonucleotides used for decarboxylation experiments.

Oligonucleotides (1.6 µM) were incubated with purified mDnmt1 (1 µM) in 1 x *Dnmt1 Reaction Buffer* (*New England Biolabs*), alongside BSA (0.1 mg/mL), with or without SAM (160 µM). The total reaction volume was 33.5 µL and each experiment was done in triplicates. Incubation was performed at 37 °C for 16 h. Samples were subsequently extracted with *Roti-Phenol/CHCl3/isoamylalcohol* (*Roth*), precipitated with EtOH (3 vol), washed twice with ice-cold 70 % EtOH and resuspended in ddH_2O. The purified, desalted DNA was then digested to nucleosides and analyzed by ESI-LC-MS/MS as detailed in Section 6.4.6. Mass spectrometry analysis was performed by Toni Pfaffeneder.

Deformylation experiments were performed using [$^{15}N_2$]-labeled fC-containing oligonucleotides (see Table 18). Strands were annealed either in the combination ODN13 + ODN14, or ODN13 + ODN15. Incubation of annealed oligonucleotides (1 µL) with mDnmt1 (1 µM) was done as described for the decarboxylation experiment, except for the final reaction volume which was 100 µL. Samples were

incubated at 37 °C for 3 h, followed by extraction using using *Roti-Phenol/CHCl3/isoamylalcohol* (*Roth*) and by desalting using *ZipTip* (*Millipore*). Samples were subsequently digested to single nucleosides and analyzed by high resolution HPLC-ESI-MS.

Name	Sequence	Oligo
ODN 13	5´ - GTA ATG fCGC TAG G	$[^{15}N_2]$-fC
ODN 14	5´- CCT AGC GCA TTA C	C counter-strand
ODN 15	5´- CCT AGmC GCA TTA C	mC counter-strand

Table 18 Oligonucleotides used for deformylation experiments

7 Synthesis of 5-hydroxymethyl-, 5-formyl- and 5-carboxycytidine-triphosphates and their incorporation into oligonucleotides by Polymerase Chain Reaction

ORGANIC LETTERS

2013
Vol. 15, No. 2
366–369

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

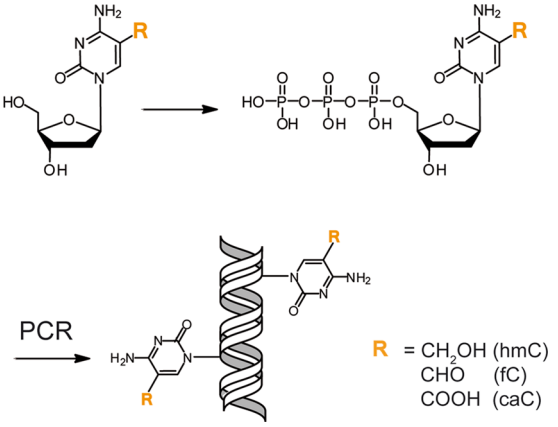
Barbara Steigenberger,[†] Stefan Schiesser,[†] Benjamin Hackner, Caterina Brandmayr, Silvia K. Laube, Jessica Steinbacher, Toni Pfaffeneder, and Thomas Carell*

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

Thomas.Carell@lmu.de

Received December 4, 2012

ABSTRACT



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

Recently three new nucleobases were discovered in DNA isolated from mouse embryonic stem cells.[1−4] The three nucleobases are oxidation products of 5-methylcytosine (mC), which is a base that regulates transcriptional activity.[5] The oxidation reaction of 5-methylcytosine to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and finally 5-carboxycytosine (caC) is now understood to be performed by 10-11-translocon proteins (TET1−3).[6] These enzymes are α-ketoglutarate dependent oxidases, which directly utilize molecular oxygen for the oxidation reaction.[7] The three new nucleobases are currently thought to be involved in epigenetic programming of cells, and they could be intermediates of a long searched for pathway of active demethylation.[8−10] Recently new sequencing methods that allow the genome wide localization of hmC[11,12] and fC[13] in genomic DNA were reported.

[†] These authors contributed equally.

(1) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
(2) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
(3) Pfaffeneder, T.; Hackner, B.; Truss, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008.
(4) He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C.-X.; Zhang, K.; He, C.; Xu, G.-L. *Science* **2011**, *333*, 1303.
(5) Law, J. A.; Jacobsen, S. E. *Nat. Rev. Genet.* **2010**, *11*, 204.
(6) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.
(7) Loenarz, C.; Schofield, C. J. *Chem. Biol.* **2009**, *16*, 580.
(8) Jurkowski, T. P.; Jeltsch, A. *ChemBioChem* **2011**, *12*, 2543.
(9) Gu, T.-P.; Guo, F.; Yang, H.; Wu, H.-P.; Xu, G.-F.; Liu, W.; Xie, Z.-G.; Shi, L.; He, X.; Jin, S.-g.; Iqbal, K.; Shi, Y. G.; Deng, Z.; Szabo, P. E.; Pfeifer, G. P.; Li, J.; Xu, G.-L. *Nature* **2011**, *477*, 606.
(10) Ladwein, K. I.; Jung, M. *Angew. Chem., Int. Ed.* **2011**, *50*, 12143.

10.1021/ol3033219
Published on Web 01/03/2013

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

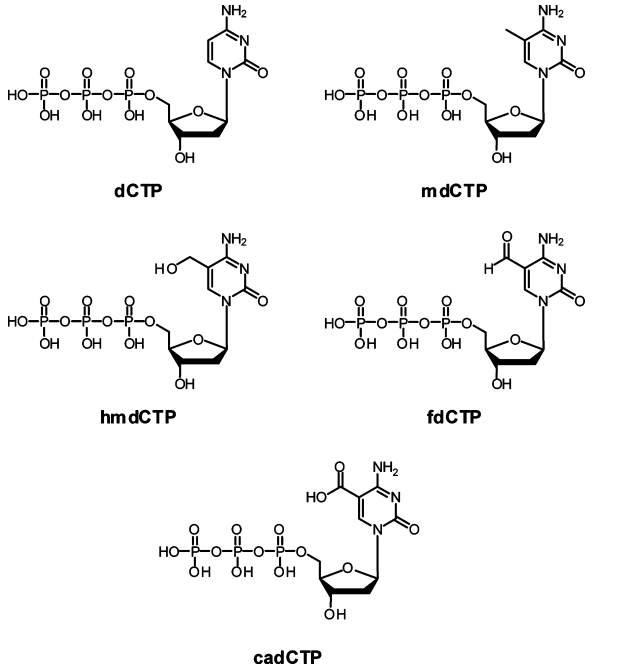
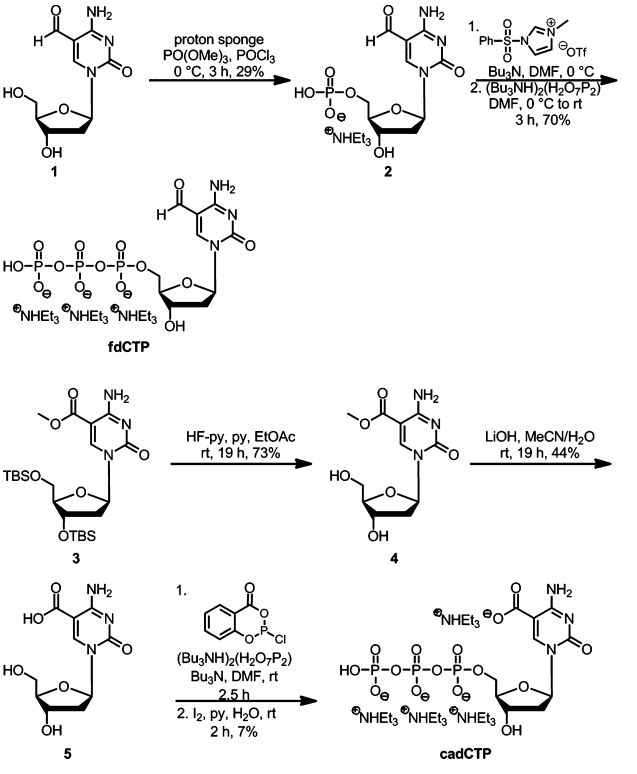


Figure 1. Depiction of the canonical DNA base dC and of mC as well as of the new epigenetic bases hmC, fC, and caC as triphosphates.

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

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

Scheme 1. Synthesis of fdCTP and cadCTP



(11) Booth, M. J.; Branco, M. R.; Ficz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S. *Science* **2012**, *336*, 934.
(12) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C.-X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J.-H.; Jin, P.; Ren, B.; He, C. *Cell* **2012**, *149*, 1368.
(13) Raiber, E.-A.; Beraldi, D.; Ficz, G.; Burgess, H. E.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, *13*, R69.
(14) Tardy-Planechaud, S.; Fujimoto, J.; Lin, S. S.; Sowers, L. C. *Nucleic Acids Res.* **1997**, *25*, 553.
(15) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671.
(16) Dai, Q.; He, C. *Org. Lett.* **2011**, *13*, 3446.
(17) Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chem.—Eur. J.* **2011**, *17*, 13782.
(18) Dai, Q.; He, C. *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc.: 2001.
(19) Jäger, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M. *J. Am. Chem. Soc.* **2005**, *127*, 15071.
(20) Kuwahara, M.; Nagahima, J.; Hasegawa, M.; Tamura, T.; Kitagata, R.; Hanawa, K.; Hososhima, S.; Kasamatsu, T.; Ozaki, H.; Sawai, H. *Nucleic Acids Res.* **2006**, *34*, 5383.
(21) Shoji, A.; Hasegawa, M.; Hososhima, S.; Kuwahara, M.; Ozaki, H.; Sawai, H. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 776.
(22) Globisch, D.; Münzel, M.; Müller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. *PLoS ONE* **2010**, *5*, e15367.
(23) Mohamady, S.; Desoky, A.; Taylor, S. D. *Org. Lett.* **2011**, *14*, 402.

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

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

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

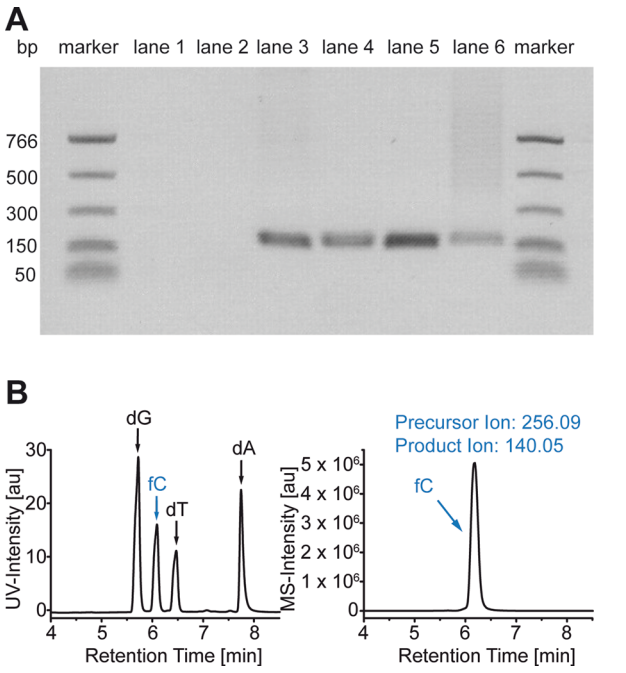


Figure 2. (A) Depiction of the results of the PCR analyzed by gel electrophoresis. Lane 1: without template. Lane 2: without any dC derivative. Lane 3: dCTP. Lane 4: protected D_2,$^{15}N_2$-hmdCTP. Lane 5: fdCTP. Lane 6: cadCTP. (B) UV trace (left) and mass trace of fC (right) of a fully digested PCR mixture with fdCTP.

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

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

(24) Guan, L.; van der Heijden, G. W.; Bortvin, A.; Greenberg, M. M. *ChemBioChem* **2012**, *12*, 2184.
(25) Ludwig, J.; Eckstein, F. *J. Org. Chem.* **1989**, *54*, 631.
(26) Caton-Williams, J.; Smith, M.; Carrasco, N.; Huang, Z. *Org. Lett.* **2011**, *13*, 4156.
(27) Raindlová, V.; Pohl, R.; Šanda, M.; Hocek, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 1064.
(28) Raindlová, V.; Pohl, R.; Hocek, M. *Chem.—Eur. J.* **2012**, *18*, 4080.

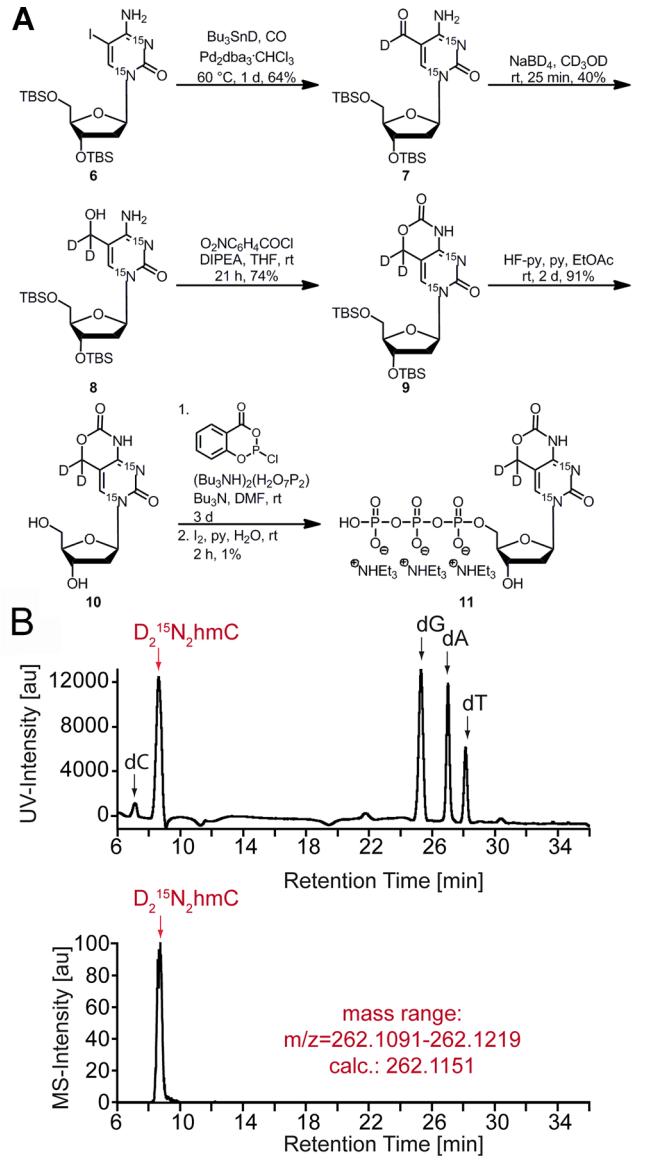


Figure 3. (A) Synthesis of the protected D_2,$^{15}N_2$-hmdCTP. (B) UV trace (top) and D_2,$^{15}N_2$-hmC mass trace (bottom) of a fully digested PCR mixture.

performed so far, which limits our ability to create DNA fragments with modified hmC, e.g. with isotopically labeled compounds as needed for mass spectrometry based quantification and proteomics studies. The synthetic challenge associated with the synthesis of hmdCTP is the benzylic hydroxyl group of the hmC heterocycle which is more reactive than the primary 5′-OH group. This makes it difficult to access the 5′-monophosphate directly from hmC. To circumvent this problem and to enable the PCR incorporation of modified hmC building blocks, we investigated the ability to insert a protected hmC derivative by PCR followed by deprotection. The synthesis of the protected hmdCTP building block and the results of the PCR study are shown in Figure 3.

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

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

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

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

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

The authors declare no competing financial interest.

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

ARTICLE
PUBLISHED ONLINE: 18 MAY 2014 | DOI: 10.1038/NCHEMBIO.1532

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

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

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

Methylcytosine is an epigenetically important nucleobase associated with the control of transcriptional activity, genomic imprinting, X-chromosome inactivation and suppression of transposable elements[1]. Controlled formation and removal of mC at specific genomic loci is critical for correct genome programming or reprogramming during cellular differentiation[2]. Recently, it was discovered that Tet proteins (Tet1–3) oxidize mC to give the oxidized C-derived nucleobases hmC[3], 5-formylcytosine (fC)[4,5] and 5-carboxycytosine (caC)[5,6], whose biological functions are still yet unclear (**Fig. 1a**)[7]. As fC and caC are both removed by thymine DNA glycosylase (Tdg)[6,8], it is currently assumed that they serve as intermediates of an active DNA demethylation process involving base excision repair. In addition to these oxidized C derivatives, cells also contain oxidized T nucleobases such as hmU and fU. These compounds are currently known as oxidative lesions that are thought to form upon the reaction of T with reactive oxygen species (ROS)[9,10]. It was recently suggested that hmU might also be produced by deamination of hmC, a hypothesis that remains controversial[11–14]. Deamination of hmC, situated in a base pair with G (hmC:G), would give rise to hmU:G mismatches, which are known substrates for the DNA glycosylases Tdg, Smug1, Mbd4, Ung2 (ref. 15), Neil1 and Nthl1 (ref. 16). Deamination of hmC:G to hmU:G followed by mismatch repair would therefore establish an alternative pathway to active demethylation (**Fig. 1a**).

To unravel the origin of oxidized nucleobases, and of hmU in particular, in DNA from mESCs, we performed isotope tracing and quantitative MS studies using the chemically synthesized isotopologs of mC, hmC, fC, caC, hmU and fU as internal standards (**Fig. 1b** and **Supplementary Results**, **Supplementary Fig. 1**). For the assessment of oxidation products that are formed by the action of ROS, we additionally quantified 8-oxo-G because 8-oxo-G is a well-established ROS reaction product formed from G[17,18]. We show here that hmU is generated enzymatically from thymidine during stem cell differentiation by the action of the Tet enzymes. A proteomic analysis provides new insight into how genomic hmU can influence the binding of chromatin remodeling proteins and transcription factors.

RESULTS

hmU is present at elevated levels in mESCs

We first created an inventory of the named nucleosides (**Supplementary Fig. 1**) in mESCs (**Fig. 1c**) and adult cortex tissue (**Fig. 1d**). In mESCs, we observed that hmC is, as expected, the most abundant oxidized pyrimidine (10% relative to mC), followed by fC (~1–2% of hmC). The ROS marker 8-oxo-G was detected at similar levels (~45% of fC), showing that nonenzymatic, ROS-induced oxidations of nucleobases are important processes, as expected. We also found relatively high levels of fU (22% of fC). Clearly detectable were also caC and hmU. hmU, which is at the center of this study, was unequivocally detected, as shown by its retention time and its specific fragmentation pattern, which were found to be identical with that of the internal standard [D_2]hmU (**Fig. 1e**). Notably, both caC and hmU were present in comparable amounts (5% of fC). In adult mouse cortex DNA (**Fig. 1d**; for other tissues see

[1]Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, München, Germany. [2]Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Berlin, Germany. [3]Department of Biomedicine, University of Basel, Basel, Switzerland. [4]Center for Integrated Protein Science at the Department of Pharmacy—Center for Drug Research, Ludwig-Maximilians-Universität München, München, Germany. [5]Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. [6]Department of Molecular Cancer Research, Cancer Genomics Netherlands, Utrecht, The Netherlands. [7]Present address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands. [8]These authors contributed equally to this work.
*e-mail: markus.mueller@cup.uni-muenchen.de or thomas.carell@cup.uni-muenchen.de

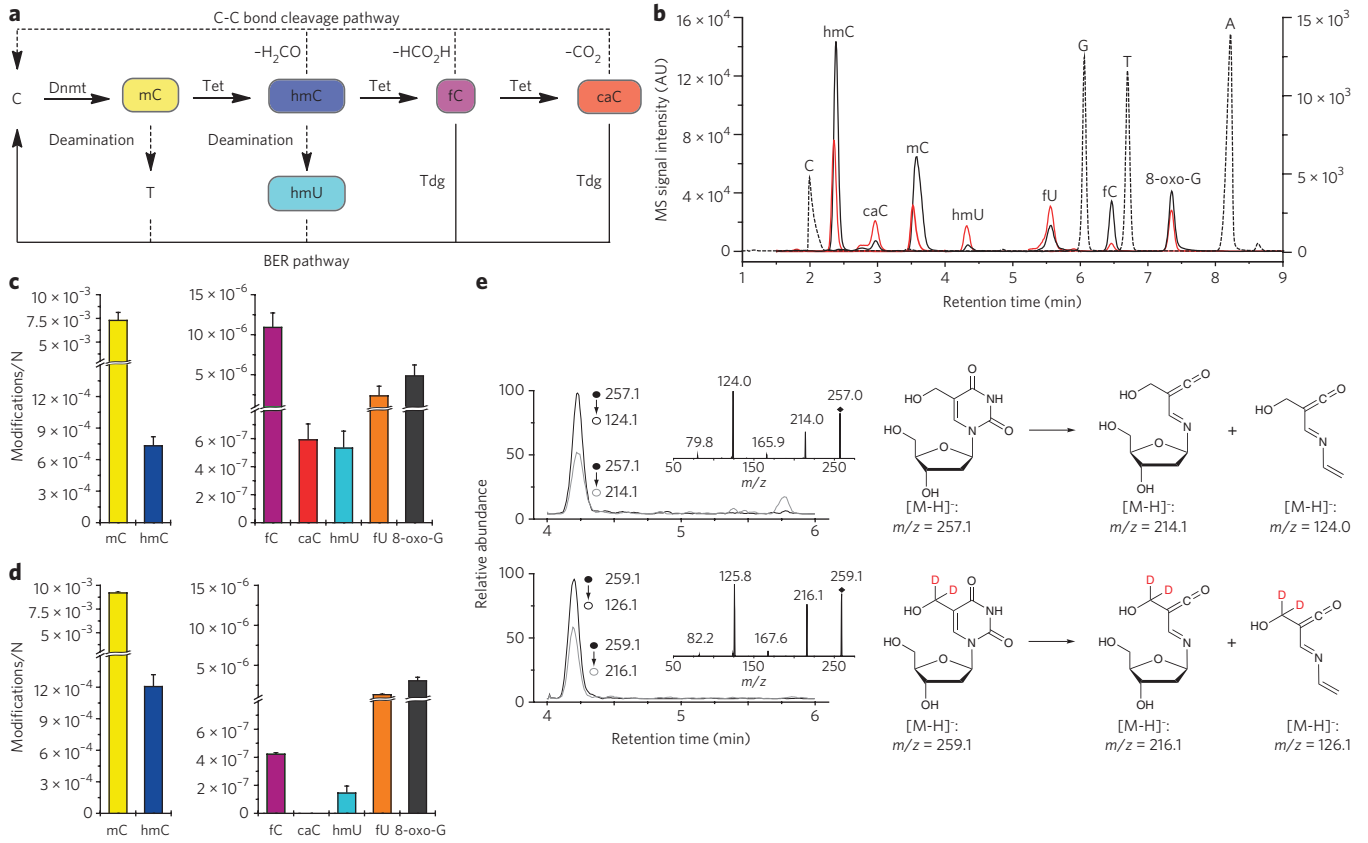


Figure 1 | Metabolism of cytosine derivatives, their detection by LC-UV-ESI-MS/MS and levels in mESCs and mouse cortex. (**a**) Potential active demethylation pathways. (**b**) Overlaid LC/UV and LC/MS/MS chromatograms of a representative DNA sample from mESCs. The dotted LC/UV chromatogram of C, G, T and A is scaled arbitrarily; the overlaid LC/MS/MS chromatograms of hmC, mC, fC and 8-oxo-G are scaled to the left *y* axis; the LC/MS/MS chromatograms of caC, hmU and fU are scaled to the right *y* axis. Red chromatograms refer to the corresponding labeled internal standards depicted in **Supplementary Figure 1**. AU, arbitrary units. (**c**,**d**) DNA modification levels per nucleoside (N) in mESCs (**c**; WT01, $n = 7$) and 3-month-old mouse cortex tissue (**d**; $n = 3$). Depicted are mean values ± s.d. (**e**) Representative LC/MS/MS chromatograms for identification and quantification of hmU. Shown are the overlaid chromatograms for the two characteristic fragment ions of hmU (top trace) and the [D_2]hmU internal standard (bottom trace) derived from a mESC DNA sample. The insets show the MS/MS full-scan spectra of synthetic hmU and [D_2]hmU matching the proposed fragmentation pathway.

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

We next investigated this in more detail and quantified the levels of hmU in three different mESC lines (WT01, J1 and R1) and in a variety of tissues (**Fig. 2a**). Clearly, the hmU levels were higher in mESCs than in somatic tissues by factors of 2–15. The detected hmU levels correspond roughly to 500–1,700 hmU bases per genome in tissue and 2,900–7,800 in mESCs (**Fig. 2a**) depending on the cell type and growth conditions. We cannot explain these differences by elevated oxidative stress levels in mESCs (**Supplementary Fig. 3**). We next analyzed the levels of C, mC and the oxidized pyrimidines hmC, fC, hmU and fU as well as 8-oxo-G in somatic tissues (**Supplementary Fig. 2**) and performed a correlation and cluster analysis to reveal relationships of the modification levels (**Fig. 2b**). The data analysis confirmed that the low levels of hmU in somatic tissues correlate strongly (significant on a level <0.001) with the levels of fU and 8-oxo-G within a separated cluster. This result showed that in somatic cells, hmU and fU are ROS-induced reaction products. If we assume that in somatic tissues hmU is exclusively formed by ROS, the data showed that in mESCs, 70–80% of the detected hmU is produced by ROS-independent processes (**Supplementary Fig. 3**).

hmU formation by oxidation of T

To analyze the origin of the oxidized nucleobases in mESCs, we performed isotope tracing experiments (**Fig. 3a–c** and **Supplementary Fig. 4**). Substitution of L-methionine with [*methyl*-$^{13}CD_3$]L-methionine ([$^{13}CD_3$]L-Met) in the growth medium is known to give the labeled *S*-adenosylmethionine cofactor, which is needed for the conversion of C to mC. Replacement of L-Met by [$^{13}CD_3$]L-Met for 5 d (2 passages) in the medium at a concentration of 0.2 mM furnished 89% labeled [$^{13}CD_3$]mC, 88% labeled [*hydroxymethyl*-$^{13}CD_2$]hmC and 93% labeled [*formyl*-^{13}CD]fC (**Fig. 3b** and **Supplementary Fig. 4**). Within the detection limit (7 and 50 molecules per 10^8 nucleosides, respectively), no incorporation of the isotopes ^{13}C and D into hmU and fU was observed, showing that hmC is not the precursor of hmU. We next added isotope-labeled [$^{13}C,^{15}N_2$]thymidine ([$^{13}C,^{15}N_2$]T) to the growth medium and observed ~76% of label incorporation into T, hmU and fU (**Fig. 3a**, **Supplementary Fig. 4** and **Supplementary Table 1**). The combined data showed that hmU is not generated by deamination of hmC but by oxidation of T (T → hmU; **Fig. 3c**). Consequently, all of the detected hmU

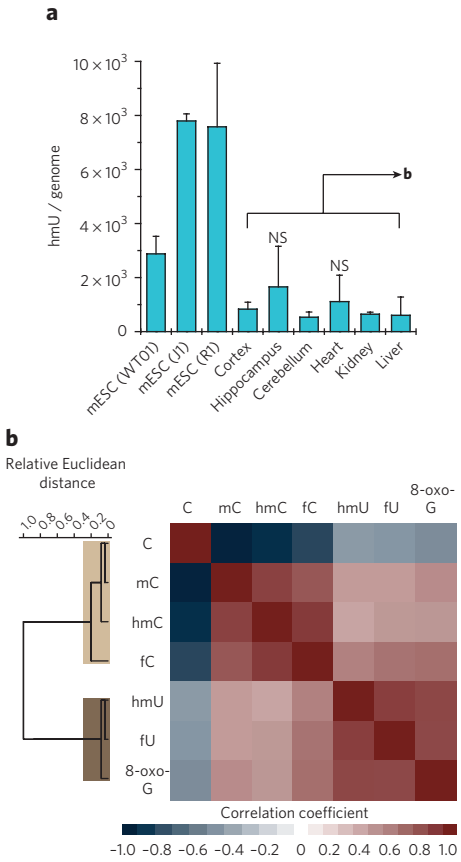


Figure 2 | hmU is present at elevated levels in mESCs compared to tissue. (**a**) hmU levels per genome in mESCs (n_{WT01} = 7 replicates, n_{J1} = 2, n_{R1} = 3) and mouse tissue (3-month-old individuals, n = 3). Levels per genome were obtained considering a mouse genome size of 2.7 × 10^9 base pairs. Depicted are mean values ± s.d. The differences between mESCs and mouse tissues are significant (P = 2.9 × 10^{-5} to 3.6 × 10^{-2}; unpaired two-tailed t-test) except for WT01 and hippocampus (P = 0.292) or heart (P = 0.069). These exceptions are due to higher hmU levels in the hippocampus and heart caused by higher background oxidation (higher 8-oxo-G levels). hmU levels normalized to oxidative background (8-oxo-G) levels are in **Supplementary Figure 3**. NS, not significant. (**b**) Unsupervised clustering analysis of Pearson correlation coefficients of 24 data sets of selected mouse organs at a defined time point (3-month-old individuals: cortex, hippocampus, cerebellum, heart, liver and kidney). Discussed correlations are strong to very strong (Pearson coefficient >0.7) and significant on a level <0.001.

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

The fact that we were unable to detect hmU derived from hmC deamination, which would be situated in a base pair with G (hmU:G), could also be explained by very fast repair. If repair of the hmU:G base pair is extremely efficient, for example, because deamination and glycosylase-based repair occurs in a tight complex of the involved enzymes, we would be unable to detect this type of hmU because of low steady state levels. Indeed, it was proposed that hmC deamination and hmU excision requires a complex of the cytidine deaminase Aid and the glycosylase Tdg[12]. To assess this possibility, we performed isotope tracing experiments using [$^{13}CD_3$]L-Met in *Tdg*$^{-/-}$ mESCs stably complemented with either empty vector (control) or a minigene expressing a catalytically incompetent Tdg at near-endogenous levels. These cells are able to form the Aid–Tdg complex, but the Tdg is inactive, which should give elevated hmU levels if deamination occurs. In both cell lines, we detected high levels of labeled hmC. In the control cells, [*hydroxymethyl*-$^{13}CD_2$] hmU was not detected. However, in the cell line complemented with inactive Tdg, some labeled [*hydroxymethyl*-$^{13}CD_2$]hmU was indeed observed, albeit only in small amounts (~7% of total hmU and ~0.06% of total hmC; **Supplementary Fig. 6**). Thus, the Tdg protein is required for deamination of hmC to hmU. In wild-type (WT) mESCs, the hmU:G mismatches are obviously repaired so quickly that they do not contribute to steady state levels of hmU. All of the detected hmU was derived from T oxidation and resided in hmU:A base pairs.

Tet enzymes form hmU in correlation with mC oxidation

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

In support of enzymatic T oxidation by Tet proteins, ectopic expression of the catalytic domain of Tet1 (Tet1cd) in HEK-293T cells led to a 65-fold increase for both hmC and hmU levels. This was not observed when a catalytic mutant of Tet1 (Tet1cm) was expressed (**Supplementary Fig. 7a**). To determine whether under these conditions hmU is generated by deamination, as previously suggested[11,12], we again replaced natural L-Met with [$^{13}CD_3$]L-Met in the medium, but we did not detect incorporation of heavy isotopes into hmU, providing evidence that the elevated hmU levels do not originate from hmC deamination (**Supplementary Table 1**). Finally, we confirmed enzymatic generation of hmU *in vitro* by incubating recombinant Tet1cd with a plasmid that was premethylated by bacterial methyltransferase M.SssI. In addition to oxidation of mC to hmC, fC and caC, we detected hmU at a prominent level of 9% relative to hmC (**Supplementary Fig. 7b**), showing that the catalytic center of the Tet enzymes clearly has the capacity to oxidize T to hmU.

To further confirm that hmU is formed in mESCs in the process of epigenetic reprogramming, we analyzed the dynamic changes of mC and all of the oxidized pyrimidines plus 8-oxo-G during differentiation. It was recently shown that mC and hmC levels sharply increase when mESCs maintained in the naive state are shifted to a primed state in serum-containing medium[23–25]. To investigate global kinetics of all of the Tet-generated oxidation products under more physiologically relevant priming conditions, we used established protocols based on serum-free N2B27 medium for differentiation of naive mESCs into states resembling that of post-implantation epiblasts[26,27]. mESCs were first grown for several passages in the presence of MEK and GSK3 inhibitors (dual inhibition or 2i conditions) and LIF to induce a hypomethylated state resembling that of the naive epiblast[23–25]. The data in **Figure 5a** are averaged from three independent differentiation experiments, each performed with two cell lines in the absence of growth factors. First, we observed that the fU and 8-oxo-G levels stayed constant (**Fig. 5a**), in line with the idea that both are ROS-derived products. In contrast, fC and caC levels peaked at about 8 h. Both hmU and hmC also peak between 8 h and 16 h. Isotope tracing with [$^{13}CD_3$]L-Met under these conditions provided no evidence for switched-on deamination of hmC to hmU (**Supplementary Table 1**). At these peak levels, we estimated that mESCs contain roughly 110,000 fC bases, 4,400 caC bases and

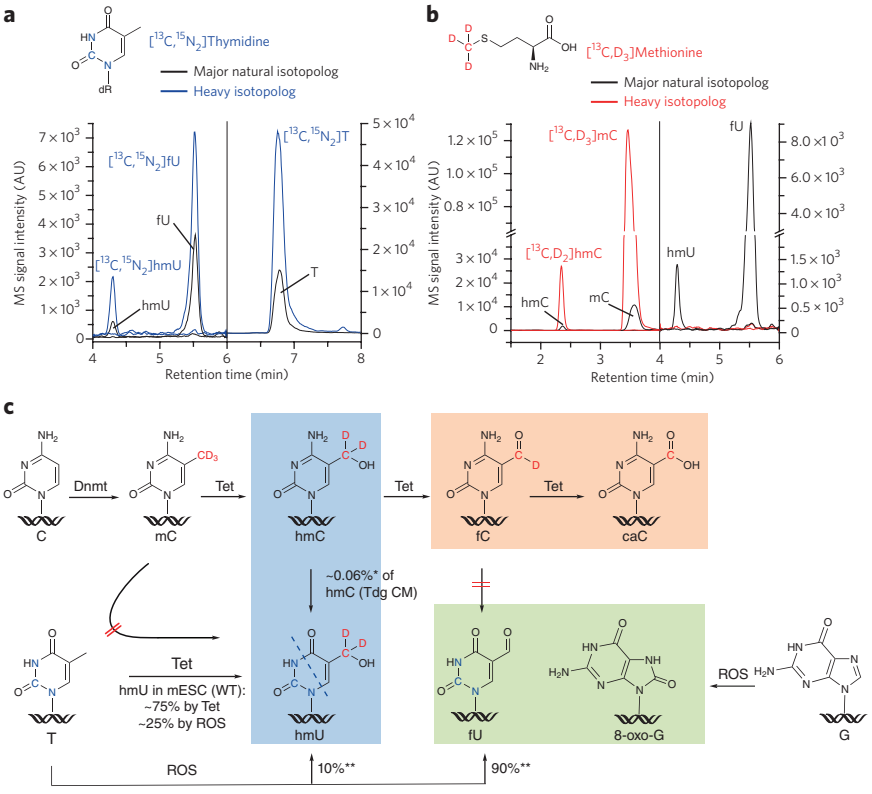


Figure 3 | hmU and fU are thymine oxidation products in WT mESCs with no detectable contribution from hmC or fC deamination. (**a**) Overlaid LC/MS/MS chromatograms of heavy ([^{13}C,^{15}N$_2$]hmU, [^{13}C,^{15}N$_2$]fU and [^{13}C,^{15}N$_2$]T; blue) and major natural (black) isotopologs of hmU, fU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

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

Recently, it was shown that exposure of naive mESCs to fibroblast growth factor 2 (FGF-2) and activin A (ActA) for 48 h under similar conditions as those described above gives a homogeneous cell population whose transcriptome closely resembles that of the post-implantation epiblast[26]. Under these conditions, mC levels increased more rapidly, approaching somatic levels within a time frame closely reflecting that observed during embryonic development (**Supplementary Fig. 9a**). Levels of hmC rose steadily throughout the 48-h time course. Whereas fC peaked at 36 h, hmU reached its maximum at 24 h (**Supplementary Fig. 9b**). qPCR data showed a robust induction of Dnmt3b that is most likely responsible for the increasing mC levels (**Supplementary Fig. 9a**). A transient peak of Tet1 expression at 24 h went in hand with rising hmU and fC levels, whereas Tet3 expression started rising slowly from the same time point and may drive further accumulation of hmC and fC (**Supplementary Fig. 9b**). Tdg was progressively upregulated together with the onset of methylation activity, whereas Smug1 showed little, if any, change (**Supplementary Fig. 10**). These data supported our view that hmU peaking is not caused by downregulation of the repair pathway.

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

hmU attracts specific readers

To obtain initial insight into a potential biological function of hmU in comparison to hmC, we screened for specific readers associating with hmU:A as well as with hmC:G and hmU:G using protein pulldown and relative quantification by LC/MS/MS. In previous stable isotope labeling by amino acids in cell culture (SILAC)-based proteomics studies with hmC, fC and caC containing oligonucleotides, we and others observed a high number of specific protein readers, arguing that the new bases influence a variety of different processes[28,29]. For this hmU study, we further developed this approach for the detection of proteins that directly interact with the modified bases hmU (as well as hmC for comparison) to get a more direct insight into their function (**Fig. 6a**). For the study, we not only included into the biotinylated DNA duplexes (24mers) an hmU (hmC) base but also equipped the counter strand with a polyethylene glycol–based linker carrying a reactive *N*-hydroxysuccinimide (NHS) ester moiety and a reductively cleavable disulfide bond in the middle. This reactive linker cannot be inserted into the DNA strand using solid phase synthesis. In addition, the NHS ester does not survive hybridization conditions. We therefore attached the linker as its azide derivative, using Cu(I)-catalyzed click chemistry, to an alkyne-bearing base present in the DNA duplex[30,31]. These DNA duplexes were subsequently incubated with nuclear extracts from mESCs. The DNA-bound protein complexes were isolated using streptavidin-coated magnetic beads (**Fig. 6a**)[32]. DNA duplexes with the canonical base pairs A:T and C:G at the respective positions served as reference strands. The NHS linker has two functions in the experiment: First, it will covalently trap the reader proteins specifically at the ε-amino groups of lysines, which allows the identification of transiently binding proteins as well. Second, because the linker is cleavable, it will tag the trapped lysine residues, leaving a defined label on those proteins that bind in close proximity to the modified bases hmC and hmU. This allowed us to distinguish protein readers that bound close to hmC and hmU from proteins that are secondary members of the complexes. The covalently trapped proteins were next tryptically digested and labeled with tandem mass tagging (TMT) isobaric tags to allow protein

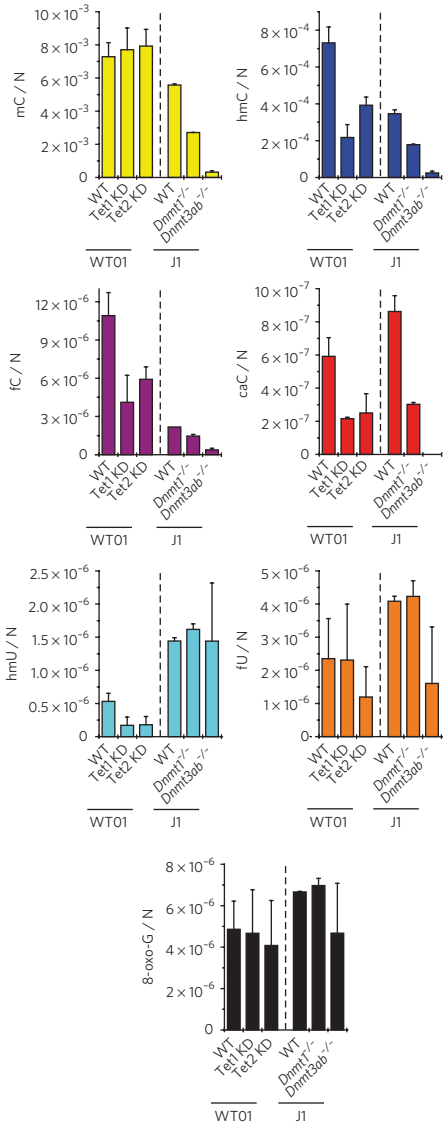


Figure 4 | Tet1 and Tet2 generate hmU in mESCs. Effect of Tet and Dnmt depletion on DNA modification levels in mESCs. J1 cell lines bearing homozygous *Dnmt*-null mutations and WT01 mESCs expressing Tet1 or Tet2 shRNAs (knockdown (KD)) were analyzed. Depicted are mean values per nucleoside (N) ± s.d.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

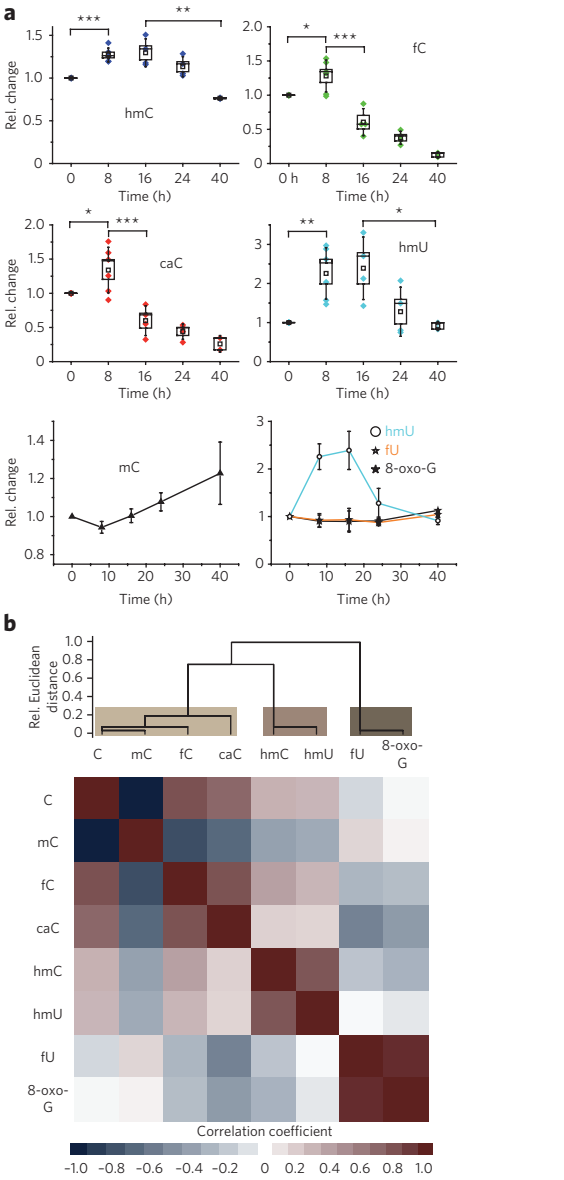


Figure 5 | hmU is produced during mESC differentiation. (**a**) Time course of DNA modification levels during early differentiation (0–40 h) of naive mESCs in the absence of growth factors. Box plot of relative modification changes for hmC, fC, caC and hmU averaged from six independent experiments using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

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

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

Among the dsRs we identified were the regulatory proteins Uhrf2, transcription factor HIVEP3 or the poly [ADP-ribose] polymerase 14 (Parp14), which showed that these proteins interacted directly with hmU. We also detected several H2A and H2B histone proteins among the dsRs and the sRs. Notably, in the group of the sRs, we observed also Dnmt3a and Dnmt3b, which are involved in gene expression regulation. Methyl CpG-binding protein 2 (MeCP2) was identified as well, but it showed only low enrichment. We also detected several transcriptional regulators, such as bromodomain PHD finger

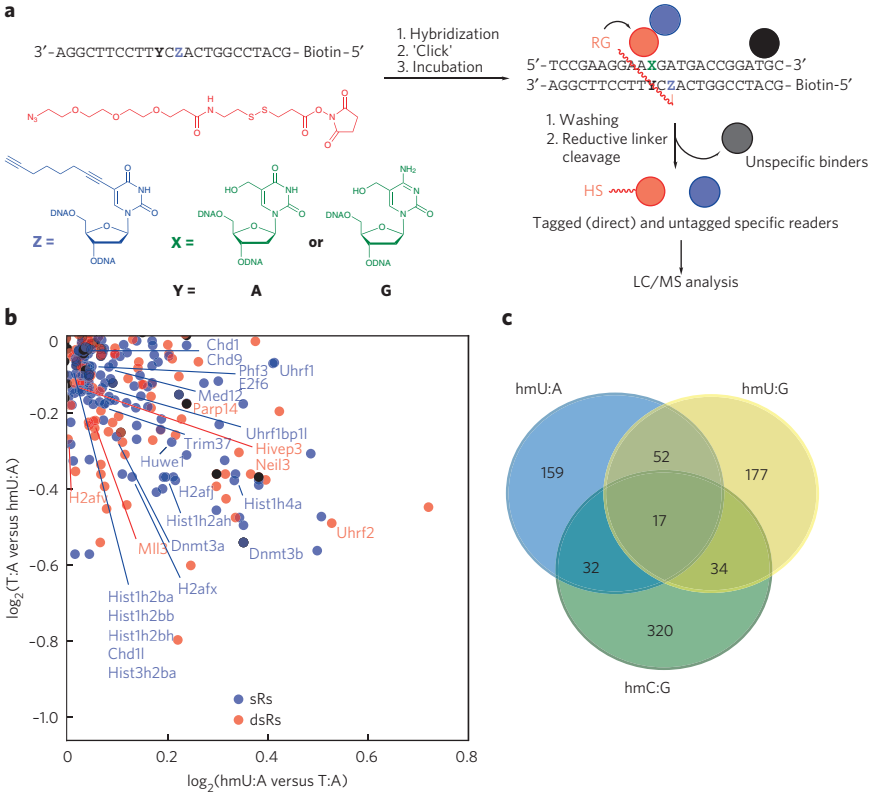


Figure 6 | Identification of hmU:A readers. (**a**) Workflow of pulldown experiments with hmU:A-containing oligonucleotides. DNA oligomers with the modified base Z containing an alkyne group for click reaction and biotin for the pulldown are hybridized with DNA oligomers containing hmU (X = hmU and Y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

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

As a first validation of the MS-identified readers, we focused on Uhrf1 and Uhrf2. Uhrf1 is a known mC and hmC binder[34], whereas we recently identified the homolog Uhrf2 as a specific binder of hmC[29]. The MS data for hmU obtained in this study showed an enrichment of Uhrf1 and Uhrf2. In addition, Uhrf2 was identified as a direct binder. To validate the MS findings, we overexpressed Uhrf1 or Uhrf2 together with Tet1cd in HEK-293T[29] and measured the levels of mC, hmC, fC, caC and also hmU. Indeed, we observed increased levels of hmU, proving a functional relation between hmU levels and Uhrf1 and Uhrf2 (**Supplementary Fig. 12**). As Uhrf2 is known to interact with Dnmt3a and Dnmt3b[35] and these two proteins are also identified as sRs of hmU, our initial data provide a consistent picture.

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

DISCUSSION

This work aimed to analyze the origin of hmU in mESCs. Although hmU as well as fU are both well-characterized oxidation products of thymidine[9,10], hmU in particular was recently discussed as an intermediate in active demethylation by deamination of hmC[11,12]. We used stable isotopologs of the main oxidation products hmC, fC, caC, hmU, fU and 8-oxo-G in combination with quantitative MS to show that the levels of hmU are strongly elevated in stem cells in comparison to somatic tissue. Notably, we did not observe equally elevated levels of fU and 8-oxo-G, both of which are formed by the reaction of DNA with ROS. This analysis suggests that hmU in stem cells is not only formed by ROS. The quantitative data allowed us to estimate that in WT stem cells, about 70–80% of the found hmU is not formed by ROS but is derived from a different process. Isotope tracing studies with labeled [$^{13}CD_3$]L-Met and [^{13}C,$^{15}N_2$] thymidine showed that the detected hmU originates exclusively from T. Previous studies already proposed a connection between Tet activity and hmU levels, but these studies did not address the origin of hmU, which led to the hypothesis that hmU could form by deamination of hmC[11,22]. To investigate the involvement of deamination[11–14], we studied *Tdg*$^{-/-}$ stem cells reconstituted with a catalytically inactive Tdg. In these cells, the putative deamination–repair complex[12] of Tdg and Aid could form, and we expected to see elevated hmU levels. Indeed, in this experiment, we were able to detect labeled hmU, showing that deamination of hmC does occur, but the levels were low. Most notably, special conditions were required to detect this deamination. In WT cells, all of the detected hmU is clearly derived from T oxidation. As such, the detected hmU is situated in a base pair with A (hmU:A). In conclusion, the majority of hmU in mESCs is produced independently of both ROS and deamination.

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

Received 15 December 2013; accepted 17 April 2014;
published online 18 May 2014

METHODS

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

References

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

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

Acknowledgments

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

Author contributions

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

Competing financial interests

The authors declare no competing financial interests.

Additional information

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

ONLINE METHODS

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

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

Oligonucleotide synthesis. Oligonucleotide synthesis was performed on an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions (DMT off) and acetyl-protected dC. Phosphoramidites (including the 5-hydroxymethyl-dU-CE phosphoramidite and the 5′-biotin phosphoramidite) and polystyrene carriers were obtained from Glen Research. 5-octadinyl-dU (Z) phosphoramidite was synthesized according to literature and characterized by ^{1}H-NMR, ^{13}C-NMR, ^{31}P-NMR and ESI-MS[43]. The crude oligonucleotide was cleaved from the resin and deprotected in 30% (v/v) ammonium hydroxide and 40% (v/v) methylamine (1:1) at 65 °C for 10 min and purified by preparative and analytical HPLC (Waters Breeze and Alliance, respectively). Separation was performed by applying a VP 250/10 Nucleosil 100-7 C18 column (flow: 5 ml/min) from Macherey-Nagel with a gradient of buffer A (0.1 M $NHEt_3OAc$ in water) and buffer B (0.1 M $NHEt_3OAc$ in 80% MeCN). DNA-containing fractions were characterized by MALDI-TOF (Bruker Autoflex II) and analytical HPLC, combined, and desalted by C18-Sep-Pak cartridges (Waters). For analytical HPLC, separation was performed by applying a CC 250/4 Nucleosil 120-3 C18 column from Macherey-Nagel (flow: 0.5 ml/min) with the aforementioned buffer system.

For the protein pulldown studies, an hmU-containing oligonucleotide was hybridized with a DNA strand modified with a 5′-biotin and a 5-octadienyl-dU. The sequences of the DNA strands are summarized in **Supplementary Table 3**. To 10 nmol of the dsDNA, 0.5 µl of a 200 mM solution of the cross-linking azide-PEG_3-S-S-NHS ester ($C_{18}H_{29}N_5O_8S_2$, Jena Bioscience, Jena, Germany) in DMSO was added. In a separate tube, CuBr was dissolved in a TBTA solution (DMSO/*t*BuOH 3:1, 100 mM) resulting in a 1:1 Cu(I):TBTA ratio. This solution was immediately added to the DNA/azide mixture. Furthermore, 45 µl of DMSO/*t*BuOH (3:1) were added, the mixture was shaken at 37 °C for 3 h, and the resulting product was purified by ethanol precipitation.

LC/MS analysis of DNA samples. Quantitative LC/UV-ESI-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. An improved method, based on earlier published work[29,42,44,45], was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature 50 °C, gas flow 15 l/min (N_2), nebulizer 30 psi, sheath gas heater 275 °C, sheath gas flow 11 l/min (N_2), capillary voltage 2,500 V in the positive ion mode, capillary voltage −2,250 V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 4** and **5**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 µm, 2.1 mm × 150 mm) at 30 °C using a gradient of water and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 ml/min: 0 → 5 min; 0 → 3.5% (v/v) MeCN; 5 → 6.9 min; 3.5 → 5% MeCN; 6.9 → 7.2 min; 5 → 80% MeCN; 7.2 → 10.5 min; 80% MeCN; 10.5 → 11.3 min; 80 → 0% MeCN; 11.3 → 13 min; 0% MeCN. The effluent up to 1.5 min and after 9 min was diverted to waste by a Valco valve. The autosampler was cooled to 10 °C. The injection volume was amounted to 29 µl. Calibration curves, method validation and data processing are in **Supplementary Note 2**. A complete compilation of LC/MS quantifications results see **Supplementary Note 3**.

DNA digestion. 5–25 µg of genomic DNA in 25 µl H_2O were digested as follows: An aqueous solution (7.5 µl) of 480 µM $ZnSO_4$, containing 42 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards (**Supplementary Note 2**) were added, and the mixture was incubated at 37 °C for 3 h. After addition of 7.5 µl of a 520 µM [Na]$_2$-EDTA solution, containing 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for another 3 h at 37 °C and then stored at −20 °C. Prior to LC/MS/MS analysis, samples with up to 15 µg DNA, for which the quantification of low amounts of caC was aspired, were filtered by using an AcroPrep Advance 96 filter plate 0.2 µm Supor (Pall Life Sciences). In contrast, samples with 15–25 µg DNA (isotope-tracing experiments) were filtered by using an AcroPrep Advance 96 filter plate 10K Omega (Pall Life Sciences).

Genomic DNA isolation. Tissues of female WT mice (C57-BL6/N) were dissected at postnatal day 90 and prepared as earlier described[19,21]. Genomic DNA was extracted using the Qiagen Blood and Cell Culture DNA Midi Kit except for mESC samples differentiated in the presence of growth factors (see below). Extraction was performed following the manufacturer's instructions for genomic DNA isolation from cell culture samples or tissue samples, respectively. All buffers until loading of the sample on Genomic-tip 100/G were additionally supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene. (BHT, 200 µM) and deferoxamine mesylate salt (desferal, 200 µM) as well as the deaminase inhibitor tetrahydrouridine (THU, 200 µM), according to published methods, to reduce background oxidation or deamination[18]. Elution buffer QF was supplemented with 200 µM BHT. Following elution, all steps were performed on ice. DNA was then precipitated with NaOAc (0.3 M final) and 0.7 volumes *i*PrOH. DNA pellets from cultured cells were washed twice with ice-cold 70% EtOH and resuspended in H_2O containing 20 µM BHT using a Qiagen TissueLyser (30 Hz, 2 min). DNA pellets from mouse tissues were resuspended in PBS buffer and additionally extracted with phenol/$CHCl_3$, precipitated, washed and resuspended as described above.

R1 mESC samples differentiated in the presence of growth factors or transfected with Smug1 esiRNAs were lysed directly in the plates with RLT buffer (Qiagen) supplemented with BHT and desferal as described above. DNA was isolated using the Zymo Quick gDNA Midi Kit according to the manufacturer's instruction, except that elution was repeated four times with 100 µl of elution buffer supplemented with BHT (200 µM). Eluted DNA was precipitated with 2 M ammonium acetate and two volumes of absolute ethanol and finally resuspended in H_2O containing 20 µM BHT. The flow-through from the spin columns was used to isolate RNA (see real-time PCR analysis).

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

doi:10.1038/nchembio.1532

thymidine in serum-primed mESCs 2i cultures of R1, cells were passaged twice (5 d) in the same serum-containing medium lacking 2i and simultaneously supplemented with 100 μM [^{13}C,$^{15}N_2$]T. For isotope tracing with heavy methionine in serum-primed mESCs, 2i cultures of R1 cells and mESC lines with targeted Tdg alleles were passaged twice (over 5 d) without 2i in L-Met-free DMEM (Life Technology) supplemented as above and with 0.2 mM of either [*methyl*-^{13}C,D_3]L-Met or natural L-Met. For mESC differentiation without growth factors, R1 cells and the C57Bl/6/129 mixed background cell line were first plated at 1×10^5 cells/cm^2 on gelatin-coated plates in N2B27 medium containing 1,000 U/ml LIF to favor attachment and initial survival[51]. After 12 h, the medium was replaced without addition of LIF (defined as time point 0 h). The medium was replaced once more at 24 h. For isotope tracing with [*methyl*-^{13}C,D_3]L-Met during mESC differentiation, R1 cells were cultured for two passages in L-Met-free N2B27 medium supplemented with LIF, 2i and 0.2 mM of either [*methyl*-^{13}C,D_3]L-Met or natural L-Met. Differentiation of R1 cells in the presence of FGF-2 and ActA was as described[26], with minor modifications. Briefly, mESCs were cultured in N2B27 medium containing 2i and 1,000 U/ml LIF for several passages and then seeded at 2.2×10^5 cells/cm^2 in N2B27 medium containing 1% KnockOut Serum Replacement (Life Technologies), 12 ng/ml FGF-2 (PeproTech) and 20 ng/ml ActA (ORF Genetics) on plates coated with a thin layer of Geltrex extracellular matrix preparation (Life Technologies). The medium was exchanged after 24 h.

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

Culture and transfection procedures for HEK-293T cells. All transfections were performed using jetPRIME transfection reagent (PEQLAB Biotechnologie GmbH) according to the manufacturer's instructions. HEK-293T cells were seeded 24 h before transfection at a density of 2.5×10^6 cells per 75 cm^2 flask and incubated in 10 ml of medium. The transfection solution (500 μl of jetPRIME buffer, a specific amount of plasmid DNA (**Supplementary Note 4**) and 20 μl of jetPRIME reagent) was added to the medium, and the cells were incubated for 48 h, with an additional medium exchange 24 h after transfection. When cotransfection of esiRNA was performed, a second transfection step (500 μl of jetPRIME buffer, 5 μg of esiRNA and 20 μl of jetPRIME reagent) was carried out 4 h after transfection of plasmid DNA. esiRNAs were purchased from Sigma (human TDG esiRNA EHU038971; human SMUG1 esiRNA EHU098861; human CDK5RAP1 esiRNAEHU079221). **Supplementary Note 4** summarizes the overexpression and knockdown procedures.

Isotope tracing with [^{13}C,$^{15}N_2$]thymidine or [^{13}C,D_3]L-Met in HEK-293T cells transfected with Tet1cd. 24 h before transfection, 2.5×10^6 cells were seeded in a 75-cm^2 flask containing 10 ml either of (for [^{13}C,$^{15}N_2$]thymidine) DMEM medium supplemented with 50 μM [^{13}C,$^{15}N_2$]T or (for [^{13}C,D_3]L-Met) DMEM medium lacking L-Met, L-cystine and pyruvate, which was supplemented with 10% dialyzed FBS, 2 mM [^{13}C,D_3]L-Met and 0.2 mM L-cystine. Transfection was performed as described above using labeled medium.

Real-time PCR analysis of mRNA expression. For analysis of Tet1 and Tet2 knockdown in mESC total RNA was prepared with Trizol (Invitrogen), cDNA synthesis was performed with Quantitect reverse transcription kit from Qiagen, and real-time PCR was performed with the Power Sybr Green PCR master mix from Applied Biosystems on an Applied Biosystems 7500 Fast system. Knockdown efficiencies relative to control samples transfected with SCR esiRNAs were 79% and 70% for Tet1 and Tet2, respectively. The primers used to estimate them are listed in **Supplementary Note 4**. For analysis of Smug1 knockdown in mESCs and EpiLC differentiation samples, total RNA was prepared with RNeasy spin columns (Qiagen), followed by DNase treatment using TURBO DNA-free (Ambion, Life Technologies); cDNA synthesis was carried out using iScript cDNA Synthesis kit (Bio-Rad); real-time PCR was performed with SsoFast EvaGreen Supermix (Bio-Rad). Smug1 knockdown efficiency relative to control samples transfected with esiRNAs targeting GFP was estimated to be 60%. Quantification of Tet, Dnmt, Tdg and Smug1 transcripts during EpiLC differentiation and Smug1 knockdown samples was performed using the primers listed in **Supplementary Note 4**. Expression levels were quantified with respect to the housekeeping gene *Gapdh* and normalized to time point 0 h.

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

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

Pulldown assay. For the pulldown assay, 250 μg (50 μl) of the crude nuclear protein extracts were filled up to 500 μl with 50 mM TEAB and 1 mM $MgCl_2$. The binding conditions were 45 mM TEAB, 1.1 mM $MgCl_2$, 2 mM HEPES, 42 mM NaCl and 20 μM EDTA. Complete Protease Inhibitor Cocktail Tablets were used from Roche Diagnostics (Indianapolis, IN, USA). The DNA oligomers with DNA-protein cross-linker were dissolved in neat DMSO, and 1 nmol was added to the protein lysate and incubated for 20 min at room temperature. Streptavidin-coated magnetic particles (Roche Diagnostics, Indianapolis, IN, USA) were washed three times with binding buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 7.4) before 200 μl of the bead slurry (equal to 2 mg beads) were added to the sample. Following 2 h incubation at room temperature under constant rotating, the beads were washed three times with 50 mM TEAB and 1 mM $MgCl_2$. The beads were reconstituted in 50 mM TEAB and 1 mM $MgCl_2$. Disulfide bonds of the cross linker were cleaved and alkylated in the process of enzymatic digestion, and the magnetic particles were removed before adding trypsin (described below).

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

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

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

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

42. Schiesser, S. *et al.* Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
43. Gierlich, J. *et al.* Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **8**, 3639–3642 (2006).
44. Cao, H. & Wang, Y. Collisionally activated dissociation of protonated 2′-deoxycytidine, 2′-deoxyuridine, and their oxidatively damaged derivatives. *J. Am. Soc. Mass Spectrom.* **17**, 1335–1341 (2006).
45. Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).
46. Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
47. Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
48. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90**, 8424–8428 (1993).
49. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
50. Montanez, E. *et al.* Kindlin-2 controls bidirectional signaling of integrins. *Genes Dev.* **22**, 1325–1330 (2008).
51. Ying, Q.L. & Smith, A.G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
52. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
53. Kittler, R., Heninger, A.K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779–784 (2005).
54. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).

doi:10.1038/nchembio.1532

9 Outlook

The investigation of modified tRNA nucleosides in higher eukaryotes presented in this work reveals that modification levels differ between mammalian tissues. This difference is proposed to depend on variations in protein synthesis needs. Additionally, *S. cerevisiae* was found to increasingly modify its tRNA population upon entry into stationary phase, a feature which is suggested to facilitate protein translation during quiescence and to ensure rapid start of the translational machinery upon exit from stationary phase. These findings suggest a regulation of the tRNA modification levels to ensure a dynamic adaptation to the translational requirements of the cell. As a next step it would be interesting to further elucidate how this regulation is established, whether it is dependent on the levels of tRNA modifying proteins or whether it derives from selective maintenance of a pool of hypomodified tRNA molecules which can be targeted for modification depending on protein synthesis needs. Furthermore, investigation using double-labelling of the tRNA will allow to elucidate whether tRNA modification turnover is coupled to overall tRNA turnover and this result will help to accurately determine the contribution of tRNA degradation to the increase in RNA modifications observed in *S. cerevisiae*. In addition, it would be interesting to determine at which point of development the tissue-dependent distribution of tRNA modifications is established, for example by monitoring the tRNA modification levels during differentiation of mouse embryonic stem cells to different cell lineages.

Further studies should additionally focus on the modifications discovered in other RNA species and their role and impact on RNA function. Quantification of these modifications and their variation in response to various environmental cues will certainly prove essential for their further characterization. Particularly helpful will also be the development of RNA modification mapping/sequencing methods which will then allow to specifically locate modified nucleosides in the various RNA molecules.

The involvement of the oxidized cytosines hmC, fC and caC as epigenetic markers or as intermediates in different pathways of active demethylation is still the focus of ongoing investigation. Results presented in this work showed that fC and caC might be involved in replication-coupled dilution (i.e. erasure) of the methylation pattern, possibly complementing the previously discovered TDG-dependent mechanism of fC and caC removal. Further studies should continue the search for a C-C cleavage-dependent demethylation pathway, which to date remains elusive. In particular, it would be beneficial to perform these studies in living cells, testing a range of cells as well as various differentiation time points. Furthermore, the generation of plasmid DNA using labelled cytosine triphosphate derivatives via the Polymerase Chain Reaction could be exploited to produce labelled substrates for a putative demethylase which can be transfected into cells. The enhanced stability of these DNA constructs would allow for the unambiguous identification of possible demethylation products.

Furthermore, the role of fC and caC as epigenetic markers with a specific function in modulation of gene expression is still unclear. Future studies should address this issue by establishing the influence of these modifications on the rate of gene transcription, as well as on recruitment of specific chromatin remodelling proteins.

The discovery of hmU as a product of Tet1 enzymatic activity suggests that hmU might play a more complex role compared to a simple oxidative lesion. It would therefore be interesting to characterize the specific binders of hmU as determined by proteome analysis. Further information could be gained by sequencing of this base in the mammalian genome. This might reveal specific loci where hmU is enriched, and therefore contribute to our understanding of its function as a product of Tet1 enzymatic activity.

The observation that hmU is largely generated by Tet1-dependent oxidation of T rather than by deamination of hmC also raises the question of whether deamination of hmC is at all involved in active demethylation. Further studies should monitor whether deamination of hmC can be detected during specific stages of cell differentiation and whether it is restricted to specific cell types. It would in fact be interesting to clarify the contribution of this pathway to the demethylation process and how it is eventually regulated. Similarly, the high sensitivity of UPLC-MS/MS-based quantification methods should be exploited to address the process of deamination of mC to generate T, which again would provide important insights into the process of active demethylation.

10 Literature

[1] a) P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 98-99; b) S. Dunin-Horkawicz, A. Czerwoniec, M. J. Gajda, M. Feder, H. Grosjean, J. M. Bujnicki, *Nucleic Acids Res.* **2006**, *34*, D145-D149.

[2] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, D195-D201.

[3] S. H. Kim, G. J. Quigley, F. L. Suddath, McPherso.A, D. Sneden, J. J. Kim, Weinzier.J, A. Rich, *Science* **1973**, *179*, 285-288.

[4] a) A. Alexandrov, I. Chernyakov, W. Gu, S. L. Hiley, T. R. Hughes, E. J. Grayhack, E. M. Phizicky, *Mol. Cell* **2006**, *21*, 87-96; b) Y. Motorin, M. Helm, *Biochemistry* **2010**, *49*, 4934-4944; c) E. M. Phizicky, J. D. Alfonzo, *FEBS Lett.* **2010**, *584*, 265-271.

[5] H. Grosjean, M. Sprinzl, S. Steinberg, *Biochimie* **1995**, *77*, 139-141.

[6] M. Yarus, *Science* **1982**, *218*, 646-652.

[7] H. J. Shi, P. B. Moore, *RNA* **2000**, *6*, 1091-1105.

[8] M. Lovmar, M. Ehrenberg, *Biochimie* **2006**, *88*, 951-961.

[9] C. G. Kurland, *Annu. Rev. Genet.* **1992**, *26*, 29-50.

[10] J. M. Ogle, D. E. Brodersen, W. M. Clemons, Jr., M. J. Tarry, A. P. Carter, V. Ramakrishnan, *Science* **2001**, *292*, 897-902.

[11] a) P. F. Agris, *Nucleic Acids Res.* **2004**, *32*, 223-238; b) P. F. Agris, *EMBO Rep.* **2008**, *9*, 629-635; c) E. M. Gustilo, F. A. P. Vendeix, P. F. Agris, *Curr. Opin. Microbiol.* **2008**, *11*, 134-140.

[12] P. F. Agris, F. A. P. Vendeix, W. D. Graham, *J. Mol. Biol.* **2007**, *366*, 1-13.

[13] M. Sprinzl, T. Hartmann, J. Weber, J. Blank, R. Zeidler, *Nucleic Acids Res.* **1989**, *17 Suppl*, r1-172.

[14] a) K. Takai, S. Yokoyama, *Nucleic Acids Res.* **2003**, *31*, 6383-6391; b) A. Weixlbaumer, F. V. Murphy, A. Dziergowska, A. Malkiewicz, F. A. P. Vendeix, P. F. Agris, V. Ramakrishnan, *Nat. Struct. Mol. Biol.* **2007**, *14*, 498-502.

[15] a) H. Sierzputowskagracz, E. Sochacka, A. Malkiewicz, K. Kuo, C. W. Gehrke, P. F. Agris, *J. Am. Chem. Soc.* **1987**, *109*, 7171-7177; b) S. Yokoyama, T. Watanabe, K. Murao, H. Ishikura, Z. Yamaizumi, S. Nishimura, T. Miyazawa, *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4905-4909.

[16] a) S. S. Ashraf, E. Sochacka, R. Cain, R. Guenther, A. Malkiewicz, P. F. Agris, *RNA* **1999**, *5*, 188-194; b) C. Yarian, M. Marszalek, E. Sochacka, A. Malkiewicz, R. Guenther, A. Miskiewicz, P. F. Agris, *Biochemistry* **2000**, *39*, 13390-13395; c) C. Yarian, H. Townsend, W. Czestkowski, E. Sochacka, A. J. Malkiewicz, R. Guenther, A. Miskiewicz, P. F. Agris, *J. Biol. Chem.* **2002**, *277*, 16391-16395.

[17] F. V. Murphy, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186-1191.

[18] a) J. Urbonavicius, Q. Qian, J. M. B. Durand, T. G. Hagervall, G. R. Bjork, *EMBO J.* **2001**, *20*, 4863-4873; b) G. R. Bjork, P. M. Wikstrom, A. S. Bystrom, *Science* **1989**, *244*, 986-989; c) G. R. Bjork, J. M. B. Durand, T. G. Hagervall, R. Leipuviene, H. K. Lundgren, K. Nilsson, P. Chen, Q. Qian, J. Urbonavicius, *FEBS Lett.* **1999**, *452*, 47-51.

[19] a) J. W. Stuart, Z. Gdaniec, R. Guenther, M. Marszalek, E. Sochacka, A. Malkiewicz, P. F. Agris, *Biochemistry* **2000**, *39*, 13396-13404; b) V. Dao, R. Guenther, A. Malkiewicz, B. Nawrot, E. Sochacka, A. Kraszewski, J. Jankowska, K. Everett, P. F. Agris, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2125-2129.

[20] L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, *Nat. Struct. Mol. Biol.* **2010**, *17*, 555-560.

[21] J. Cabello-Villegas, M. E. Winkler, E. P. Nikonowicz, *J. Mol. Biol.* **2002**, *319*, 1015-1034.

[22] a) G. Kawai, Y. Yamamoto, T. Kamimura, T. Masegi, M. Sekine, T. Hata, T. Iimori, T. Watanabe, T. Miyazawa, S. Yokoyama, *Biochemistry* **1992**, *31*, 1040-1046; b) S. Yokoyama, K. Watanabe, T. Miyazawa, *Adv. Biophys.* **1987**, *23*, 115-147.

[23] V. Perret, A. Garcia, J. Puglisi, H. Grosjean, J. P. Ebel, C. Florentz, R. Giegé, *Biochimie* **1990**, *72*, 735-743.
[24] J. R. Sampson, O. C. Uhlenbeck, *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1033-1037.
[25] M. Helm, R. Giegé, C. Florentz, *Biochemistry* **1999**, *38*, 13338-13346.
[26] a) J. Anderson, L. Phan, A. G. Hinnebusch, *Proc. Natl Acad. Sci. U. S. A.* **2000**, *97*, 5173-5178; b) S. Kadaba, A. Krueger, T. Trice, A. M. Krecic, A. G. Hinnebusch, J. Anderson, *Genes Dev.* **2004**, *18*, 1227-1240.
[27] L. Kotelawala, E. J. Grayhack, E. M. Phizicky, *RNA* **2008**, *14*, 158-169.
[28] T. Carell, C. Brandmayr, A. Hienzsch, M. Muller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131.
[29] J. Pütz, C. Florentz, F. Benseler, R. Giege, *Nat. Struct. Biol.* **1994**, *1*, 580-582.
[30] K. Nakanishi, L. Bonnefond, S. Kimura, T. Suzuki, R. Ishitani, O. Nureki, *Nature* **2009**, *461*, 1144-1148.
[31] a) R. Basavappa, P. B. Sigler, *EMBO J.* **1991**, *10*, 3105-3111; b) S. Kiesewetter, G. Ott, M. Sprinzl, *Nucleic Acids Res.* **1990**, *18*, 4677-4681.
[32] M. Thompson, R. A. Haeusler, P. D. Good, D. R. Engelke, *Science* **2003**, *302*, 1399-1401.
[33] E. Bertrand, F. Houser-Scott, A. Kendall, R. H. Singer, D. R. Engelke, *Genes Dev.* **1998**, *12*, 2463-2468.
[34] a) A. K. Hopper, D. A. Pai, D. R. Engelke, *FEBS Lett.* **2010**, *584*, 310-317; b) T. Ohira, T. Suzuki, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10502-10507.
[35] K. Nishikura, E. M. Derobertis, *J. Mol. Biol.* **1981**, *145*, 405-420.
[36] H. Grosjean, Z. SzweykowskaKulinska, Y. Motorin, F. Fasiolo, G. Simos, *Biochimie* **1997**, *79*, 293-302.
[37] H.-Q. Jiang, Y. Motorin, Y.-X. Jin, H. Grosjean, *Nucleic Acids Res.* **1997**, *25*, 2694-2701.
[38] T. Yoshihisa, K. Yunoki-Esaki, C. Ohshima, N. Tanaka, T. Endo, *Mol. Biol. Cell* **2003**, *14*, 3266-3279.
[39] a) H. H. Shaheen, A. K. Hopper, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11290-11295; b) A. Takano, T. Endo, T. Yoshihisa, *Science* **2005**, *309*, 140-142.
[40] A. K. Hopper, H. H. Shaheen, *Trends Cell Biol.* **2008**, *18*, 98-104.
[41] J. M. Wohlgamuth-Benedum, M. A. T. Rubio, Z. Paris, S. J. Long, P. Poliak, J. Lukes, J. D. Alfonzo, *J. Biol. Chem.* **2009**, *284*, 23947-23953.
[42] a) D. M. Thompson, C. Lu, P. J. Green, R. Parker, *RNA* **2008**, *14*, 2095-2103; b) S. Yamasaki, P. Ivanov, G. F. Hu, P. Anderson, *J. Cell Biol.* **2009**, *185*, 35-42.
[43] D. M. Thompson, R. Parker, *Cell* **2009**, *138*, 215-219.
[44] a) C. T. Y. Chan, M. Dyavaiah, M. S. DeMott, K. Taghizadeh, P. C. Dedon, T. J. Begley, *PLoS Genet.* **2010**, *6*, e1001247; b) N. Netzer, J. M. Goodenbour, A. David, K. A. Dittmar, R. B. Jones, J. R. Schneider, D. Boone, E. M. Eves, M. R. Rosner, J. S. Gibbs, A. Embry, B. Dolan, S. Das, H. D. Hickman, P. Berglund, J. R. Bennink, J. W. Yewdell, T. Pan, *Nature* **2009**, *462*, 522-526.
[45] C. T. Chan, Y. L. Pang, W. Deng, I. R. Babu, M. Dyavaiah, T. J. Begley, P. C. Dedon, *Nat. Commun.* **2012**, *3*, 937.
[46] U. Begley, M. Dyavaiah, A. Patil, J. P. Rooney, D. DiRenzo, C. M. Young, D. S. Conklin, R. S. Zitomer, T. J. Begley, *Mol. Cell* **2007**, *28*, 860-870.
[47] M. Frenkel-Morgenstern, T. Danon, T. Christian, T. Igarashi, L. Cohen, Y.-M. Hou, L. J. Jensen, *Mol. Syst. Biol.* **2012**, *8*.
[48] E. M. Novoa, L. Ribas de Pouplana, *Trends Genet.* **2012**, *28*, 574-581.
[49] K. Watanabe, M. Shinma, T. Oshima, S. Nishimura, *Biochem. Biophys. Res. Commun.* **1976**, *72*, 1137-1144.
[50] a) J. A. Kowalak, J. J. Dalluge, J. A. McCloskey, K. O. Stetter, *Biochemistry* **1994**, *33*, 7869-7876; b) L. Droogmans, M. Roovers, J. M. Bujnicki, C. Tricot, T. Hartsch, V. Stalon, H. Grosjean, *Nucleic Acids Res.* **2003**, *31*, 2148-2156.

[51] M. Roovers, J. Wouters, J. M. Bujnicki, C. Tricot, V. Stalon, H. Grosjean, L. Droogmans, *Nucleic Acids Res.* **2004**, *32*, 465-476.

[52] G. R. Bjork, K. Jacobsson, K. Nilsson, M. J. O. Johansson, A. S. Bystrom, O. P. Persson, *EMBO J.* **2001**, *20*, 231-239.

[53] E. M. Novoa, M. Pavon-Eternod, T. Pan, L. Ribas de Pouplana, *Cell* **2012**, *149*, 202-213.

[54] W. Reik, *Nature* **2007**, *447*, 425-432.

[55] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251-260.

[56] a) K. Luger, J. C. Hansen, *Curr. Opin. Struct. Biol.* **2005**, *15*, 188-196; b) G. Felsenfeld, M. Groudine, *Nature* **2003**, *421*, 448-453.

[57] P. Ball, *Nature* **2003**, *421*, 421-422.

[58] a) T. Kouzarides, *Cell* **2007**, *128*, 693-705; b) B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41-45.

[59] G. E. Zentner, S. Henikoff, *Nat. Struct. Mol. Biol.* **2013**, *20*, 259-266.

[60] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, *Nat. Struct. Mol. Biol.* **2007**, *14*, 1025-1040.

[61] S. K. Ooi, C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S. P. Lin, C. D. Allis, X. Cheng, T. H. Bestor, *Nature* **2007**, *448*, 714-717.

[62] a) J. Wang, S. Hevi, J. K. Kurash, H. Lei, F. Gay, J. Bajko, H. Su, W. Sun, H. Chang, G. Xu, F. Gaudet, E. Li, T. Chen, *Nat. Genet.* **2009**, *41*, 125-129; b) P. O. Esteve, H. G. Chin, J. Benner, G. R. Feehery, M. Samaranayake, G. A. Horwitz, S. E. Jacobsen, S. Pradhan, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5076-5081.

[63] F. Fuks, P. J. Hurd, D. Wolf, X. Nan, A. P. Bird, T. Kouzarides, *J. Biol. Chem.* **2003**, *278*, 4035-4040.

[64] B. Li, M. Carey, J. L. Workman, *Cell* **2007**, *128*, 707-719.

[65] B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, E. S. Lander, *Cell* **2006**, *125*, 315-326.

[66] C. Coulondre, J. H. Miller, P. J. Farabaugh, W. Gilbert, *Nature* **1978**, *274*, 775-780.

[67] R. S. Illingworth, A. P. Bird, *FEBS Lett.* **2009**, *583*, 1713-1720.

[68] A. M. Deaton, A. Bird, *Genes Dev.* **2011**, *25*, 1010-1022.

[69] R. Straussman, D. Nejman, D. Roberts, I. Steinfeld, B. Blum, N. Benvenisty, I. Simon, Z. Yakhini, H. Cedar, *Nat. Struct. Mol. Biol.* **2009**, *16*, 564-571.

[70] A. Kuroda, T. A. Rauch, I. Todorov, H. T. Ku, I. H. Al-Abdullah, F. Kandeel, Y. Mullen, G. P. Pfeifer, K. Ferreri, *PLoS One* **2009**, *4*.

[71] L. Lopez-Serra, M. Esteller, *Br. J. Cancer* **2008**, *98*, 1881-1885.

[72] P. A. Jones, *Nat. Rev. Genet.* **2012**, *13*, 484-492.

[73] E. Li, C. Beard, R. Jaenisch, *Nature* **1993**, *366*, 362-365.

[74] W. Reik, A. Lewis, *Nat. Rev. Genet.* **2005**, *6*, 403-410.

[75] A. Hellman, A. Chess, *Science* **2007**, *315*, 1141-1143.

[76] A. Doi, I.-H. Park, B. Wen, P. Murakami, M. J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G. Q. Daley, A. P. Feinberg, *Nat. Genet.* **2009**, *41*, 1350-U1123.

[77] a) A. M. Lindroth, X. F. Cao, J. P. Jackson, D. Zilberman, C. M. McCallum, S. Henikoff, S. E. Jacobsen, *Science* **2001**, *292*, 2077-2080; b) R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315-322.

[78] a) M. Esteller, *Nat. Rev. Genet.* **2007**, *8*, 286-298; b) C. P. Walsh, J. R. Chaillet, T. H. Bestor, *Nat. Genet.* **1998**, *20*, 116-117.

[79] a) A. Meissner, *Nat. Biotechnol.* **2010**, *28*, 1079-1088; b) W. Reik, W. Dean, J. Walter, *Science* **2001**, *293*, 1089-1093.

[80] A. Portela, M. Esteller, *Nat. Biotechnol.* **2010**, *28*, 1057-1068.

[81] S. E. Goelz, B. Vogelstein, S. R. Hamilton, A. P. Feinberg, *Science* **1985**, *228*, 187-190.
[82] M. Esteller, *Hum. Mol. Genet.* **2007**, *16*, R50-R59.
[83] T. K. Kelly, D. D. De Carvalho, P. A. Jones, *Nat. Biotechnol.* **2010**, *28*, 1069-1078.
[84] a) U. Aapola, R. Lyle, K. Krohn, S. E. Antonarakis, P. Peterson, *Cytogenet. Cell Genet.* **2001**, *92*, 122-126; b) H. Gowher, K. Liebert, A. Hermann, G. L. Xu, A. Jeltsch, *J. Biol. Chem.* **2005**, *280*, 13341-13348; c) K. Hata, M. Okano, H. Lei, E. Li, *Development* **2002**, *129*, 1983-1993; d) Y.-G. Hu, R. Hirasawa, J.-L. Hu, K. Hata, C.-L. Li, Y. Jin, T. Chen, E. Li, M. Rigolet, E. Viegas-Pequignot, H. Sasaki, G.-L. Xu, *Hum. Mol. Genet.* **2008**, *17*, 2654-2664; e) D. Bourc'his, G. L. Xu, C. S. Lin, B. Bollman, T. H. Bestor, *Science* **2001**, *294*, 2536-2539.
[85] A. Hermann, S. Schmitt, A. Jeltsch, *J. Biol. Chem.* **2003**, *278*, 31717-31721.
[86] M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C. L. Hsieh, X. Y. Zhang, K. G. Golic, S. E. Jacobsen, T. H. Bestor, *Science* **2006**, *311*, 395-398.
[87] S. Klimasauskas, S. Kumar, R. J. Roberts, X. D. Cheng, *Cell* **1994**, *76*, 357-369.
[88] a) M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257; b) M. Okano, S. P. Xie, E. Li, *Nat. Genet.* **1998**, *19*, 219-220.
[89] M. Kaneda, M. Okano, K. Hata, T. Sado, N. Tsujimoto, E. Li, H. Sasaki, *Nature* **2004**, *429*, 900-903.
[90] H. Leonhardt, A. W. Page, H. U. Weier, T. H. Bestor, *Cell* **1992**, *71*, 865-873.
[91] T. H. Bestor, V. M. Ingram, *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80*, 5559-5563.
[92] L. S. H. Chuang, H. I. Ian, T. W. Koh, H. H. Ng, G. L. Xu, B. F. L. Li, *Science* **1997**, *277*, 1996-2000.
[93] C. Frauer, A. Rottach, D. Meilinger, S. Bultmann, K. Fellinger, S. Hasenoeder, M. Wang, W. Qin, J. Soeding, F. Spada, H. Leonhardt, *PLoS One* **2011**, *6*.
[94] J. Song, O. Rechkoblit, T. H. Bestor, D. J. Patel, *Science* **2011**, *331*, 1036-1040.
[95] a) H. Lei, S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch, E. Li, *Development* **1996**, *122*, 3195-3205; b) E. Li, T. H. Bestor, R. Jaenisch, *Cell* **1992**, *69*, 915-926.
[96] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *Chembiochem* **2011**, *12*, 206-222.
[97] A. Jeltsch, *Chembiochem* **2002**, *3*, 274-293.
[98] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930-935.
[99] N. W. Penn, Bojanows.K, R. Yura, R. Suwalski, C. Oriley, *Biochem. J.* **1972**, *126*, 781-&.
[100] a) Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303-1307; b) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303; c) T. Pfaffeneder, B. Hackner, M. Truss, M. Muenzel, M. Mueller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012.
[101] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129-U1151.
[102] a) W. A. Pastor, L. Aravind, A. Rao, *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 341-356; b) L. Tan, Y. G. Shi, *Development* **2012**, *139*, 1895-1902.
[103] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, *Cell Cycle* **2009**, *8*, 1698-1710.
[104] a) Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Mol. Cell* **2011**, *42*, 451-464; b) H. Zhang, X. Zhang, E. Clark, M. Mulcahey, S. Huang, Y. G. Shi, *Cell Res.* **2010**, *20*, 1390-1393.
[105] Y. Xu, C. Xu, A. Kato, W. Tempel, J. G. Abreu, C. Bian, Y. Hu, D. Hu, B. Zhao, T. Cerovina, J. Diao, F. Wu, H. H. He, Q. Cui, E. Clark, C. Ma, A. Barbara, G. J. Veenstra, G. Xu, U. B. Kaiser, X. S. Liu, S. P. Sugrue, X. He, J. Min, Y. Kato, Y. G. Shi, *Cell* **2012**, *151*, 1200-1213.
[106] G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature.* **2011**, *473*, 398-402.

[107] K. P. Koh, A. Yabuuchi, S. Rao, Y. Huang, K. Cunniff, J. Nardone, A. Laiho, M. Tahiliani, C. A. Sommer, G. Mostoslavsky, R. Lahesmaa, S. H. Orkin, S. J. Rodig, G. Q. Daley, A. Rao, *Cell Stem Cell* **2011**, *8*, 200-213.

[108] M. M. Dawlaty, K. Ganz, B. E. Powell, Y.-C. Hu, S. Markoulaki, A. W. Cheng, Q. Gao, J. Kim, S.-W. Choi, D. C. Page, R. Jaenisch, *Cell Stem Cell* **2011**, *9*, 166-175.

[109] M. Munzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460-6468.

[110] a) W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394-397; b) K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343-348.

[111] H. Wu, A. C. D'Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *Nature* **2011**, *473*, 389-U578.

[112] a) M. Ko, H. S. Bandukwala, J. An, E. D. Lamperti, E. C. Thompson, R. Hastie, A. Tsangaratou, K. Rajewsky, S. B. Koralov, A. Rao, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 14566-14571; b) Z. Li, X. Cai, C.-L. Cai, J. Wang, W. Zhang, B. E. Petersen, F.-C. Yang, M. Xu, *Blood* **2011**, *118*, 4509-4518.

[113] M. Ko, Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, A. Rao, *Nature* **2010**, *468*, 839-843.

[114] a) T. P. Gu, F. Guo, H. Yang, H. P. Wu, G. F. Xu, W. Liu, Z. G. Xie, L. Shi, X. He, S. G. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabo, G. P. Pfeifer, J. Li, G. L. Xu, *Nature* **2011**, *477*, 606-610; b) K. Iqbal, S. G. Jin, G. P. Pfeifer, P. E. Szabo, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3642-3647; c) M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, *Nat. Commun.* **2011**, *2*; d) A. Inoue, L. Shen, Q. Dai, C. He, Y. Zhang, *Cell Res.* **2011**, *21*, 1670-1676.

[115] M. A. Hahn, R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S.-G. Jin, Y. Jiang, G. P. Pfeifer, Q. Lu, *Cell Rep.* **2013**, *3*, 291-300.

[116] a) Q. Chen, Y. Chen, C. Bian, R. Fujiki, X. Yu, *Nature* **2013**, *493*, 561-564; b) R. Deplus, B. Delatte, M. K. Schwinn, M. Defrance, J. Mendez, N. Murphy, M. A. Dawson, M. Volkmar, P. Putmans, E. Calonne, A. H. Shih, R. L. Levine, O. Bernard, T. Mercher, E. Solary, M. Urh, D. L. Daniels, F. Fuks, *EMBO J.* **2013**, *32*, 645-655; c) P. Vella, A. Scelfo, S. Jammula, F. Chiacchiera, K. Williams, A. Cuomo, A. Roberto, J. Christensen, T. Bonaldi, K. Helin, D. Pasini, *Mol. Cell* **2013**, *49*, 645-656.

[117] D. Globisch, M. Muenzel, M. Mueller, S. Michalakis, M. Wagner, S. Koch, T. Brueckl, M. Biel, T. Carell, *PLoS One* **2010**, *5*.

[118] a) C. X. Song, C. Yi, C. He, *Nat. Biotechnol.* **2012**, *30*, 1107-1116; b) M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368-1380; c) M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934-937; d) C.-X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678-691.

[119] a) H. Stroud, S. Feng, S. M. Kinney, S. Pradhan, S. E. Jacobsen, *Genome Biol.* **2011**, *12*; b) H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *Genes Dev.* **2011**, *25*, 679-684.

[120] E. A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69.

[121] L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H.-L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692-706.

[122] a) M. R. Branco, G. Ficz, W. Reik, *Nat. Rev. Genet.* **2012**, *13*, 7-13; b) S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607-620.

[123] H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang, X. Cheng, *Nucleic Acids Res.* **2012**, *40*, 4841-4849.

[124] a) J. A. Hackett, R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down, M. A. Surani, *Science* **2013**, *339*, 448-452; b) J. J. Vincent, Y. Huang, P. Y. Chen, S. Feng, J. H. Calvopina, K. Nee, S. A. Lee, T. Le, A. J. Yoon, K. Faull, G. Fan, A. Rao, S. E. Jacobsen, M. Pellegrini, A. T. Clark, *Cell Stem Cell* **2013**.
[125] J.-K. Zhu, in *Ann. Rev. Genet., Vol. 43*, **2009**, pp. 143-166.
[126] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334-35338.
[127] a) S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67-79; b) J. U. Guo, Y. Su, C. Zhong, G. L. Ming, H. Song, *Cell* **2011**, *145*, 423-434.
[128] a) K. Rai, I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, B. R. Cairns, *Cell* **2008**, *135*, 1201-1212; b) C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101-U1126.
[129] C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, R. M. Kohli, *Nat. Chem. Biol.* **2012**, *8*, 751-758.
[130] a) P. O. Falnes, R. F. Johansen, E. Seeberg, *Nature* **2002**, *419*, 178-182; b) S. C. Trewick, T. F. Henshaw, R. P. Hausinger, T. Lindahl, B. Sedgwick, *Nature* **2002**, *419*, 174-178.
[131] J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes, A. A. Axhemi, *Biochim. Biophys. Acta* **2005**, *1723*, 256-264.
[132] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Muller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516-6520.
[133] C. C. Chen, K. Y. Wang, C. K. Shen, *J. Biol. Chem.* **2012**, *287*, 33116-33121.
[134] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Muenzel, M. Wagner, M. Mueller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Mueller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159.
[135] a) H. H. Ng, P. Jeppesen, A. Bird, *Mol. Cell. Biol.* **2000**, *20*, 1394-1406; b) H. H. Ng, Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg, A. Bird, *Nat. Genet.* **1999**, *23*, 58-61; c) V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird, L. C. Sowers, *Nucleic Acids Res.* **2004**, *32*, 4100-4108.
[136] M. W. Kellinger, C.-X. Song, J. Chong, X.-Y. Lu, C. He, D. Wang, *Nat. Struct. Mol. Biol.* **2012**, *19*, 831-833.
[137] A. Thalhammer, A. S. Hansen, A. H. El-Sagheer, T. Brown, C. J. Schofield, *Chem. Commun.* **2011**, *47*, 5325-5327.
[138] A. Hienzsch, Doctoral thesis, Ludwig-Maximilians-Universität München Munich, Germany, **2012**.
[139] M. A. Preston, S. D'Silva, Y. Kon, E. M. Phizicky, *RNA* **2013**, *19*, 243-256.
[140] I. Sethy, R. D. Moir, M. Librizzi, I. M. Willis, *J. Biol. Chem.* **1995**, *270*, 28463-28470.
[141] M. Ciesla, J. Towpik, D. Graczyk, D. Oficjalska-Pham, O. Harismendy, A. Suleau, K. Balicki, C. Conesa, O. Lefebvre, M. Boguta, *Mol. Cell. Biol.* **2007**, *27*, 7693-7702.
[142] T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2009**, *48*, 7932-7934.
[143] a) S. Y. Chan, D. R. Appling, *J. Biol. Chem.* **2003**, *278*, 43051-43059; b) D. Thomas, A. Becker, Y. Surdin-Kerjan, *J. Biol. Chem.* **2000**, *275*, 40718-40724.
[144] J. R. Warner, S. A. Morgan, R. W. Shulman, *J. Bacteriol.* **1976**, *125*, 887-891.
[145] E. M. Phizicky, A. K. Hopper, *Genes Dev.* **2010**, *24*, 1832-1860.
[146] M. Nwagwu, M. Nana, *J. Embryol. Exp. Morphol.* **1980**, *56*, 253-267.
[147] U. Karnahl, C. Wasternack, *Int. J. Biochem.* **1992**, *24*, 493-497.
[148] K. C. Perry, D. J. Cove, *Physiol. Plant.* **1986**, *67*, 680-684.
[149] a) I. Clark, J. W. Mackenzie, J. R. McCoy, W. Lin, *Recent Res. Cancer* **1983**, *84*, 388-400; b) S. Weissman, M. Lewis, M. Karon, A. Z. Eisen, *J. Lab. Clin. Med.* **1962**, *60*, 40-&.

[150] A. Czerwoniec, S. Dunin-Horkawicz, E. Purta, K. H. Kaminska, J. M. Kasprzak, J. M. Bujnicki, H. Grosjean, K. Rother, *Nucleic Acids Res.* **2009**, *37*, D118-121.
[151] D. M. Thompson, R. Parker, *J. Cell Biol.* **2009**, *185*, 43-50.
[152] C. M. Chan, C. Zhou, R. H. Huang, *Science* **2009**, *326*, 247-247.
[153] R. H. Huang, *Biochemistry* **2012**, *51*, 4087-4095.
[154] M. L. Whitney, R. L. Hurto, H. H. Shaheen, A. K. Hopper, *Mol. Biol. Cell* **2007**, *18*, 2678-2686.
[155] a) J. V. Gray, G. A. Petsko, G. C. Johnston, D. Ringe, R. A. Singer, M. Werner-Washburne, *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 187-206; b) S. Zaman, S. I. Lippman, X. Zhao, J. R. Broach, *Annu. Rev. Genet.* **2008**, *42*, 27-81.
[156] L. Galdieri, S. Mehrotra, S. Yu, A. Vancura, *OMICS* **2010**, *14*, 629-638.
[157] a) E. K. Fuge, E. L. Braun, M. Wernerwashburne, *J. Bacteriol.* **1994**, *176*, 5802-5813; b) W. H. Mager, A. J. J. Dekruijff, *Microbiol. Rev.* **1995**, *59*, 506-&; c) Y. Sanchez, S. L. Lindquist, *Science* **1990**, *248*, 1112-1115; d) M. Wernerwashburne, J. Becker, J. Kosicsmithers, E. A. Craig, *J. Bacteriol.* **1989**, *171*, 2680-2688.
[158] C. Brandmayr, M. Wagner, T. Brueckl, D. Globisch, D. Pearson, A. C. Kneuttinger, V. Reiter, A. Hienzsch, S. Koch, I. Thoma, P. Thumbs, S. Michalakis, M. Mueller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 11162-11165.
[159] L. M. Dickson, A. J. P. Brown, *Mol. Gen. Genet.* **1998**, *259*, 282-293.
[160] a) R. Hershberg, D. A. Petrov, in *Annu. Rev. Genet., Vol. 42*, **2008**, pp. 287-299; b) J. B. Plotkin, G. Kudla, *Nat. Rev. Genet.* **2011**, *12*, 32-42.
[161] T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, Y. Pilpel, *Cell* **2010**, *141*, 344-354.
[162] M. J. Martinez, S. Roy, A. B. Archuletta, P. D. Wentzell, S. Santa Anna-Arriola, A. L. Rodriguez, A. D. Aragon, G. A. Quinones, C. Allen, M. Werner-Washburne, *Mol. Biol. Cell* **2004**, *15*, 5295-5305.
[163] M. Radonjic, J. C. Andrau, P. Lijnzaad, P. Kemmeren, T. T. Kockelkorn, D. van Leenen, N. L. van Berkum, F. C. Holstege, *Mol. Cell* **2005**, *18*, 171-183.
[164] N. P. Hoyle, L. M. Castelli, S. G. Campbell, L. E. Holmes, M. P. Ashe, *J. Cell. Biol.* **2007**, *179*, 65-74.
[165] J. Lui, S. G. Campbell, M. P. Ashe, *Biochem. Soc. Trans.* **2010**, *38*, 1131-1136.
[166] T. D. A. L. V., in *Current Protocols in Molecular Biology, Vol. 23*, John Wiley, New York, **1993**, pp. 13.11.11-13.11.17.
[167] V. Reiter, Doctoral thesis, Ludwig-Maximilians-Universität München Munich, Germany, **2013**.
[168] R. Kambampati, C. T. Lauhon, *Biochemistry* **2003**, *42*, 1109-1117.
[169] T. Numata, Y. Ikeuchi, S. Fukai, T. Suzuki, O. Nureki, *Nature* **2006**, *442*, 419-424.
[170] I. Moukadiri, S. Prado, J. Piera, A. Velazquez-Campoy, G. R. Bjork, M. E. Armengod, *Nucleic Acids Res.* **2009**, *37*, 7177-7193.
[171] a) T. G. Hagervall, C. G. Edmonds, J. A. McCloskey, G. R. Bjork, *J. Biol. Chem.* **1987**, *262*, 8488-8495; b) J. M. Bujnicki, Y. Oudjama, M. Roovers, S. Owczarek, J. Caillet, L. Droogmans, *RNA* **2004**, *10*, 1236-1242.
[172] D. Pearson, T. Carell, *Nucleic Acids Res.* **2011**, *39*, 4818-4826.
[173] a) A. Kitamura, T. Sengoku, M. Nishimoto, S. Yokoyama, Y. Bessho, *Protein Sci.* **2011**, *20*, 1105-1113; b) J. Kim, S. C. Almo, *BMC Struct. Biol.* **2013**, *13*, 5.
[174] Y. Taya, Nishimur.S, *Biochem. Biophys. Res. Commun.* **1973**, *51*, 1062-1068.
[175] a) M. Roovers, Y. Oudjama, K. H. Kaminska, E. Purta, J. Caillet, L. Droogmans, J. M. Bujnicki, *Proteins* **2008**, *71*, 2076-2085; b) O. Dym, D. Eisenberg, *Protein Sci.* **2001**, *10*, 1712-1728.
[176] E. C. Settembre, P. C. Dorrestein, J. H. Park, A. M. Augustine, T. P. Begley, S. E. Ealick, *Biochemistry* **2003**, *42*, 2971-2981.
[177] a) J. Sanchis, L. Fernandez, J. D. Carballeira, J. Drone, Y. Gumulya, H. Hoebenreich, D. Kahakeaw, S. Kille, R. Lohmer, J. J. P. Peyralans, J. Podtetenieff, S. Prasad, P. Soni, A. Taglieber, S. Wu, F. E. Zilly, M. T. Reetz, *Appl. Microbiol. Biotechnol.* **2008**, *81*, 387-397; b) W.-C. Tseng, J.-W. Lin, T.-Y. Wei, T.-Y. Fang, *Anal. Biochem.* **2008**, *375*, 376-378.

[178] a) A. Mattevi, M. A. Vanoni, F. Todone, M. Rizzi, A. Teplyakov, A. Coda, M. Bolognesi, B. Curti, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 7496-7501; b) L. Pollegioni, K. Diederichs, G. Molla, S. Umhau, W. Welte, S. Ghisla, M. S. Pilone, *J. Mol. Biol.* **2002**, *324*, 535-546.
[179] K. Rippe, *B.I.F. Futura* **1997**, 20-26.
[180] a) H. Sasaki, Y. Matsui, *Nat. Rev. Genet.* **2008**, *9*, 129-140; b) M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **2008**, *9*, 465-476.
[181] A. Meissner, T. S. Mikkelsen, H. C. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. L. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, E. S. Lander, *Nature* **2008**, *454*, 766-U791.
[182] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501-502.
[183] D. Bruniquel, R. H. Schwartz, *Nat. Immunol.* **2003**, *4*, 235-240.
[184] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, *Nat. Chem. Biol.* **2012**, *8*, 328-330.
[185] B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366-369.
[186] C. S. Schmidt, S. Bultmann, D. Meilinger, B. Zacher, A. Tresch, K. C. Maier, C. Peter, D. E. Martin, H. Leonhardt, F. Spada, *PLoS One* **2012**, *7*.
[187] C. Frauer, H. Leonhardt, *Nucleic Acids Res.* **2009**, *37*.
[188] K. Williams, J. Christensen, K. Helin, *EMBO Rep.* **2012**, *13*, 28-35.
[189] J. Song, M. Teplova, S. Ishibe-Murakami, D. J. Patel, *Science* **2012**, *335*, 709-712.
[190] a) J. D. Dignam, *Methods Enzymol.* **1990**, *182*, 194-203; b) J. D. Dignam, R. M. Lebovitz, R. G. Roeder, *Nucleic Acids Res.* **1983**, *11*, 1475-1489.
[191] A. Holmberg, A. Blomstergren, O. Nord, M. Lukacs, J. Lundeberg, M. Uhlen, *Electrophoresis* **2005**, *26*, 501-510.

11 Contributions

Declaration of contribution to „Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with Translational Efficiency " (see Chapter 3)

The project was started by Dr. Tobias Brückl and Dr. Daniel Globish, who performed the initial measurement using pig tissues. Together with Mirko Wagner, I performed the quantification studies on mouse tissues and the *in vitro* translation assay using mouse samples, as well as the quantification of Ψ in mouse and pig tissue and the later refinement of the quantification of m^1G modification in pig tissues. I contributed to writing of the manuscript together with Mirko Wagner, Dr. Markus Müller and Prof T. Carell. I contributed and prepared Figures 1 and 2, as Supplementary Figures S3-S4, S7-S10 and Tables S2-S3, S6-S8.

This article was reprinted with permission from C. Brandmayr,[1] M. Wagner,[1] T. Bruckl,[1] D. Globisch, D. Pearson, A. C. Kneuttinger, V. Reiter, A. Hienzsch, S. Koch, I. Thoma, P. Thumbs, S. Michalakis, M. Muller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 11162-11165. "*Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with Translational Efficiency*". Copyright © 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Declaration of contribution to „Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction" (see Chapter 7)

I performed the incorporation and deprotection of labelled, protected dhmCTP together with Stefan Schiesser, as well as the incorporation of dcaCTP. I contributed to Figures 2 and 3 and helped with writing of the manuscript and of the supplementary information relating to the PCR studies.

This article was reprinted with permission from B. Steigenberger,[1] S. Schiesser,[1] B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366-369. "*Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction*". Copyright © 2013 American Chemical Society.

Declaration of contribution to „Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA" (see Chapter 8)

I performed all the experiments in HEK-293T cells, together with Mirko Wagner. I additionally performed the qPCR studies for analysis of SMUG1 expression levels in mESC knock-down samples and for analysis of glycosylases, Tet and Dnmt proteins during differentiation of EpiLC samples. I contributed to writing of the manuscript and of the supplementary information.

The article was reprinted with permission from T. Pfaffeneder,[1] F. Spada,[1] M. Wagner,[1] C. Brandmayr, S. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Muller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schar, M. Muller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574-581. "*Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA"*. Copyright © 2014 Nature America, Inc.

12 Abbreviations

A	Adenosine
A-site	Aminoacyl tRNA binding site
AID	Activation-induced deaminase
Am	2´-O-Methyladenosine
Ar	2´-O-Ribosyladenosine
ASL	Anticodon stem and loop
BER	Base excision repair
bp	Base pair
BSA	Bovine serum albumine
C	Cytosine
caC	5-Carboxycytosine
Cm	2´-O-Methylcytosine
CpG	dC-dG dinucleotide
DEAE	Diethylaminoethylcellulose
DMEM	Dulbecco´s Modified Eagle´s Medium
DNA	Deoxyribonucleic acid
ds	Double stranded
DSL	Dihydrouridine stem and loop
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylendiamine tetraacetate
ES cells	Embryonic stem cells
ESI	Electrospray Ionisation
EtOH	Ethanol
FAD	Flavin adenine dinucleotide
fC	5-formylcytosine
Gm	2´-O-Methylguanosine
h	Hour
hmC	5-Hydroxymethylcytosine
hmU	5-Hydroxymethyluracil
HPLC	High performance liquid chromatography
I	Inosine
i^6A	N^6-isopentenyladenosine
IPTG	Isopropylthiogalactoside
m^1A	1-Methyladenosine
m^1G	1-Methylguanosine
m^2A	2-Methyladenosine
m^2G	N^2-Methylguanosine
$m^2{}_2G$	N^2,N^2-Dimethylguanosine
m^7G	7-Methylguanosine
mC	5-Methylcytosine
mcm^5s^2U	5-Methoxycarbonylmethyl-2-thiouridine
MeCN	acetonitrile
min	Minutes

mnm^5s^2U	5-Methylaminomethyl-2-thiouridine
MS	Mass spectrometry
ms^2i^6A	2-Methylthio-N^6-isopentenyladenosine
OD	Optical density
OHyW	Hydroxywybutosine
P-site	Peptidyl tRNA binding site
PCR	Polymerase chain reaction
PDB	Protein Data Bank
Q	Queuosine
RNA	Ribonucleic acid
rpm	Rotation per minute
RS	tRNA synthase
S. cerevisiae	*Saccharomyces cerevisiae*
SAM	*S*-Adenosylmethionine
SDS	Sodium dodecylsulfate
SMUG1	Single-strand-selective monofunctional uracil-DNA glycosylase 1
t^6A	N^6-THeronylcarbamoyladenosine
Tet	Ten-eleven translocation
Tris	Tris(-hydroxymethyl)-aminomethane
tRNA	Transfer ribonucleic acid
TSL	Thymidine stem and loop
U	Uridine
UV	Ultraviolet
vol.	Volume
Ψ	Pseudouridine
yW	Wybutosine

12. Abbreviations

13 Supplementary Information

13.1 Supplementary material for Chapter 3

C. Brandmayr,[1] M. Wagner,[1] T. Bruckl,[1] D. Globisch, D. Pearson, A. C. Kneuttinger, V. Reiter, A. Hienzsch, S. Koch, I. Thoma, P. Thumbs, S. Michalakis, M. Muller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 11162-11165. "*Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with Translational Efficiency*."

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with Translational Efficiency**

*Caterina Brandmayr, Mirko Wagner, Tobias Brückl, Daniel Globisch, David Pearson, Andrea Christa Kneuttinger, Veronika Reiter, Antje Hienzsch, Susanne Koch, Ines Thoma, Peter Thumbs, Stylianos Michalakis, Markus Müller, Martin Biel, and Thomas Carell**

anie_201203769_sm_miscellaneous_information.pdf

Supplementary Materials and Methods

Porcine and murine tissue samples. Pork tissue was obtained from a slaughterhouse immediately after sacrifice and processed within 4 h. Prolonged waiting times were observed to result in impaired results. Heart, kidneys, liver and spleen were briefly washed with demineralized water. Cerebellum, cerebrum, spinal cord and lung were washed rigorously and superficial blood vessels were removed. Each organ was sampled at three different positions from two animals to give 6 samples in total. Tissue samples were cut out omitting surface areas. Mouse organs were prepared from C57BL/6N mice (5 weeks old) and frozen in liquid nitrogen. The organs were kept at –80°C until tRNA extraction was performed.

tRNA extraction. All extraction steps were performed on ice or at 4 °C. All extraction, desalting, and chromatography steps were performed with minor changes as described previously [1]. Pork tissue samples (5 g) or whole cerebellums, buffer 1 (15 mL, 0.01 M $Mg(OAc)_2$, 0.05 M NaOAc, 0.15 M NaCl, pH 4.5) and ice were placed into a Waring Blender. The mixture was homogenized and transferred to a 50 mL Falcon tube and extracted three times with phenol. A similar lysis protocol was used for whole mouse organs. The final purified dry tRNA pellet was dissolved in MilliQ water (200-1000 µL) for enzymatic digestion. In case the resulting tRNA concentration proved to be too low for the subsequent digestion (<140 ng/µl) another EtOH precipitation step was conducted. Purity of the tRNA samples was proven by gel electrophorsis (Fig. S10).

Enzymatic digestion. The enzymatic digestion was performed in a two step procedure as described previously [1]. All labeled nucleosides were added, followed by centrifugation of the sample (12100 g, 15 min). The supernatant was lyophilized to a total volume of 105 µL. Each pork and mouse tissue experiment was performed at least in triplicate with three independent concentrations of the appropriate labeled nucleosides. The concentrations of standard solutions were chosen to be in the expected range of the sample nucleoside concentration.

LC-ESI-MS. The samples (100 µL injection volume) were analyzed by LC-ESI-MS on a *Thermo Finnigan LTQ Orbitrap XL* and were chromatographed by a *Dionex Ultimate 3000*

HPLC system with a flow of 0.15 mL/min over an Uptisphere120-3HDO column from *Interchim*. The column temperature was maintained at 30 °C. Eluting buffers were buffer C (2 mM $HCOONH_4$ in H_2O (pH 5.5)) and buffer D (2 mM $HCOONH_4$ in H_2O/MeCN 20/80 (pH 5.5)). The gradient was 0 → 41.25 min; 0 % → 6 % buffer D; 41.25 → 80 min; 8 % → 60 % buffer D; 80 → 82 min; 60 % → 100 % buffer D; 82 → 100 min; 100 % buffer D; 100 → 105 min; 100 → 0 % buffer D; 105 → 115 min; 0 % buffer D. The elution was monitored at 260 nm (*Dionex Ultimate 3000 Diode Array Detector)*. The chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of *m/z* 200-1000 with a resolution of 30000. Parameters of the mass spectrometer were tuned with a freshly mixed solution of adenosine (5 µM) in buffer C. The parameters used in this section were sheath gas flow rate, 16 arb; auxiliary gas flow rate, 10 arb; sweep gas flow rate, 4 arb; spray voltage, 5.0 kV; capillary temperature, 200 °C; capillary voltage, 25 V, tube lens 60 V.

LC-MS of RNase A digests. RNase A digests were performed by incubation of tRNA (10 µg) with RNase A (1 or 10 µg, Fermentas) in 100 mM ammonium acetate buffer at a total volume of 100 µL at 37 ºC for 2 h. Whole digested samples were analyzed by the same LC-ESI-MS system as for the quantitative analysis with a flow of 0.15 mL/min (*Thermo Finnigan LTQ Orbitrap XL*; *Dionex Ultimate 3000 HPLC* system; Uptisphere120-3HDO column from *Interchim*). The column temperature was maintained at 30 °C. We used the same eluting buffers C and D. The gradient was 0 min → 10 min; 0 % buffer D; 10 → 20 min; 0 % → 10 % buffer D; 20 min → 42 min; 10 % → 40 % buffer D; 42 min → 50 min; 40 % → 100 %; 50 min → 60 min; 100 % buffer D; 60 min → 70 min; 100 % → 0 % buffer D; 70 min → 80 min; 0 % buffer D. The parameters used in this section were sheath gas flow rate, 30 arb; auxiliary gas flow rate, 10 arb; sweep gas flow rate, 4 arb; spray voltage, 2.3 kV; capillary temperature, 200 °C; capillary voltage, -20 V, tube lens -93 V. Ions were scanned by use of a negative polarity mode over a full-scan range of *m/z* 200-1500 with a resolution of 30000. Parameters of the mass spectrometer were tuned with a solution of the RNA 7-mer AUUCCCG (5 µM) in buffer C.

Separation of mitochondria and cytosol. Isolation of mitochondria from porcine organs was done by fractionated centrifugation at low and high speed according to the Mitochondria Isolation Kit from Sigma-Aldrich.[2]

Cytochrome C oxidase assay. The respiratory activity of mitochondria in whole cell lysates and in enriched fractions of pork tissue was measured using the cytochrome C oxidase Assay Kit of Sigma-Aldrich.[3]

***In vitro* translation assay.** For *in vitro* translation activity measurements we used the combined transcription and translation reticulocyte assay kit from Promega (TNT Coupled Reticulocyte Lysate System). tRNAs present in the reticulocyte lysate were removed by chromatography using an ethyanolamine-Sepharose column according a previous report [4]. The chromatographic step was performed at 4 °C and collected fractions were frozen at -80 °C. Briefly, we supplemented the lysate with 20 µM hemin, 50 mM KCl, and 0.5 mM $MgCl_2$ prior to chromatography. The column was equilibrated with buffer H (75 mM KCl, 1.6 mM $MgCl_2$, 10 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10 mM HEPES pH = 7.2). The supplemented extract (700 µL) was loaded on the column and eluted with 1.0 mL of buffer H. Fractions of 50 µL were collected and assayed for tRNA dependent translational activity before usage. The column was regenerated between different filtration steps with buffer I (500 mM KCl, 1,6 mM $MgCl_2$, 10 mM NaCl, 0,1 mM EDTA, 1 mM DTT, 10 mM HEPES pH = 7.2) and stored at 4 °C. The collected tRNA-depleted fractions were used for *in vitro* translation experiments and stored at -80 °C.

For the *in vitro* translation assay, we used the non-radioactive luciferase control reaction with TNT RNA Polymerase T7 and Luciferase control DNA as described in the Promega Kit with minor changes. Each assay contained TNT reaction buffer (0.5 µL), T7 TNT RNA Polymerase (0.25 µL), amino acid mixture minus leucine (1 mM), amino acid mixture minus methionine (1mM), RNasin ribonuclease inhibitor (10 U), Luciferase T7 control DNA (0.25 µg). Afterwards the appropriate tRNA (12.5 ng) and RNase-free water were added to a total assay volume of 6.25 µL, followed by addition of tRNA-depleted lysate (6.25 µL).

These translation reactions were incubated at 30 °C and a 1 µL aliquot was removed every 2-3 min from each fraction starting at 12 min. This sample was transferred into a 96 well plate, mixed well with 25 µL of Luciferase Assay substrate and analyzed immediately.

Luminescence was measured with a TECAN Microplate Reader Genios Pro in 10 cycles for 100 ms each cycle. Reproducible luminescence slopes were obtained after normalization to the liver value and the results are shown in Fig S7. The data were averaged and plotted against time. A linear fit of each initial slope was performed and normalized to the highest value (usually liver). Every measurement was repeated at least in triplicate and every value represents mean value with s.d. Care was taken to use different tRNA-depleted fractions.

Calibration curves. Mass calibration curves of the labeled and corresponding unlabeled synthesized nucleosides were obtained at five different concentration ratios. For each concentration an average value of three independent measurements was determined (Fig. S2). Each labeled nucleoside solution was mixed with three different concentrations of the corresponding unlabeled nucleosides. The areas of labeled and unlabeled nucleosides from LC-MS measurements were determined using the *Qualbrowser* program by extraction of the accurate mass range with a mass filter (Table S1) from the total ion current (TIC). The linear fits of the determined area ratios with the amount ratios gave R^2-values of minimum 0.9992. The linear fit equations were used for calculation of the exact nucleoside contents in bulk tRNA samples. Synthetic labeled nucleosides were added to the digest solutions and the areas of labeled and unlabeled nucleosides were determined as described above. The amount of each nucleoside (Table S2-7) was calculated from the obtained area ratios and the linear fit equations of the calibration curves.

Ψ calibration curves (see Fig. S3) were obtained by addition of three different known amounts of non labeled Ψ standard to three of four liver tRNA extract samples previously digested as described above. The samples were measured by HPLC-ESI-MS together with the non-spiked liver tRNA extract sample. The ion current area of the non-spiked sample was subtracted from the ion current areas of the spiked samples and the resulting values were plotted against the amount of Ψ added. The derived linear fit equation was used for quantification of Ψ in the investigated samples.

Intra- and Inter-assay tests. The intra-assay test was performed for representative nucleosides m^1A, i^6A, ms^2i^6A, and m^1G. The determined values of labeled to unlabeled nucleosides using the calibration curves of a sample after enzymatic digestion showed good reproducibility for each nucleoside (N=5); 2.5% for m^1A, 0.4% for i^6A, 0.7% for ms^2i^6A, and

2.6% for m^1G. The single area values of each labeled or unlabeled nucleoside in the intra-assay test resulted in an average value of 4.1%. The inter-assay test gave an area ratio reproducibility (N=6) of 6.3% for m^1A, 1.0% for i^6A, 1.8% for ms^2i^6A, and 4.3% for m^1G on six subsequent days. The determined single area values for each nucleoside resulted in strong variations and large average error of 34.5%. Using the area ratio with the calibration curves we gained perfect reproducibility (Table S10). No memory effect was observed during blank LC/MS experiments performed after several measurements of a sample. The blank analyses were not contaminated by carry-over.

Quantification of the Ψ nucleoside in murine and porcine tissues. Biological samples from different murine and porcine tissues were analyzed by HPLC-ESI-MS without addition of standards and quantification of Ψ was obtained from the ion current area of the unlabeled Ψ in each sample (accurate mass range is given in Table S1) and the linear fit equation of the calibration curve (see Fig.S3). The HPLC-ESI-MS protocol used was similar to that used for quantification by isotope-dilution (see above). Care was taken to perform all the measurements and the data collection for the calibration curve during the same day so to avoid variations in the mass detection accuracy.

Influence of codon bias on the *in vitro* translation activity experiment. Observed differences in translation activity could be attributed both to a variable composition of isoacceptor tRNAs and differences in tRNA modification levels. A tissue-specific codon bias has in fact been statistically proven in mammals by Plotkin *et al.*,[5] but they also show that the same tissues of evolutionary closely related species feature the same codon usage. From our analysis of the tRNA ensembles of cerebrum and spleen we can deduce that in our experiments codon bias only plays a minor role. As shown in our data, the porcine spleen tRNA set shows a higher level of modification than the cerebrum tRNAs. This is in good accordance with our observation that the *in vitro* translation activity of porcine spleen tRNA is higher than that of cerebrum. In the case of mouse we observe the opposite behavior. The murine spleen tRNA set is less modified than its cerebrum tRNAs, which goes in line with a lower translation activity of murine spleen tRNA *in vitro* compared to cerebrum.

If codon usage strongly biased the *in vitro* translation results presented here, we would expect similar outcomes for both organisms regarding the translational activity in a certain tissue. This is not the case.

Supplementary Figures

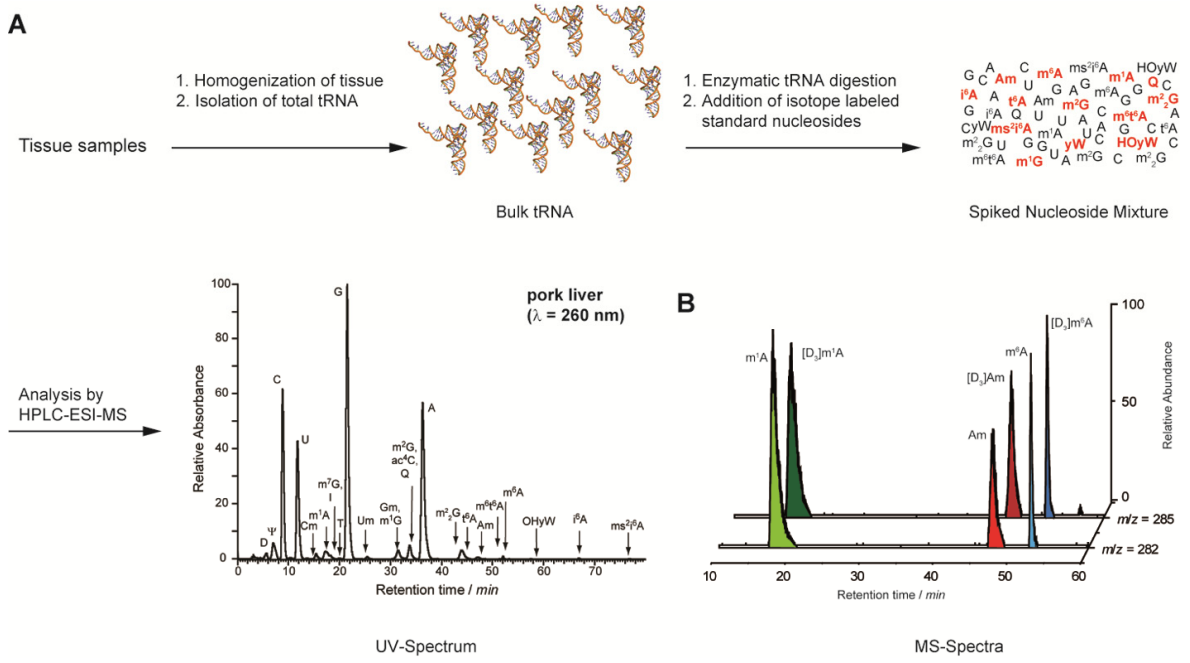


Fig. S1. Illustration of the isotope dilution method used to quantify modified tRNA nucleosides. A) The general method used, starting with homogenization of different tissue samples and ending with LC-MS analysis. The red colour letters represent the added isotope-labeled derivatives. B) Example of the two ion currents obtained for the natural monomethylated adenosines m^1A, Am and m^6A and the corresponding labeled nucleosides d_3-m^1A, d_3-Am and d_3-m^6A.

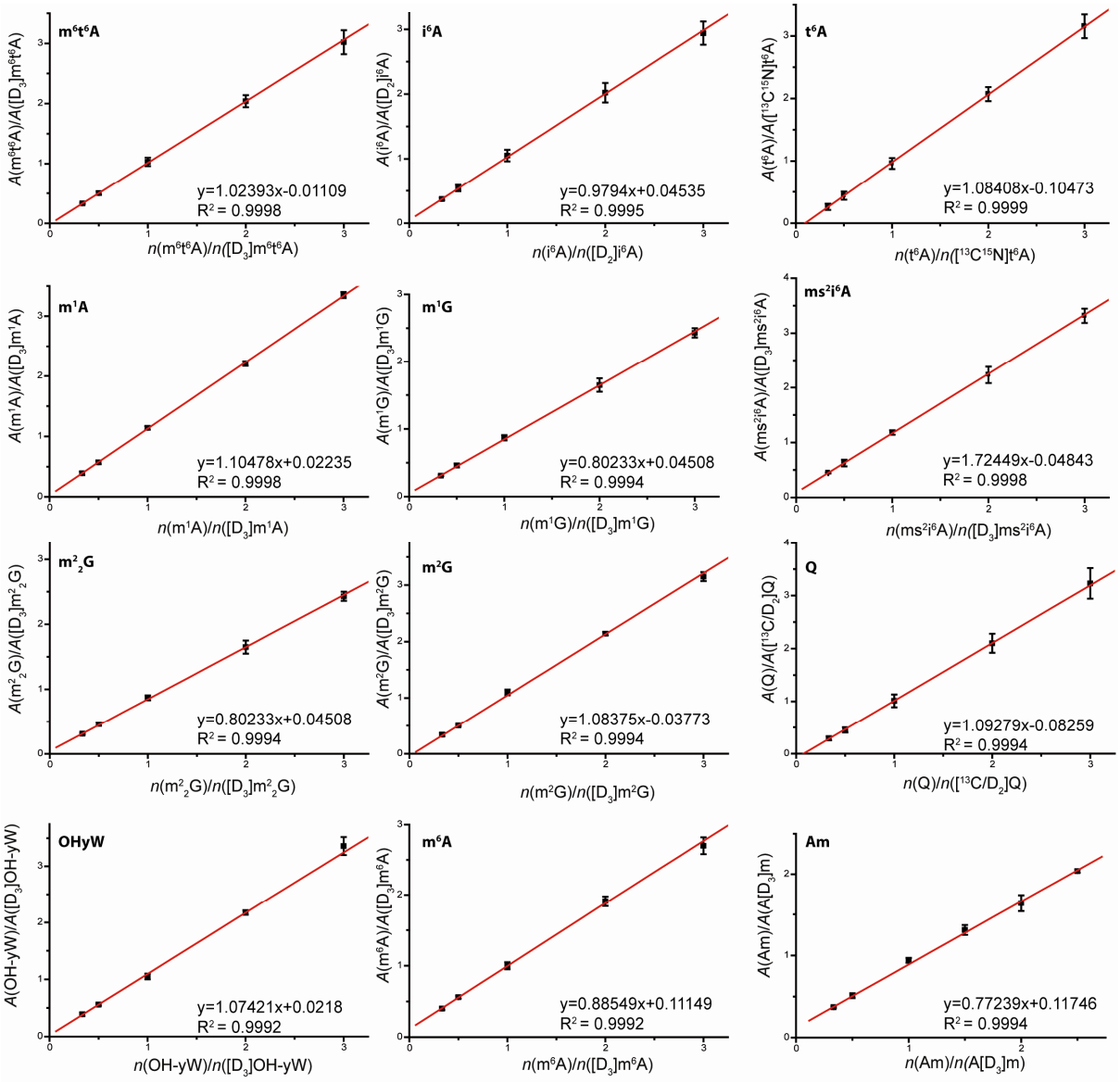


Fig. S2. Calibration curves for the nucleosides synthesized and used for quantification: m^6t^6A, i^6A, t^6A, m^1A, m^1G, ms^2i^6A, $m^2{}_2G$, m^2G, Q, OHyW, m^6A, and Am, with an average R^2 value of 0.9995. Quantitative data for the modification m^6A are not presented due to rearrangement from m^1A to m^6A. Inter-and intra-assays proved constant values of m^1A and varying values of m^6A. The modifications m^1G and Gm were quantified with d_3-m^1G because of overlapping UV and mass signals.

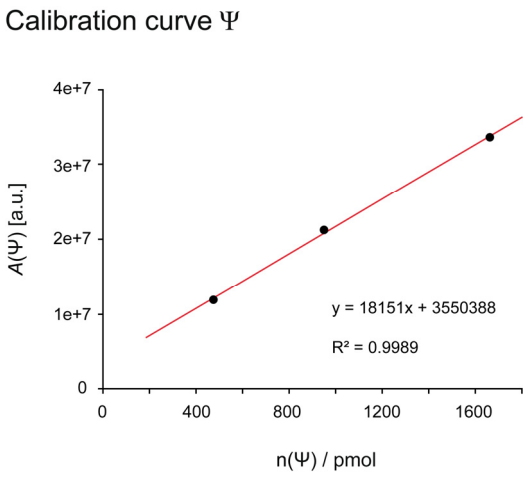


Fig. S3. Calibration curve for the tRNA nucleoside Ψ with a R^2 value of 0.9989 (for details, see above). Care was taken to perform data collection for the calibration curve and the measurements of the biological samples during the same day so to avoid variations in the mass detection accuracy.

A

Mouse

Nucleosides / Tissues	m^5C	Ψ
Liver	1888.0	1986.9
Cerebrum	1831.5	1639.2
Lung	1747.8	1436.9
Kidney	1316.5	924.5
Spleen	1368.5	1284.7

B

Pork

Nucleosides / Tissues	m^1G	m^5C	Ψ
Liver	746.7	2742.5	2263.1
Spleen	627.1	2687.5	2146.1
Kidney	628.8	2764.5	1886.8
Lung	575.8	2452.7	1907.6
Heart	535.1	2138.1	1776.1
Cerebrum	448.5	1936.4	1330.4

Fig. S4. Quantitative data for the tRNA nucleosides m^1G, m^5C and Ψ in various representative murine (*A*) and porcine (*B*) tissues. All tRNA nucleoside values are given per 1000 tRNA molecules (‰). These data reveal a similar, tissue-dependent extent of modification for the investigated modified nucleosides, corresponding to the trend shown in the main paper. The Ψ values were determined by the method described above without addition of an isotope-labeled derivative. A separate quantification of m^1G for porcine tissues (*B*) was possible using the optimized LC-method. Color codes in (*A*) and (*B*) are based on quantile calculations; red: highest value, yellow: 50% quantile, green: lowest value. For intermediate values appropriate shades of color were calculated, standard deviations are listed in Tables S6-S7.

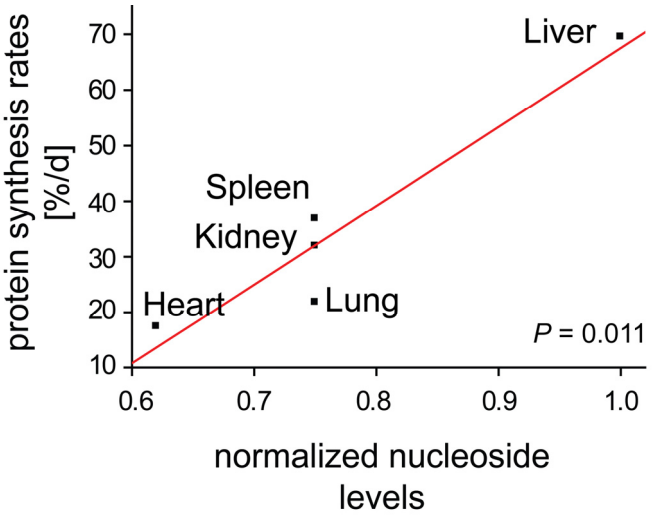


Fig. S5. Correlation of *in vivo* protein synthesis rates with normalized nucleoside levels. This correlation shows a high significance with $P = 0.011$ [6].

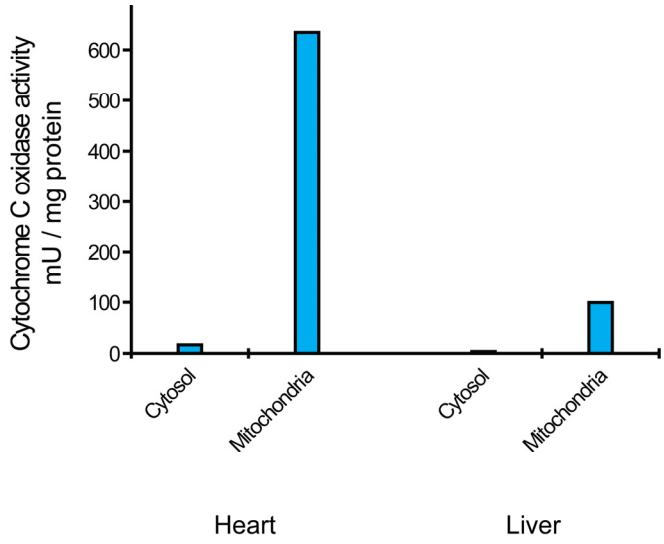


Fig. S6. Cytochrome C oxidase activity after separation of mitochondria and cytosol in porcine tissues. Activity of cytochrome C oxidase was taken as a measure of the mitochondrial content in each fraction. Values for the activity are increased in mitochondrial fractions and decreased in cytosolic fractions for the two representative tissues heart and liver, therefore providing evidence for an enrichment of mitochondria in mitochondrial fractions and suggesting an almost mitochondria-free cytosolic fraction.

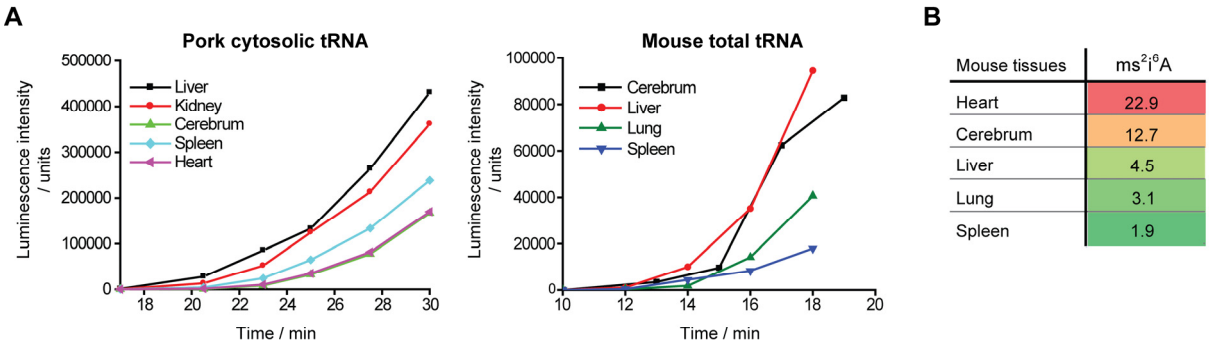


Fig. S7. (A) Representative *in vitro* translation experiment read out of pork cytosolic tRNAs (liver, kidney, cerebrum, spleen, heart) and mouse total tRNAs (cerebrum, liver, lung and spleen). Mouse tissues were selected based on low mitochondrial content, as shown by the relatively low content of ms^2i^6A modification (see (B)), which is known to only occur in mitochondrial tRNA.[3] Heart is listed to provide a comparison with a tissue known to contain a high proportion of mitochondrial tRNA. This experiment was performed at least in triplicate. Initial rates were calculated in the range of 17-25 min and normalized to liver in the case of pork samples. For mouse tRNA ensembles, rates were calculated from data points between 12-16 minutes and normalized to cerebrum.

Tissue	normalized nucleoside levels[a]	normalized *in vitro* translation activity (total tRNA)	normalized *in vitro* translation activity (cytosolic tRNA)
Liver	1.00	0.95 ± 0.02	1.00
Cerebellum	0.99 ± 0.23	1.00	0.96 ± 0.05
Spleen	0.75 ± 0.08	0.77 ± 0.22	0.72 ± 0.12
Kidney	0.75 ± 0.09	0.79 ± 0.07	0.87 ± 0.01
Cerebrum	0.68 ± 0.17	0.64 ± 0.02	0.41 ± 0.04
Heart	0.62 ± 0.14	0.90 ± 0.04	0.55 ± 0.14

Fig. S8. Translational activity of tRNA sets from pork tissues. Average normalized nucleoside levels of pork liver, cerebellum, spleen, kidney, cerebrum, and heart (calculated from LC-MS data presented in Fig. 2*B*) and relative *in vitro* translation activities of cytosolic and total tRNA. All values are normalized to the highest value. Standard deviations (mean ± s.d.) are given for the other tissues. Note: while the error values here are relatively large, these represent the variation over all modified nucleosides. The measurements for each nucleoside have low errors (~5%), and they show the same relationship as the averaged set.

A

Tissue	normalized nucleoside levels	normalized *in vitro* translation activity (total tRNA)		
		replicate 1	replicate 2	replicate 3
Liver	1.00	0.42 ± 0.05	0.58 ± 0.23	0.72 ± 0.19
Cerebrum	0.969 ± 0.05	1.00	0.39 ± 0.17	-
Lung	0.855 ± 0.06	0.19 ± 0.09	0.21 ± 0.12	0.30 ± 0.10
Spleen	0.675 ± 0.09	0.01 ± 0.001	0.08 ± 0.00	0.31 ± 0.002

B

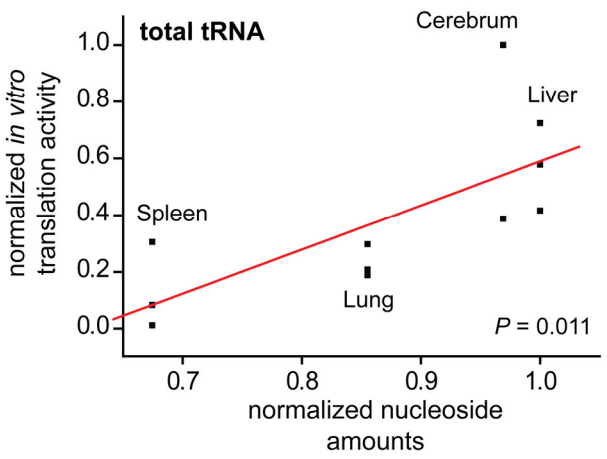


Fig. S9. Translational activity of tRNA sets from mouse tissues. (A) Average normalized nucleoside levels of mouse liver, cerebrum, lung and spleen (calculated from LC-MS data presented in Fig. 2*A*) and corresponding *in vitro* translation rates for each replicate. (B) Linear fit of normalized in vitro translational acitivity of total murine tRNAs and normalized nucleoside levels showing a considerable correlation (r = 0.723, *P* = 0.011). Each data set was normalized to its corresponding highest value.

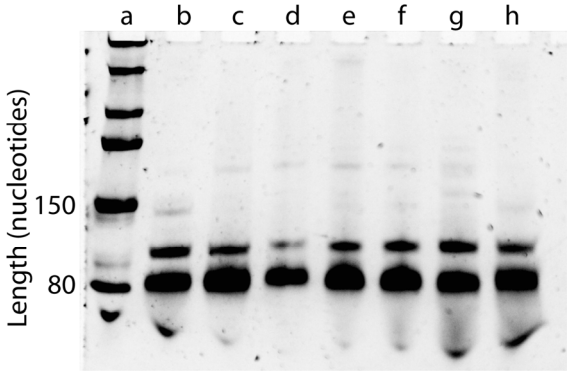


Fig. S10. RNA PAGE gel for mouse tRNA extracts from different tissues. a) Marker, b) Liver, c) Cerebrum, d) Cerebellum, e) Brain stem, f) Lung, g) Spleen, h) Kidney. As previously reported by others, there is a small but constant contamination of 5S rRNA. This contamination is observed for all tissues and therefore does not bias the results reported in this publication [7].

Supplementary Tables

Nucleoside	Nucleosides mass range *m/z*	Labeled nucleosides mass range *m/z*
Am, m^1A	282.1142-282.1262	285.1335-285.1435
t^6A	413.1315-413.1475	418.1420-418.1580
i^6A	336.1606-336.1716	338.1740-338.1840
ms^2i^6A	382.1484-382.1594	385.1679-385.1789
$m^2{}_2G$	312.1248-312.1368	315.1441-315.1561
m^2G, m^1G, Gm	298.1076-298.1196	301.1276-301.1396
Q	410.1640-410.1730	413.1778-413.1898
m^6t^6A	427.1532-427.1622	430.1707-430.1807
OHyW	525.1879-525.2029	528.2065-528.2195
m^5C	258.1034-258.1134	261.1223-261.1323
Ψ	245.0708-245.0828	-

Table S1. High resolution mass ranges of natural and corresponding labeled nucleosides used for quantification. Modifications 3'-adjacent to the anticodon loop: hydroxywybutosine (OHyW), *N*^{6}-isopentenyladenosine (i^6A), *N*^{6}-methyl-*N*^{6}-threonylcarbamoyladenosine (m^6t^6A), *N*^{6}-threonylcarbamoyladenosine (t^6A), and 1-methylguanosine (m^1G). Modifications in the wobble position: queuosine (Q) and 2'-*O*-methylguanosine (Gm). Modifications in other positions: 1-methyladenosine (m^1A), 2'-*O*-methyladenosine (Am), m^1G, *N*^{2}-methylguanosine (m^2G), *N*^{2},*N*^{2}-dimethylguanosine ($m^2{}_2G$), 5-methylcytosine (m^5C), pseudouridine (Ψ) and Gm.

Mouse Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G	m^6t^6A	m^1A
Kidney	Mean value	21.0	202.7	36.0	389.5	779.7	70.8	22.5	517.9	19.8	715.7
	Standard deviation	3.6	5.8	0.9	12.3	40.3	6.2	4.3	23.0	1.3	66.0
	Standard deviation in %	17.2	2.9	2.5	3.2	5.2	8.8	19.3	4.4	6.7	9.2
Lung	Mean value	28.7	271.7	45.0	495.2	946.2	81.3	21.7	640.0	24.8	876.5
	Standard deviation	18.2	18.3	1.9	30.5	48.1	7.4	3.4	21.5	1.8	54.2
	Standard deviation in %	63.2	6.7	4.3	6.2	5.1	9.2	15.9	3.4	7.3	6.2
Spleen	Mean value	47.8	210.3	30.6	397.0	754.3	47.7	26.9	487.0	20.9	653.6
	Standard deviation	1.4	7.8	2.1	3.2	43.9	17.4	11.4	26.3	0.9	60.1
	Standard deviation in %	2.9	3.7	6.8	0.8	5.8	36.5	42.4	5.4	4.3	9.2
Liver	Mean value	18.9	304.0	49.8	614.6	1102.7	62.5	37.7	735.0	30.2	1016.7
	Standard deviation	9.6	7.5	1.0	14.2	34.8	7.7	9.1	14.2	0.7	150.2
	Standard deviation in %	50.7	2.5	2.1	2.3	3.2	12.3	24.3	1.9	2.3	14.8
Cerebrum	Mean value	23.6	290.3	50.6	555.8	1104.0	109.7	27.6	694.1	24.2	972.4
	Standard deviation	1.7	10.8	3.0	23.0	33.8	11.9	2.4	24.2	0.6	39.9
	Standard deviation in %	7.4	3.7	5.9	4.1	3.1	10.9	8.6	3.5	2.5	4.1
Cerebellum	Mean value	20.7	274.8	41.1	524.4	1026.3	95.0	18.6	646.6	24.4	1011.0
	Standard deviation	1.4	14.9	1.7	18.1	67.7	9.1	3.3	35.3	0.9	118.9
	Standard deviation in %	6.6	5.4	4.2	3.5	6.6	9.6	17.6	5.5	3.5	11.8
Heart	Mean value	26.4	241.2	36.4	419.8	801.6	83.4	12.8	526.8	17.5	800.7
	Standard deviation	9.5	26.6	5.3	44.9	108.1	13.6	2.7	61.1	2.7	113.2
	Standard deviation in %	35.9	11.0	14.4	10.7	13.5	16.3	20.8	11.6	15.4	14.1

Mouse Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G	m^6t^6A	m^1A
Brain stem	Mean value	19.3	262.5	40.4	514.5	1025.7	93.5	16.9	644.3	23.2	944.3
	Standard deviation	3.6	9.5	2.7	23.4	72.1	6.7	3.4	39.2	0.7	111.0
	Standard deviation in %	18.7	3.6	6.6	4.5	7.0	7.2	20.5	6.1	3.1	11.8

Table S2. Modification numbers per 1000 tRNAs in different murine tissues. The table lists the average values calculated from the modification content of two sets of five animals. For each sample at least three independent digests and measurements were performed. 10,3 % mean standard deviation was obtained for all nucleosides excluding Am. 22,3 % mean standard deviation was obtained for Am. As noted in the Table S3, the overall mean standard deviation across nucleosides and tissues for the two individual sample sets is below 5%.

Mouse Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G	m^6t^6A	m^1A
Lung A	Mean value	49.6	255.8	43.0	479.9	938.6	90.1	23.9	634.8	24.4	852.3
	Standard deviation	0.6	15.5	0.8	27.9	62.0	1.2	3.1	30.3	2.0	59.1
	Standard deviation in %	1.3	6.1	1.8	5.8	6.6	1.3	13.0	4.8	8.2	6.9
Lung B	Mean value	13.1	283.6	46.5	506.6	951.9	75.3	20.0	643.9	25.2	900.0
	Standard deviation	2.1	8.5	0.9	27.2	33.0	2.0	2.6	9.1	1.6	39.4
	Standard deviation in %	16.1	3.0	1.9	5.4	3.5	2.6	13.0	1.4	6.2	4.4
Spleen A	Mean value	47.2	217.4	32.2	397.8	797.7	67.7	37.7	508.6	21.5	646.3
	Standard deviation	0.2	4.7	0.4	3.7	8.5	0.7	2.2	18.7	0.8	30.2
	Standard deviation in %	0.4	2.1	1.4	0.9	1.1	1.0	5.7	3.7	3.5	4.7
Spleen B	Mean value	48.3	203.2	29.1	396.5	711.0	32.7	16.1	465.5	20.3	660.9
	Standard deviation	1.8	0.9	1.9	2.6	6.0	1.7	4.8	10.4	0.6	78.7
	Standard deviation in %	3.7	0.5	6.6	0.7	0.8	5.2	29.7	2.2	3.0	11.9
Liver A	Mean value	28.4	310.6	50.1	612.6	1107.8	68.8	45.7	744.1	30.4	895.9
	Standard deviation	0.7	4.4	0.4	3.2	39.5	0.7	2.0	13.6	0.9	9.1
	Standard deviation in %	2.5	1.4	0.8	0.5	3.6	1.0	4.3	1.8	2.9	1.0
Liver B	Mean value	9.4	297.4	49.5	616.5	1097.7	54.2	29.7	725.8	30.0	1137.6
	Standard deviation	1.9	2.7	1.4	19.7	28.6	3.7	5.9	7.1	0.3	125.8
	Standard deviation in %	20.0	0.9	2.7	3.2	2.6	6.8	20.0	1.0	0.8	11.1
Cerebrum A	Mean value	22.2	298.6	52.8	572.2	1131.5	113.8	28.8	713.0	24.3	963.2
	Standard deviation	0.9	5.0	1.2	8.4	15.3	4.8	1.7	9.3	0.7	46.4
	Standard deviation in %	3.9	1.7	2.2	1.5	1.4	4.3	6.0	1.3	2.7	4.8

Mouse Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G	m^6t^6A	m^1A
Cerebrum B	Mean value	25.3	279.2	47.7	533.9	1067.3	104.2	25.2	668.9	23.9	984.6
	Standard deviation	0.7	5.2	1.8	17.4	2.7	15.8	1.4	11.6	0.4	24.2
	Standard deviation in %	2.9	1.8	3.9	3.3	0.2	15.1	5.7	1.7	1.6	2.5
Cerebellum A	Mean value	19.3	291.2	42.8	544.7	1096.1	102.7	20.7	684.4	25.2	1082.0
	Standard deviation	0.6	6.3	0.5	5.1	23.3	6.6	2.8	19.5	0.4	129.9
	Standard deviation in %	3.0	2.2	1.3	0.9	2.1	6.4	13.3	2.9	1.8	12.0
Cerebellum B	Mean value	21.8	262.6	39.9	509.2	974.0	87.3	17.0	618.2	23.8	957.8
	Standard deviation	0.7	3.0	1.2	3.8	35.1	2.2	2.6	3.8	0.5	74.0
	Standard deviation in %	3.1	1.1	3.0	0.8	3.6	2.5	15.5	0.6	2.0	7.7
Heart A	Mean value	17.0	266.2	41.1	461.9	905.8	93.9	15.7	584.3	20.1	891.9
	Standard deviation	0.6	8.0	2.4	12.8	30.2	7.7	0.3	21.7	0.9	82.1
	Standard deviation in %	3.5	3.0	5.8	2.8	3.3	8.2	1.7	3.7	4.4	9.2
Heart B	Mean value	35.7	216.3	31.6	377.6	697.4	69.4	10.7	469.4	14.9	709.4
	Standard deviation	2.3	10.4	2.0	17.8	27.1	3.4	1.2	20.1	0.6	47.1
	Standard deviation in %	6.6	4.8	6.3	4.7	3.9	4.9	11.7	4.3	3.7	6.6
Brain stem A	Mean value	14.5	274.8	43.5	543.1	1091.9	87.0	20.9	692.7	23.8	1023.2
	Standard deviation	1.2	3.6	1.5	0.4	32.1	-	1.0	3.3	0.9	40.7
	Standard deviation in %	8.6	1.3	3.4	0.1	2.9	-	5.0	0.5	3.6	4.0
Brain stem B	Mean value	21.7	256.4	38.9	500.2	992.5	95.1	14.8	620.2	22.8	904.8
	Standard deviation	1.3	4.0	1.6	14.3	63.2	6.6	2.2	23.4	0.4	113.9
	Standard deviation in %	5.9	1.6	4.2	2.8	6.4	6.9	14.8	3.8	1.5	12.6

Mouse Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G	m^6t^6A	m^1A
Kidney A	Mean value	24.6	204.2	36.4	376.3	810.6	74.4	25.9	540.1	20.7	770.0
	Standard deviation	0.6	7.6	0.6	2.9	26.6	4.5	0.5	5.0	0.9	9.8
	Standard deviation in %	2.3	3.7	1.7	0.8	3.3	6.0	2.1	0.9	4.4	1.3
Kidney B	Mean value	17.5	201.1	35.6	399.4	748.8	66.3	19.1	495.7	18.6	675.0
	Standard deviation	0.9	2.4	1.0	5.5	25.0	5.3	3.8	6.3	0.6	60.6
	Standard deviation in %	5.2	1.2	2.7	1.4	3.3	7.9	19.8	1.3	3.2	9.0

Table S3. Modification numbers per 1000 tRNAs in different murine tissues. The table lists the average values calculated from the modification content of each individual set of five animals. For each sample at least three independent digests and measurements were performed. 4,8 % mean standard deviation was obtained for all nucleosides excluding Am. 5,0 % mean standard deviation was obtained for Am.

Pork Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G+Gm	m^6t^6A	m^1A
Heart	Mean value	45.9	175.4	29.0	372.9	673.1	29.6	12.0	638.2	11.0	810.9
	Standard deviation	3.3	8.2	2.7	22.7	61.3	0.0	0.0	3.0	0.4	19.1
	Standard deviation in %	7.2	4.7	9.4	6.1	9.1	0.0	0.3	0.5	3.5	2.4
Liver	Mean value	20.5	214.6	39.6	562.7	1,176.5	55.2	33.6	1,059.6	21.9	941.6
	Standard deviation	4.8	6.4	2.7	26.1	54.7	4.6	2.9	43.6	1.8	23.6
	Standard deviation in %	23.2	3.0	6.8	4.6	4.7	8.4	8.7	4.1	8.0	2.5
Kidney	Mean value	68.4	177.8	30.7	451.5	896.0	32.3	21.1	816.3	15.8	871.6
	Standard deviation	19.1	7.7	0.5	25.4	55.0	13.8	4.7	40.3	1.5	35.5
	Standard deviation in %	28.0	4.3	1.8	5.6	6.1	42.8	22.3	4.9	9.5	4.1
Spleen	Mean value	80.4	182.1	30.7	464.3	793.6	42.7	26.1	690.5	13.9	833.6
	Standard deviation	12.6	8.9	1.0	19.6	29.5	5.9	2.2	13.4	0.4	45.2
	Standard deviation in %	15.6	4.9	3.2	4.2	3.7	13.8	8.4	1.9	2.9	5.4
Lung	Mean value	58.5	184.1	33.6	427.2	879.3	33.9	21.7	793.6	13.9	876.3
	Standard deviation	16.3	8.5	2.9	27.0	45.4	11.3	1.5	28.2	1.1	16.1
	Standard deviation in %	27.8	4.6	8.5	6.3	5.2	33.3	7.0	3.6	7.9	1.8
Spine marrow	Mean value	61.8	198.6	36.0	457.0	970.7	50.8	26.0	799.5	11.8	911.0
	Standard deviation	1.7	2.7	0.1	2.8	13.1	0.3	0.1	1.5	0.1	11.4
	Standard deviation in %	2.7	1.4	0.3	0.6	1.3	0.6	0.3	0.2	1.0	1.2
Cerebellum	Mean value	67.8	237.3	41.7	566.8	1,036.4	83.5	29.0	829.4	17.5	1,102.8
	Standard deviation	6.1	0.6	0.3	1.2	1.2	4.8	1.9	13.8	0.0	6.7
	Standard deviation in %	9.0	0.2	0.6	0.2	0.1	5.7	6.5	1.7	0.2	0.6

Pork Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G+Gm	m^6t^6A	m^1A
Cerebrum	Mean value	139.9	165.8	26.1	406.2	682.2	57.2	20.3	586.8	10.5	732.6
	Standard deviation	15.9	2.8	1.5	7.3	34.6	3.8	1.2	0.1	0.6	26.3
	Standard deviation in %	11.3	1.7	5.9	1.8	5.1	6.6	5.7	0.0	5.9	3.6

Table S4. Modification numbers per 1000 tRNAs in different porcine tissues. The table lists the average values calculated from the modification content of two different animals. From each animal at least two independent samples were investigated. For each sample at least three independent digests and measurements were performed. 5.9 % mean standard deviation was obtained for all nucleosides excluding Am. 14.7 % mean standard deviation was obtained for Am.

Pork Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G+Gm	m^6t^6A	m^1A
Heart 1	Mean value	48.2	169.6	30.9	356.8	629.8	29.6	12.0	636.0	11.2	797.4
	Standard deviation	9.7	6.1	1.1	4.0	18.3	1.1	0.5	34.6	0.5	10.4
	Standard deviation in %	20.1	3.6	3.7	1.1	2.9	3.9	3.8	5.4	4.2	1.3
Heart 2	Mean value	43.5	181.3	27.1	389.0	716.5	29.6	12.1	640.3	10.7	824.5
	Standard deviation	12.1	7.1	3.0	40.7	85.2	0.9	1.1	38.8	1.1	28.5
	Standard deviation in %	27.8	3.9	11.0	10.5	11.9	3.1	9.0	6.1	10.5	3.5
Liver 1	Mean value	23.8	210.1	37.7	544.2	1,137.8	58.5	35.7	1,028.7	20.6	958.3
	Standard deviation	2.2	2.2	0.4	18.1	0.8	4.6	1.6	7.7	0.8	39.4
	Standard deviation in %	9.4	1.0	1.1	3.3	0.1	7.9	4.6	0.8	3.7	4.1
Liver 2	Mean value	17.1	219.1	41.5	581.1	1,215.2	51.9	31.5	1,090.4	23.1	925.0
	Standard deviation	1.2	12.1	1.9	39.5	86.9	1.9	3.4	39.5	0.9	30.0
	Standard deviation in %	6.8	5.5	4.6	6.8	7.2	3.7	10.9	3.6	3.9	3.2
Kidney 1	Mean value	81.9	172.4	30.3	433.5	857.1	22.5	17.8	787.9	14.8	846.5
	Standard deviation	23.7	19.1	2.8	30.1	89.1	2.1	2.1	43.0	2.0	25.5
	Standard deviation in %	28.9	11.1	9.4	6.9	10.4	9.4	11.6	5.5	13.3	3.0
Kidney 2	Mean value	54.8	183.3	31.1	469.5	934.9	42.1	24.5	844.8	16.9	896.7
	Standard deviation	15.8	2.9	2.9	3.7	10.7	0.1	3.8	3.8	0.4	40.4
	Standard deviation in %	28.8	1.6	9.2	0.8	1.1	0.2	15.4	0.4	2.5	4.5
Spleen 1	Mean value	89.3	175.8	30.0	450.4	772.8	38.6	24.5	681.0	13.6	801.7
	Standard deviation	6.7	2.6	0.0	14.4	16.6	0.7	2.2	1.6	0.4	6.6
	Standard deviation in %	7.5	1.5	0.1	3.2	2.1	1.7	9.1	0.2	3.0	0.8

Pork Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G+Gm	m^6t^6A	m^1A
Spleen 2	Mean value	71.5	188.4	31.4	478.2	814.5	46.9	27.6	699.9	14.2	865.5
	Standard deviation	11.2	9.7	1.9	30.7	76.0	0.7	0.3	87.3	1.5	102.8
	Standard deviation in %	15.6	5.1	5.9	6.4	9.3	1.5	1.0	12.5	10.5	11.9
Lung 1	Mean value	47.0	178.1	31.6	408.0	847.2	25.9	20.7	773.7	13.2	864.9
	Standard deviation	1.2	4.3	0.1	5.3	3.2	4.6	0.7	0.1	1.2	3.7
	Standard deviation in %	2.5	2.4	0.3	1.3	0.4	17.6	3.5	0.0	8.8	0.4
Lung 2	Mean value	70.0	190.1	35.7	446.3	911.3	41.9	22.8	813.6	14.7	887.7
	Standard deviation	5.0	5.0	2.2	19.5	32.7	0.5	0.6	24.6	0.2	25.3
	Standard deviation in %	7.2	2.6	6.3	4.4	3.6	1.3	2.5	3.0	1.3	2.8
Spine marrow 1	Mean value	63.0	200.5	36.1	459.0	961.4	51.0	26.1	798.5	11.9	919.1
	Standard deviation	7.6	4.3	1.3	8.3	28.4	3.1	1.1	17.2	0.8	56.6
	Standard deviation in %	12.1	2.1	3.5	1.8	3.0	6.1	4.2	2.2	7.1	6.2
Spine marrow 2	Mean value	60.7	196.7	35.9	454.9	979.9	50.6	26.0	800.6	11.7	903.0
	Standard deviation	17.5	2.8	1.4	14.3	20.1	2.2	2.0	2.8	0.6	25.5
	Standard deviation in %	28.9	1.4	3.8	3.2	2.1	4.3	7.6	0.3	5.4	2.8
Cerebellum 1	Mean value	72.1	237.7	41.9	567.6	1,035.6	86.8	30.3	839.1	17.6	1,098.1
Cerebellum 2	Mean value	63.5	236.9	41.5	566.0	1,037.2	80.1	27.6	819.6	17.5	1,107.5
Cerebrum 1	Mean value	151.1	163.8	25.0	401.0	657.7	59.9	21.2	586.9	10.1	714.1
	Standard deviation	27.9	10.8	1.7	30.5	44.2	5.1	1.0	22.3	0.8	61.6
	Standard deviation in %	18.5	6.6	6.7	7.6	6.7	8.6	4.7	3.8	8.2	8.6

Pork Tissues		Am	t^6A	i^6A	$m^2{}_2G$	m^2G	Q	OHyW	m^1G+Gm	m^6t^6A	m^1A
Cerebrum 2	Mean value	128.7	167.8	27.2	411.3	706.7	54.6	19.5	586.7	11.0	751.2
	Standard deviation	0.1	7.8	0.7	6.1	19.0	1.5	0.7	18.7	0.4	38.1
	Standard deviation in %	0.1	4.6	2.5	1.5	2.7	2.8	3.8	3.2	3.8	5.1

Table S5. Modification numbers per 1000 tRNAs in different porcine tissues. The table lists the average values calculated from the modification content of at least two independent samples from one animal (with the exception of cerebellum, for which only one sample was analyzed). For each sample at least three independent digests and measurements were performed. 6.7 % mean standard deviation was obtained for all nucleosides excluding Am. 17.0 % mean standard deviation was obtained for Am.

Mouse Tissues		m^5C	Ψ
Liver	Mean value	1888.0	1986.9
	Standard deviation	68.3	49.4
	Standard deviation in %	3.6	2.5
Cerebrum	Mean value	1831.5	1639.2
	Standard deviation	84.2	61.8
	Standard deviation in %	4.6	3.8
Lung	Mean value	1747.8	1436.9
	Standard deviation	237.6	-
	Standard deviation in %	13.6	-
Kidney	Mean value	1316.5	924.5
	Standard deviation	81.8	65.6
	Standard deviation in %	6.2	7.1
Spleen	Mean value	1368.5	1284.7
	Standard deviation	98.1	2.7
	Standard deviation in %	7.2	0.2

Table S6. Modification numbers per 1000 tRNAs in different murine tissues. The table lists the average values calculated from the modification content of two sets of five animals (with the exception of cerebrum Ψ). For each sample at least two independent digests and measurements were performed. 5.4 % mean standard deviation was obtained for the two nucleosides. As noted in the Table S3, the overall mean standard deviation across nucleosides and tissues for the two individual sample sets is below 5 %.

Pork Tissues		m^1G	m^5C	Ψ
Liver	Mean value	746.7	2742.5	2263.1
	Standard deviation	3.2	253.6	347.0
	Standard deviation in %	0.4	9.2	15.3
Spleen	Mean value	627.1	2687.5	2146.1
	Standard deviation	7.8	182.1	161.5
	Standard deviation in %	1.2	6.8	7.5
Kidney	Mean value	628.8	2764.5	1886.8
	Standard deviation	10.6	217.0	270.2
	Standard deviation in %	1.7	7.9	14.3
Lung	Mean value	575.8	2452.7	1907.6
	Standard deviation	8.8	307.1	163.9
	Standard deviation in %	1.5	12.5	8.6
Heart	Mean value	535.1	2138.1	1776.1
	Standard deviation	3.1	269.5	203.0
	Standard deviation in %	0.6	12.6	11.4
Cerebrum	Mean value	448.5	1936.4	1330.4
		2.6	318.4	48.0
		0.6	16.4	3.6

Table S7. Modification numbers per 1000 tRNAs in different porcine tissues. The table lists the average values calculated from at least two independent digests and measurements. 7.3 % mean standard deviation was obtained for the three nucleosides. As noted in the Table S5, the overall mean standard deviation across nucleosides and tissues for individual sample sets is around 7 %.

		average modification level	Am
Mouse tissues	Liver	1	0,135
	Cerebrum	0,970	0,168
Pork tissues	Liver	0,910	0,147
Pork tissues	Cerebellum	0,900	0,485
	Cerebellum	0,880	0,148
	Lung	0,855	0,205
	Brain stem	0,851	0,138
Pork tissues	Spine marrow	0,746	0,442
	Heart	0,713	0,188
	Kidney	0,704	0,150
Pork tissues	Kidney	0,694	0,489
Pork tissues	Lung	0,691	0,418
Pork tissues	Spleen	0,689	0,575
	Spleen	0,676	0,342
Pork tissues	Cerebrum	0,608	1
Pork tissues	Heart	0,570	0,328

Table S8. Color-coded table listing the average modification levels and Am content in porcine and murine tissues. Values for each modification (listed in Fig. 2*A* and 2*B*) were normalized for the highest value across both organisms. The average modification level was calculated by averaging the normalized modification values.

Number	tRNA fragments	Liver (modified/unmodified)	Heart (modified/unmodified)
1	A-m^2G-Cp	1.00	0.62
2	G-m^2G-Up	0.49	0.41
3	m^1G-m^2G-Cp	0.59	0.35
4	t^6A-ACp	1.28	0.85
5	A-m^1A-AUp	20.3	3.69
6	G-m^1A-AACp	6.01	2.20
7	A-i^6A-ACp	1.07	0.14
8	G-m^1A-GCp	0.25	0.24
9	G-m^1A-Up	0.27	0.28
10	A-ms^2i^6A-AGCp	Traces (<0.1)	0.33

Table S9. Representative modified tRNA fragments analyzed after RNase A digestion of porcine tRNA, calculated using Sprinzl tRNA database [8] sequences and the Mongo Oligo Mass Calculator program (http://library.med.utah.edu/masspec/mongo.htm). The mass area ratios of modified to unmodified fragments from total tRNA of liver and heart are shown using the z=-2 peak. Analyzed fragments 1-7 clearly represent higher modification in liver than in heart. Fragments 8 and 9 have similar modified to unmodified ratios. Fragment 10 represents a mitochondrial tRNA fragment containing m^2i^6A, which is present in heart and only in traces in liver. These relative non-quantitative data are in strict accordance to the quantitative values described in Fig. 2.

Intra-assay	$A(m^1A)$	$A([D_3]m^1A)$	m^1A /1000 tRNAs	$A(i^6A)$	$A([D_2]i^6A)$	i^6A /1000 tRNAs
1	4,910,905	3,228,128	231.7	20,679,891	24,434,161	81.5
2	4,610,522	3,000,111	234.1	22,055,338	26,075,813	81.5
3	4,519,200	2,977,821	231.1	22,851,651	26,747,503	82.3
4	4,296,971	2,795,390	234.1	22,336,638	26,299,470	81.8
5	4,256,358	2,959,025	218.9	23,127,819	27,224,872	81.8
Mean value	4,518,791	2,992,095	230.0	22,210,267	26,156,364	81.8
RSD %	5.2	4.6	2.5	3.8	3.6	0.4

Intra-assay	$A(ms^2i^6A)$	$A([D_3]ms^2i^6A)$	ms^2i^6A /1000 tRNAs	$A(m^1G)$	$A([D_3]m^1G)$	m^1G /1000 tRNAs
1	7,565,152	6,021,535	34.4	1,555,016	1,882,259	135.5
2	8,516,277	6,814,035	34.2	1,629,176	1,901,322	140.8
3	8,736,921	6,919,317	34.6	1,564,160	1,880,098	136.5
4	8,595,797	6,922,770	34.0	1,530,903	1,860,304	134.9
5	9,057,064	7,301,879	34.0	1,614,978	1,842,588	144.2
Mean value	8,494,242	6,795,907	34.3	1,578,847	1,873,314	138.4
RSD %	5.9	6.2	0.7	2.4	1.1	2.6

Inter-assay	$A(m^1A)$	$A([D_3]m^1A)$	m^1A /1000 tRNAs	$A(i^6A)$	$A([D_2]i^6A)$	i^6A /1000 tRNAs
1	5,203,410	3,331,919	237.9	14,545,585	17,095,570	82.0
2	2,987,553	1,958,336	232.3	12,575,227	14,579,015	83.2
3	3,225,581	2,095,535	234.4	12,151,803	14,540,493	80.4
4	1,951,525	1,265,851	234.8	10,028,667	11,761,222	82.2
5	4,256,358	2,959,025	218.9	23,127,819	27,224,872	81.8
6	3,249,988	2,223,235	222.5	21,114,201	24,952,572	81.5
Mean value	3,479,069	2,305,650	230.1	15,590,550	18,358,957	81.8
RSD %	29.4	29.3	6.3	31.0	31.1	1.0

Inter-assay	$A(ms^2i^6A)$	$A([D_3]ms^2i^6A)$	ms^2i^6A /1000 tRNAs	$A(m^1G)$	$A([D_3]m^1G)$	m^1G /1000 tRNAs
1	5,974,676	4,846,268	33.7	926,498	1,054,171	144.6
2	5,461,095	4,327,040	34.6	705,140	894,683	128.9
3	3,881,613	3,101,702	34.3	641,656	776,294	135.5
4	3,758,364	2,908,332	35.5	439,096	544,671	132.0
5	9,057,064	7,301,879	34.0	1,614,978	1,842,588	144.2
6	8,707,594	6,786,037	35.2	1,452,283	1,713,325	139.2
Mean value	6,140,068	4,878,543	34.5	963,275	1,137,622	137.4
RSD %	34.1	34.4	1.8	44.7	42.1	4.3

Table S10. Intra- and Inter-assay test of representative nucleosides m^1A, i^6A, ms^2i^6A and m^1G.

Data set	Mean	Standard deviation	n	Shapiro Wilk W	Shapiro Wilk significance
Normalized murine nucleosides	0.87	0.15	4	0.90	0.46
Normalized porcine nucleosides	0.75	0.13	6	0.85	0.15
Protein synthesis rate (this study, all murine tRNA)	0.38	0.29	11	0.94	0.52
Protein synthesis rate (this study, cytosolic porcine tRNA)	0.75	0.24	6	0.93	0.57
Protein synthesis rate (this study, all porcine tRNA)	0.84	0.13	6	0.96	0.82
Protein synthesis rate (this study, all porcine tRNA excluding heart)	0.83	0.15	5	0.95	0.70
Protein Synthesis rate (Suryawan, %/d)	35.5	20.5	5	0.878	0.26

Table S11: Descriptive Statistics. All calculations were carried out using the SPSS or Statistica statistics programs. Correlations are calculated as Pearson's r coefficients with 2-tailed significance values. This test was chosen as the data sets used in correlations appear to be parametric as evidenced by non-significant scores in the Shapiro-Wilk test. The small sample sizes are recognized as a possible source of error and correspondingly it cannot be excluded that more complex relationships than simple linear correlations may be present.

Protein synthesis rate	r	Significance (P value)	n
This study (cytosolic porcine tRNA)	0.861*	0.028	6
This study (all porcine tissues, total tRNA)	0.672	0.143	6
This study (all porcine tissues except heart, total tRNA)	0.965**	0.008	5
This study (all murine tissues, total tRNA)	0.723*	0.011	11
Suryawan	0.956*	0.011	5

Table S12. Correlations of normalized nucleoside levels (excluding ms^2i^6A) with protein synthesis rates. * significant to 5 % level, ** significant to 1 % level.

Supplementary Information References

[1] T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2009**, *48*, 7932-7934.

[2] D. Globisch, D. Pearson, A. Hienzsch, T. Brückl, M. Wagner, I. Thoma, P. Thumbs, V. Reiter, A. C. Kneuttinger, M. Müller, S. A. Sieber, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 9739-9742.

[3] V. Reiter, D. M. S. Matschkal, M. Wagner, D. Globisch, A. C. Kneuttinger, M. Müller, T. Carell, *Nucleic Acids Res.* **2012**, *40*, 6235-6240.

[4] R. J. Jackson, S. Napthine, I. Brierley, *RNA* **2001**, *7*, 765-773.

[5] J. B. Plotkin, H. Robins, A. J. Levine, *Proc. Natl. Acad. Sci. U S A* **2004**, *101*, 12588-12591.

[6] A. Suryawan, P. M. J. O'Connor, J. A. Bush, H. V. Nguyen, T. A. Davis, *Amino Acids* **2009**, *37*, 97-104.

[7] C. T. Y. Chan, M. Dyavaiah, M. S. DeMott, K. Taghizadeh, P. C. Dedon, T. J. Begley, *PLoS Genet.* **2010**, *6*, e1001247.

[8] F. Jühling, M. Mörl, K. Hartmann Roland, M. Sprinzl, F. Stadler Peter, J. Pütz, *Nucleic Acids Res.* **2009**, *37*, D159-D162.

13.2 Supplementary material for Chapter 7

B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366-369. "*Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction.*"

Supporting Information

Synthesis of 5-hydroxymethyl-, 5-formyl- and 5-carboxycytidine-triphosphates and their incorporation into oligonucleotides by PCR

Barbara Steigenberger,[#] Stefan Schiesser,[#] Benjamin Hackner, Caterina Brandmayr, Silvia K. Laube, Jessica Steinbacher, Toni Pfaffeneder, and Thomas Carell[*]

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

e-mail: Thomas.Carell@lmu.de

General Information S2

Procedures, Analytical and Spectroscopic Data S6

General methods:

All non-aqueous reactions were performed using flame- or oven-dried glassware under an atmosphere of dry nitrogen. Commercial reagents from *Sigma-Aldrich* or *Acros* were used as received unless otherwise noted. Non-aqueous reagents were transferred under nitrogen with a syringe or cannula. Solutions were concentrated *in vacuo* on a *Heidolph* rotary evaporator with a *Vario PC2001* diphragm pump by *Vacuubrand*. Chromatographic purification of products was accomplished using flash column chromatography on *Merck* Geduran Si 60 (40–63 µm) silica gel (normal phase). Thin layer chromatography (TLC) was performed on *Merck* 60 (silica gel F_{254}) plates. Visualization of the developed chromatogram was performed using fluorescence quenching or anisaldehyde staining. ^{1}H-, ^{13}C- and ^{15}N-NMR spectra were recorded in deuterated solvents on *Bruker ARX 300*, *Varian VXR400S*, *Varian Inova 400* and *Bruker AMX 600* spectrometers and calibrated to the residual solvent peak. High-resolution ESI spectra were obtained on the mass spectrometer *Thermo Finnigan* LTQ FT-ICR. IR measurements were performed on a *Perkin Elmer Spectrum BX FT-IR* spectrometer (*Perkin Elmer*) with a diamond-ATR (Attenuated Total Reflection) setup. Melting points were determined with a *Büchi Melting Point B540*. The concentration of the purified DNA was determined with a *NanoDrop ND-1000* spectrophotometer (*Peqlab*). Extinction coefficients at 260 nm were calculated by addition of the extinction coefficients of the individual nucleobases. These are dA 15.0 L/mmol·cm, dC 7.1 L/mmol·cm, dG 12.0 L/mmol·cm and dT 8.4 L/mmol·cm. For the dC derivatives, the dC value was used. Nuclease S1 (*Aspergillus oryzae*) was obtained from *Roche*, snake venom phosphodiesterase I (*Crotalus adamanteus*) from *USB corporation* and antarctic phosphatase from *New England Biolabs*.

High performance liquid chromatography:

HPLC was performed on *Waters* or *Merck-Hitachi* units. These were in detail: analytical HPLC: *Waters* Alliance (2695 Separation Module, 2996 Photodiode Array Detector), *Merck* analytical (L-7400 UV detector, L-7100 pump), preparative HPLC: *Waters* Breeze (2487 Dual λ Array Detector, 1525 Binary HPLC Pump), *Merck* preparative (L-7150 pump, L-7420 UV detector, *Rheodyne* P/N 77 25i injection valve, ERC-3415 solvent degasser). For analytical HPLC Nucleosil 120-3 C18 from *Macherey Nagel* were used, for preparative HPLC Nucleosil 100-7 C18, VP 250/10 C18 and VP 250/32 C18 also from *Macherey Nagel* were used.

Buffer systems: Buffer A: 0.1 M triethylammonium acetate

Buffer B: 0.1 M triethylammonium acetate in 80% MeCN

Buffer system for HPLC-MS and UPLC-MS/MS: Buffer A: 0.01% formic acid in H_2O

Buffer B: 0.01% formic acid in MeCN

Polymerase chain reaction:

The PGL3_pOct4_eGFP template was a generous gift of *H. Leonhart*.[3] The primers were obtained from *Metabion* and had the following sequences, which resulted in a 150 bp fragment:

Forward Primer: 5′-TCCCGTCCTAAGGGTTGTCCTGTC-3′

Reverse Primer: 5′-ACCCTCTAGCCTTGACCTCTGGC-3′

Thermostable pyrophosphatase and natural triphosphates were obtained from *New England Biolabs*. KOD XL-polymerase was obtained from *Novagen*, and Vent (exo$^-$) from *New England Biolabs*. PCRs were carried out on *Eppendorf realplex*^{4} thermocycler and analyzed by 1–1.5% agarose gels that were run with a horizontal cell (Sub-Cell *Bio-Rad*) at 110 V using ethidium bromide for staining.

The protected D_2,$^{15}N_2$-hmC containing DNA fragments were deprotected by shaking in 0.1 M NaOH in a water/methanol 1:4 mixture at room temperature for 1 h.

Prior to analysis the reactions were cleaned up using *Nucleo Spin*® clean up Kits obtained from *Macherey Nagel*.

A typical PCR in a total volume of 50 µl contained 5 ng of template, 0.25 µM of each Primer, 2.5 U KOD XL-polymerase for cadCTP and D_2,$^{15}N_2$-hmC, respectively, 2.0 U Vent (exo$^-$) for fdCTP, 2.0 U of thermostable pyrophosphatase and 200 µM of each dNTP (corresponding natural dNTP substituted by modified triphosphates) and was incubated according to the following protocols.

Table S1. PCR program for amplification with D_2,$^{15}N_2$-hmC:

Step		duration	Temperature
initial denaturation		2 min	95°C
30 cycles	denaturation	15 s	95 °C
	annealing	15 s	56 °C
	extension	30 s	72 °C
final extension		5 min	72 °C

Table S2. PCR program for amplification with dfCTP:

Step		duration	Temperature
initial denaturation		2 min	95 °C
30 cycles	denaturation	15 s	95 °C
	annealing	15 s	55 °C
	extension	15 s	75 °C
final extension		5min	75 °C

Table S3. PCR program for amplification with cadCTP

Step		duration	Temperature
initial denaturation		2 min	95 °C
31 cycles	denaturation	30 s	95 °C
	annealing	20 s	55 °C
	extension	15 s	72 °C
final extension		5 min	72 °C

LC-MS analysis of PCR products:

Three independent PCRs were carried out with each modified dNTP and analyzed by LC-MS based on a further development, for which we are currently finishing a manuscript.1,2 In the following we shortly summarize the parameters of the method.

LC-MS/MS analysis was performed on an *Agilent* 6490 triple quadrupol mass spectrometer coupled to an Agilent 1290 UHPLC system with UV-detection. The transitions of the nucleosides were analyzed in the positive ion selected reaction monitoring mode (SRM) operating under unit mass resolution conditions (Table S4).

Table S4. Compound-dependent parameters for LC-MS/MS.

Compound	**Precursor Ion**	**Product Ion**
5-caC	272.09	156.04
5-fC	256.09	140.05

For the analysis of biological samples, we used a C8 column from *Agilent* (1.8 µm, 2.1 mm × 150 mm). The compounds were separated by a gradient using water and acetonitrile with 0.0075% formic acid. The column temperature was maintained at 30 °C. The flow rate was 400 µL min^{-1} and the injection volume 29 µL.

LC-HRMS-analysis with isotope labeled triphosphates were carried out using a *Thermo Finnigan* LTQ Orbitrap XL and chromatographed by a *Dionex* Ultimate 3000 HPLC system. The flow was set to 0.15 mL/min over an Uptisphere120-3HDO (3 µm, 2.1 mm × 15 mm) column from *Interchim*.

Results of LC-HRMS analysis:

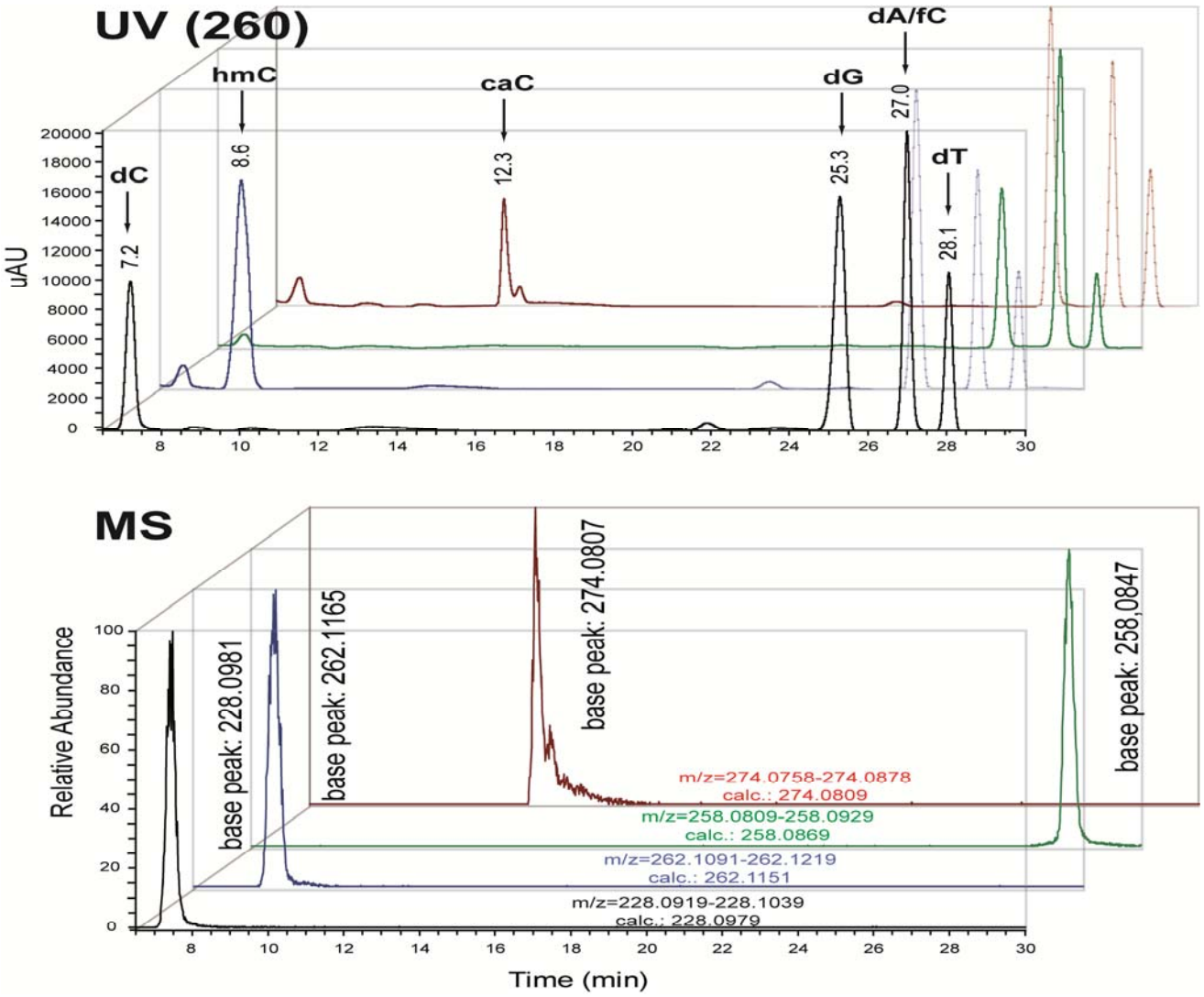


Figure S1: UV 260 nm and mass traces of the total digests of the $^{15}N_2$-labeled xdCTP measured with HRMS. Positive control with dCTP (black), D_2,$^{15}N_2$-hmdCTP (blue), $^{15}N_2$-fdCTP (green) and $^{15}N_2$-cadCTP (red).

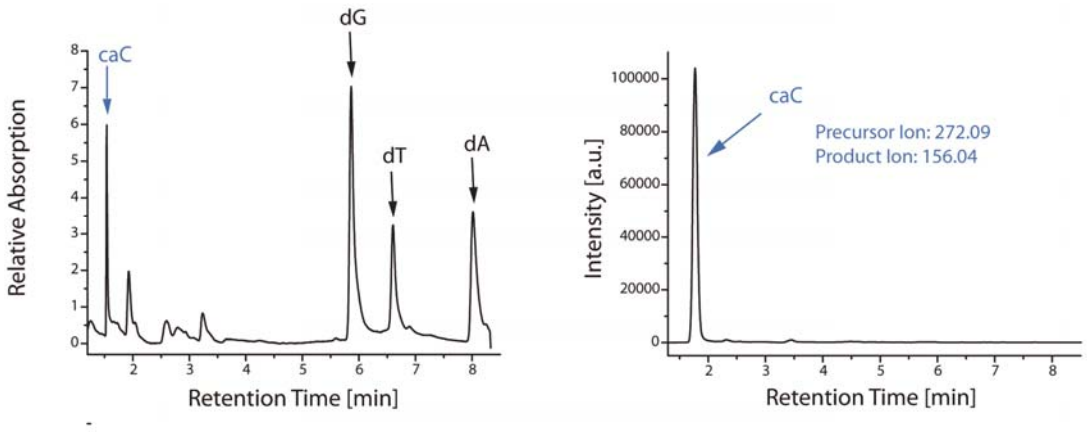
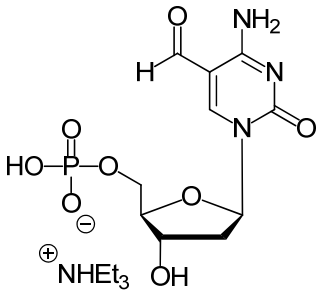


Figure S2: UV 260 nm and mass traces of the total digests of the cadCTP measured with HRMS.

Synthesis of hmC, fC and caC-triphosphates:

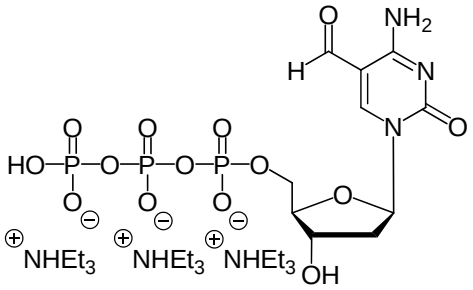
5-Formyl-2´-deoxy-cytidine-5´-monophosphate triethylammonium salt (2)



5-Formyl-2´-deoxy-cytidine (**1)** (13 mg, 0.051 mmol, 1.0 eq.), and proton sponge (16.4 mg, 0.076 mmol, 1.5 eq.) were dissolved in dry trimethylphosphite (1 mL). The solution was cooled to 0 °C and phosphorus oxychloride (5.4 µL, 0.061 mmol, 1.2 eq.) was added. After stirring for 3 h the reaction was quenched by addition of triethylammonium bicarbonate buffer (3 mL, pH 8). RP-HPLC purification (0% → 20% B in 45 min) gave **2** as a colorless solid (6.4 mg, 0.014 mmol, 29 %).

^{1}H-NMR (200 MHz, D_2O): δ (ppm) = 9.49 (s, 1H, C**H**O), 8.79 (s, 1H, N-C**H**=C), 6.17 (t, *J*=6.1 Hz, 1H, O-C**H**-N), 4.39–4.42 (m, 1H, C**H**-O), 4.10–4.12 (m, 1H, C**H**-CH_2-O), 3.71–3.87 (m, 2H, C**H$_2$**-O), 2.24–2.61 (m, 1H, N-CH-C**H$_2$**), 2.39–2.46 (m, 1H, N-CH-C**H$_2$**); **^{31}P-NMR** (162 MHz, D_2O): δ (ppm) = 0.2 (s); **HRMS** (ESI-): calc. for $C_{10}H_{13}O_8N_3P^-$ $[M\text{-}H^+]^-$: 334.0446, found: 334.0449.

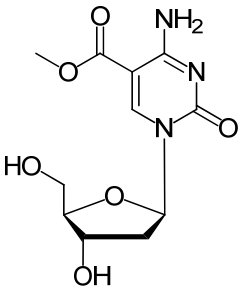
5-Formyl-2´-deoxy-cytidine-5´-triphosphate triethylammonium salt (fdCTP)



Dry tributylamine (13 µL, 56 µmol, 4.7 eq.) and 1-Methyl-3-benzenesulfonylimidazolium triflate (5.3 mg, 14 µmol, 1.2 eq.) was added to a stirred solution of 5-formyl-2´-deoxy-cytidine-5´-monophosphate **2** (4 mg, 12.0 µmol, 1.0 eq) in dry DMF (0.45 mL) at 0 °C. The resulting colorless solution was added to pyrophosphate (8.7 mg, 24 µmol, 2.0 eq.) in dry DMF (0.45 mL). The ice bath was removed and the reaction mixture stirred at room temperature for 3 h. The mixture was quenched by addition of triethylammonium acetate buffer (1 M) and the solvent was removed *in vacuo*. The triphosphate was isolated by ion exchange chromatography at 4 °C unsing a DEAE-cellulose column with a gradient from 100% water to 0.5 M TEAB (pH 7.5). Further purification was achieved by FPLC (0.1 M, 1 M TEAB, 0-100% TEAB over 30 min) by using a MonoQ 5/50 GL anion exchange column (GE) gave **fdCTP** as a colorless solid (6.9 mg, 8.4 µmol, 70 %).

^{1}H-NMR (400 MHz, D_2O): δ (ppm) = 9.53 (s, 1H, C**H**O), 8.74 (s, 1H, N-C**H**=C), 6.11 (t, *J*=5.9 Hz, 1H, O-C**H**-N), 4.48–4.58 (m, 1H, O-C**H**), 4.13–4.19 (m, 3H, O-CH-C**H**-C**H$_2$**-O), 2.26–2.46 (m, 2H, O-CH-C**H$_2$**-CH-N); **^{31}P-NMR** (162 MHz, D_2O, ppm): δ = -10.5 (m, P_γ), -11.7 (d, *J*=19.8 Hz, P_α), -23.3 (t, *J*=16.4 Hz, P_β); **HRMS** (ESI+): calc. for $C_{11}H_{15}N_3P_3O_{14}$ $[M\text{-}H^+]^-$: 493.9927, found: 493.9920.

5-Carboxylmethyl-2´-deoxy-cytidine (4)

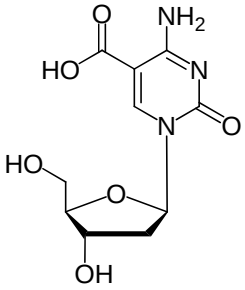


In a polypropylene tube 5-carboxylmethyl-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxy-cytidine (**3**) (266 mg, 0.52 mmol, 1.0 eq.) was dissolved in 30 mL EtOAc and HF-pyridine (70% HF, 270 µL, 7.8 mmol, 5.7 eq.) was added. After stirring the reaction mixture for 14 h at room temperature, additional HF-pyridine (100 µL, 1.1 mmol, 2 eq.) was added and the reaction mixture was stirred for 5 h

at room temperature. The reaction was quenched by addition of methoxytrimethylsilane (2.0 mL, 14 mmol, 27 eq.). After stirring for 30 min, the precipitate was collected by centrifugation (4500 rpm, 20 min). The crude product was purified by column chromatography (CH_2Cl_2/MeOH, 10:1 → 4:1) to yield 5-carboxylmethyl-2´-deoxy-cytidine **4** as a colorless solid (109 mg, 0.38 mmol, 73%).

^{1}H-NMR (400 MHz, CD_3OD) δ = 9.13 (s, 1H, N-C**H**=C), 6.19 (t, *J*=5.9 Hz, 1H, N-C**H**-O), 4.36– 4.39 (m, 1H, C**H**-OH), 4.06 (q, *J*=6.8 Hz, C**H**-CH_2-O), 3.92–3.88 (m, 5H, C$\mathbf{H_2}$-OH, OC$\mathbf{H_3}$), 2.44–2.49 (m, 1H, N-CH-C$\mathbf{H_2}$), 2.18–2.25 (m, 1H, N-CH-C$\mathbf{H_2}$);**^{13}C-NMR** (100 MHz, CD_3OD) δ = 166.6 (**C**OO), 165.2 (**C**-NH_2), 156.5 (N-**C**O-N), 149.9 (**C**H=C), 96.9 (**C**-CO), 89.2 (N-**C**H-CH_2), 88.5 (**C**H-CH_2-O), 71.2 (**C**H-OH), 62.28 (**C**H_2-OH), 52.4 (O**C**H_3), 42.6 (O-**C**H-CH_2); **HRMS** (ESI+): calc. for $C_{11}H_{16}N_2O_6$, $[M+H]^+$: 286.1034, found: 286.1032; **Melting range:** 160–170 °C; **IR** (cm^{-1}): ν = 3402 (m), 2924 (m), 1712 (s), 1632 (ss), 1501 (s), 1438 (s), 1319 (s), 1256 (s), 1191 (s), 1092 (ss), 1059 (ss), 991 (s), 991 (s), 870 (m), 790 (ss), 721 (s), 693 (s).

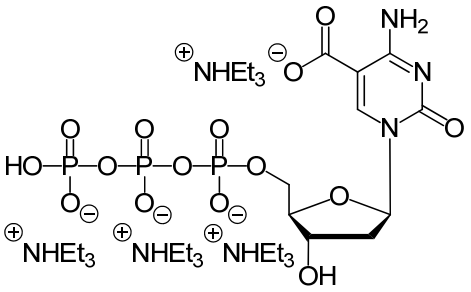
5-Carboxyl-2´-deoxycytidine (5)



5-Carboxylmethyl-2´-deoxy-cytidine (**4**) (139 mg, 487.0 mmol, 1.0 eq.) was dissolved in a mixture of water (7.2 mL) and MeCN (38.0 mL). Then LiOH (139 mg, 5.8 mmol, 12.0 eq.) was added and the mixture was stirred at room temperature for 19 h. After adjusting to pH 4.0 with 2 M hydrochloric acid, **5** precipitated as a colorless solid (58 mg, 0.214 µmol, 44%).

^{1}H-NMR (400 MHz, D_2O): δ (ppm) = 8.81 (s, 1H, N-C**H**=C), 6.30 (t, *J*=6.2 Hz, 1H, N-C**H**-O), 4.51–4.53 (m, 1H, C**H**-OH), 4.18–4.15 (m, 1H, C**H**-CH_2-O), 3.92 (dd, *J*=12.7 Hz, 3.4 Hz, 2H, C$\mathbf{H_2}$-OH), 2.60–2.40 (m, 2H, N-CH-C$\mathbf{H_2}$); **^{13}C-NMR** (100 MHz, D_2O): δ =167.9 (**C**O-O), 159.6 (**C**-NH_2), 149.7 (**C**=O-N), 148.2 (**C**H=C), 101.2 (**C**=CO), 87.2 (**C**H-CH_2-O), 86.9 (N-**C**H-CH_2), 70.1 (**C**H-OH), 60.8 (**C**H_2-OH), 39.5 (O-**C**H-CH_2); **HRMS** (ESI-): calc. for $C_{10}H_{12}N_3O_6$ $[M-H^+]^-$: 270.0732, found: 270.0732; **melting range**: > 250 °C decomposition; **IR** (cm^{-1}): ν = 3254 (m), 2934 (m), 1718 (s), 1653 (m), 1540 (m), 1437 (s), 1363 (s), 1279 (s), 1235 (m), 1198 (s), 1090 (s), 1050 (ss), 995 (s), 869 (m), 783 (s), 758 (s), 732 (s), 684 (ss).

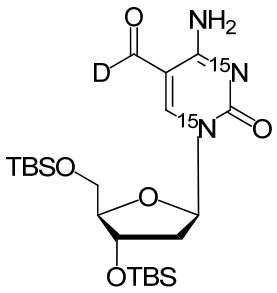
5-Carboxyl-2´-deoxy-cytidine-5´-triphosphate triethylammonium salt (cadCTP)



Tributylammonium pyrophosphate (37.5 mg, 0.08 mmol, 2.0 eq.) was dissolved in DMF (0.10 mL) and tributylamine (0.15 mL) was added. The resulting mixture was added to 2-chloro-1,3,2-benzodioxaphosphorin-4-one (16 mg, 0.08 mmol, 2.0 eq.) in DMF (0.11 mL). After 30 min at room temperature the stirred solution was added to 5-Carboxyl-2´-deoxycytidine (**5**) (10.5 mg, 38.4 µmol, 1.0eq.). After stirring at room temperature for 2.5 h a solution of 3 % iodine in a pyridine/water mixture (9:1, 0.3 mL) was added and stirred at room temperature for 1.5 h. The crude product was precipitated by addition of a 3 M sodium chloride solution (0.8 mL) and ethanol (15 mL) at -80 °C for 14 h. RP-HPLC purification (0% → 15% B in 75 min) gave **cadCTP** as a colorless solid (2.69 mg, 2.7 µmol, 8%).

^{1}H-NMR (400 MHz, D_2O): δ (ppm) = 8.32 (s, 1H, N-C**H**=C), 6.11 (t, *J*=5.6 Hz, 1H, O-C**H**-N), 4.46–4.47 (m, 1H, C**H**-O), 4.29–4.23 (m, 3H, CH-C**H**-C**H$_2$**-O), 2.52–2.39 (m, 2H, N-CH-C**H$_2$**). **^{31}P-NMR** (162 MHz, D_2O): δ (ppm) = -10.1 (d, *J*=19.6 Hz), -11.4 (d, *J*=19.8 Hz), -23.4 (t, *J*=19.8 Hz, P_β); **HRMS** (ESI+): calc. for $C_{10}H_{17}N_3P_3O_{15}$ $[M+H^+]^+$: 511.9867, found: 511.9861.

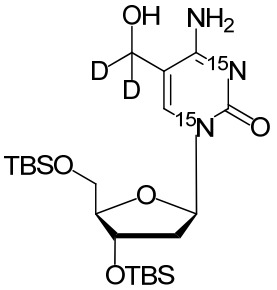
5-Deuteroformyl-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (7)



5´-Iodo-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (**6**)[3] (0.20 g, 0.34 mmol, 1.0 eq.), $Pd_2(dba)_3 \cdot CHCl_3$ (56.0 mg, 0.05 mmol, 0.2 eq.) and triphenylphosphine (85.0 mg, 0.32 mmol, 0.9 eq.) were dissolved in d^8-toluene (7.8 mL) in a high pressure glas autoclave. The autoclave was flushed with CO to remove residual air and after heating to 60 °C, the CO pressure was set to 3.5 bar. Tributyltin deuteride (0.18 mL, 0.68 mmol, 2.0 eq.) in d^8-toluene (0.82 mL) were added within 10 h and the reaction was stirred for additional 14 h at 60 °C and 3.5 bar CO pressure. After evaporation to dryness,

the product was purified by silica gel column chromatography (iHex/EtOAc 1:1) to give product **7** as an orange solid (106 mg, 0.22 mmol, 64%). **^{1}H-NMR** (300 MHz, $CDCl_3$, ppm): δ = 8.52 (s, 1H, N-C**H**=C), 8.16 (s br, 1H, N**H$_2$**), 6.32 (s br, 1H, N**H$_2$**), 6.19 (dt, *J*=6.2, 0.9 Hz, 1H, N-C**H**-O), 4.35–4.31 (m, 1H, C**H**-O-Si), 4.03 (dd, *J*=5.9, 2.6 Hz, 1H, C**H**-CH_2-O), 3.94 (dd, *J*=11.6, 2.6 Hz, 1H, C**H$_2$**-O-Si), 3.76 (dd, *J*=11.6, 2.5 Hz, 1H, C**H$_2$**-O-Si), 2.64–2.56 (m, 1H, N-CH-C**H$_2$**), 2.10–2.01 (m, 1H, N-CH-C**H$_2$**), 0.88 (s, 9H, Si-C-C**H$_3$**), 0.87 (s, 9H, Si-C-C**H$_3$**), 0.08 (s, 3H, Si-C**H$_3$**), 0.07 (s, 3H, Si-C**H$_3$**), 0.06 (s, 3H, Si-C**H$_3$**), 0.05 (s, 3H, Si-C**H$_3$**). **^{13}C-NMR** (151 MHz, $CDCl_3$, ppm): δ = 187.1 (t, *J*=26.6 Hz, COD), 162.9 (d, *J*=5.9 Hz, **C**-NH_2), 153.5 (dd, *J*=9.4, 7.2 Hz, N-**C**=O), 153.1 (d, *J*=14.1 Hz, **C**H=C), 105.1 (**C**-CO), 88.8 (O-**C**H-CH_2-O), 88.0 (d, *J*=9.5 Hz, N-**C**H-O), 71.8 (**C**H-O-Si), 62.8 (**C**H_2-O-Si), 43.2 (N-CH-**C**H_2), 26.2 (Si-C-**C**H_3), 25.9 (Si-C-**C**H_3), 18.7 (Si-**C**), 18.2 (Si-**C**), -4.3 (Si-**C**H_3), -4.7 (Si-**C**H_3), -5.0 (Si-**C**H_3), -5.1 (Si-**C**H_3). **^{15}N-NMR** (40 MHz, $CDCl_3$, ppm): δ =-172.7, -207.3. **HRMS** (ESI+): calc. for $C_{22}H_{41}DN^{15}N_2O_5Si_2$ $[M+H^+]^+$: 487.2660, found: 487.2661. **Melting range:** 137–139 °C. **IR** (cm^{-1}): ν= 3366 (w), 2954 (m), 2927 (m), 2857 (m), 1639 (s), 1504 (m), 1471 (m), 1462 (m), 1251 (m), 1084 (s), 829 (s), 775 (s).

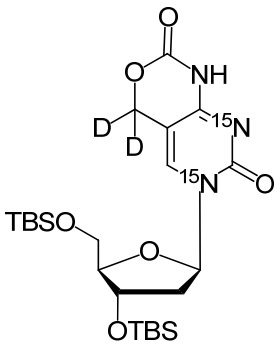
5-Dideuterohydroxymethyl-3′,5′-(*tert*-butyl-dimethylsilyl)-2′-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (8)



To 5-deuteroformyl-3′,5′-(*tert*-butyl-dimethylsilyl)-2′-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (**7**) (77.0 mg, 0.16 mmol, 1.0 eq.) in CD_3OD (5.3 mL) were added sodium borodeuteride (7.5 mg, 0.18 mmol, 1.1 eq.) and cerium(III) chloride (0.12 g, 0.49 mmol, 3.0 eq.). The resulting orange solution was stirred at room temperature for 25 min and then quenched with a saturated ammonium chloride solution (53 mL). The aqueous phase was extracted with EtOAc (53 mL), the resulting organic phase was washed with a saturated ammonium chloride solution (2 × 53 mL) and dried over magnesium sulfate. The solvent was evaporated to dryness and the product purified by silica gel column chromatography (DCM/MeOH 20:1) to obtain **8** as a colorless solid (31.0 mg, 0.06 mmol, 40%). **^{1}H-NMR** (600 MHz, $CDCl_3$, ppm): δ = 7.57 (d, *J*=1.1 Hz, 1H, N-C**H**=C), 6.10 (t, *J*=6.4 Hz, 1H, N-C**H**-O), 4.28 (dt, *J*=6.6, 3.5 Hz, 1H, C**H**-O-Si), 3.89 (q, *J*=3.1 Hz, 1H, C**H**-CH_2-O), 3.80 (dd, *J*=11.2, 3.2 Hz, 1H, C**H$_2$**-O-Si), 3.70 (dd, *J*=11.2, 3.0 Hz, 1H, C**H$_2$**-O-Si), 2.38–2.33 (m, 1H, N-CH-C**H$_2$**), 1.97–1.89 (m, 1H, N-CH-C**H$_2$**), 0.87 (s, 9H, Si-

C-C**H**$_3$), 0.85 (s, 9H, Si-C-C**H**$_3$), 0.06 (s, 3H, Si-C**H**$_3$), 0.05 (s, 3H, Si-C**H**$_3$), 0.03 (s, 3H, Si-C**H**$_3$), 0.03 (s, 3H, Si-C**H**$_3$). **^{13}C-NMR** (151 MHz, $CDCl_3$, ppm): δ = 165.5 (dd, *J*=6.1, 1.4 Hz, **C**-NH_2), 156.5 (dd, *J*=12.6, 8.1 Hz, N-**C**=O), 138.8 (d, *J*=13.2 Hz, **C**H=C), 106.1 (d, *J*=0.7 Hz, **C**-CD_2), 88.0 (O-**C**H-CH_2-O), 86.5 (d, *J*=11.2 Hz, N-**C**H-O), 71.9 (**C**H-O-Si), 62.9 (**C**H_2-O-Si), 59.1 (s br, **C**D_2), 42.4 (N-CH-**C**H_2), 26.1 (Si-C-**C**H_3), 26.0 (Si-C-**C**H_3), 18.6 (Si-**C**), 18.2 (Si-**C**), -4.3 (Si-**C**H_3), -4.7 (Si-**C**H_3), -5.1 (Si-**C**H_3), -5.2 (Si-**C**H_3). **^{15}N-NMR** (40 MHz, $CDCl_3$, ppm): δ =-173.6, -223.3. **HRMS** (ESI+): calc. for $C_{22}H_{42}D_2N^{15}N_2O_5Si_2$ $[M+H^+]^+$: 490.2880, found: 490.2878. **Melting range:** 95–97 °C. **IR** (cm^{-1}): ν= 3202 (w), 2928 (w), 2856 (w), 1657 (m), 1604 (m), 1471 (m), 1253 (m), 1075 (m), 1029 (m), 832 (s), 775 (s).

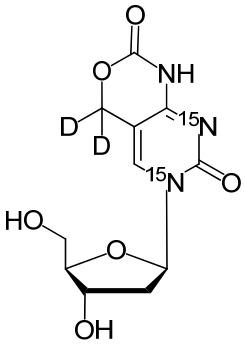
5-Dideutero-3´,5´-(*tert*-butyl-dimethylsilyl)-4,5-(1,3-[3*H*,6*H*]oxazin-2-one)-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (9)



To a solution of 5-dideuterohydroxymethyl-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (**8**) (23 mg, 46.96 µmol, 1.0 eq.) in THF (9.3 mL) were added 4-Nitrophenylchloroformate (10.6 mg, 52.59 µmol, 1.12 eq.) and DIPEA (16.5 µL, 111.3 µmol, 2.4 eq.). After stirring for 17 h at room temperature additional 4-Nitrophenylchloroformate (2.1 mg, 10.42 µmol, 0.2 eq.) and DIPEA (3.4 µL, 22.90 µmol, 0.49 eq.) were added. After stirring at room temperature for additional 4 h, the solvent was removed *in vacuo*. Purification by silica gel column chromatography (DCM/MeOH 100:1 → 80:1) gave **9** as a colorless solid (18 mg, 34.90 µmol, 74%). **^{1}H-NMR** (200 MHz, $CDCl_3$, ppm): δ = 8.16 (s, 1H, N-C**H**=C), 6.20 (t, *J*=5.9 Hz, 1H, N-C**H**-O), 4.35–4.29 (m, 1H, C**H**-O-Si), 3.99–3.89 (m, 2H, C**H**-CH_2-O, C**H**$_2$-O-Si), 3.75 (dd, *J*=11.3, 1.7 Hz, 1H, C**H**$_2$-O-Si), 2.61–2.48 (m, 1H, N-CH-C**H**$_2$), 2.10–1.96 (m, 1H, N-CH-C**H**$_2$), 0.89 (s, 9H, Si-C-C**H**$_3$), 0.86 (s, 9H, Si-C-C**H**$_3$), 0.09 (s, 3H, Si-C**H**$_3$), 0.08 (s, 3H, Si-C**H**$_3$), 0.05 (s, 3H, Si-C**H**$_3$), 0.04 (s, 3H, Si-C**H**$_3$). **^{13}C-NMR** (100 MHz, $CDCl_3$, ppm): δ = 159.7 (dd, *J*=6.9, 1.0 Hz, **C**-NH), 154.6 (dd, *J*=11.9, 6.2 Hz, N-**C**=O), 149.9 (d, *J*=2.8 Hz, O-**C**=O), 138.7 (d, *J*=13.6 Hz, **C**H=C), 96.1 (**C**-CD_2), 88.5 (O-**C**H-CH_2-O), 87.7 (d, *J*=9.8 Hz, N-**C**H-O), 71.4 (**C**H-O-Si), 64.5 (qi, *J*=23.3 Hz, CD_2), 62.6 (**C**H_2-O-Si), 42.8 (N-CH-**C**H_2), 26.1 (Si-C-**C**H_3), 25.9 (Si-

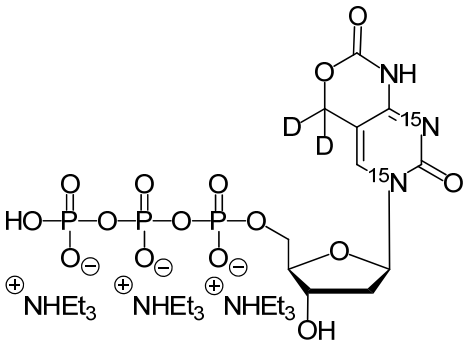
C-**C**H_3), 18.6 (Si-**C**), 18.3 (Si-**C**), -4.3 (Si-**C**H_3), -4.7 (Si-**C**H_3), -5.1 (Si-**C**H_3), -5.2 (Si-**C**H_3). **^{15}N-NMR** (40 MHz, $CDCl_3$, ppm): δ =-152.7, -210.2. **HRMS** (ESI+): calc. for $C_{23}H_{40}D_2N^{15}N_2O_6Si_2$ $[M+H^+]^+$: 516.2671, found: 516.2668. **Melting range:** 92–94 °C. **IR** (cm^{-1}): ν= 2952 (w), 2928 (w), 2855 (w), 1757 (m), 1658 (m), 1553 (m), 1471 (m), 1253 (m), 1112 (m), 1064 (m), 1028 (m), 832 (s), 774 (s).

5-Dideutero-4,5-(1,3-[3*H*,6*H*]oxazin-2-one)-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (10)



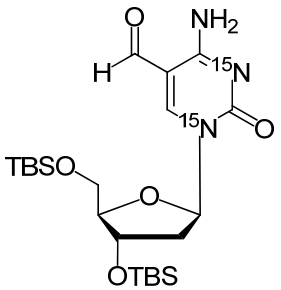
To a solution of 5-dideutero-3´,5´-(*tert*-butyl-dimethylsilyl)-4,5-(1,3-[3*H*,6*H*]oxazin-2-one)-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (**9**) (299 mg, 0.58 mmol, 1.0 eq.) in EtOAc (19 mL) were added pyridine (226 µL, 2.79 mmol, 4.8 eq.) and 70% HF-pyridine (234 µL, 8.70 mmol, 15.0 eq.). After stirring at room temperature for 19 h additional pyridine (226 µL, 2.79 mmol, 4.8 eq.) and 70% HF-pyridine (234 µL, 8.70 mmol, 15.0 eq.) were added. After stirring for additional 20 h the reaction was quenched by addition of methoxytrimethylsilane (6.3 mL, 45.70 mmol, 78.8 eq.) and stirring was continued for 30 min. After evaporation to dryness the crude product was purified by silica gel column chromatography (DCM/MeOH 10:1 → 5:1) to obtain **10** as a colorless solid (151 mg, 0.53 mmol, 91%). **^{1}H-NMR** (200 MHz, CD_3OD, ppm): δ = 8.39 (s, 1H, N-C**H**=C), 6.20 (dt, *J*=6.2, 1.3 Hz, 1H, N-C**H**-O), 4.36 (dt, *J*=6.2, 3.9 Hz, 1H, C**H**-OH), 4.03–3.97 (m, 1H, C**H**-CH_2-O), 3.89–3.70 (m, 2H, C**H_2**-OH), 2.56–2.43 (m, 1H, N-CH-C**H_2**), 2.23–2.09 (m, 1H, N-CH-C**H_2**). **^{13}C-NMR** (151 MHz, d^6-DMSO, ppm): δ = 160.1 (dd, *J*=6.8, 1.7 Hz, **C**-NH), 154.3 (dd, *J*=11.3, 5.5 Hz, N-**C**=O), 150.6 (d, *J*=3.1 Hz, O-**C**=O), 138.8 (d, *J*=13.2 Hz, **C**H=C), 96.7 (**C**-CD_2), 87.9 (CH-**C**H-CH_2-O), 86.1 (d, *J*=10.3 Hz, N-**C**H-O), 70.0 (O-**C**H), 63.7 (qi, *J*=23.1 Hz, CD_2), 61.0 (**C**H_2-OH), 40.7 (N-CH-**C**H_2). **^{15}N-NMR** (40 MHz, d^6-DMSO, ppm): δ =-153.0, -213.0. **HRMS** (ESI-): calc. for $C_{11}H_{10}D_2N^{15}N_2O_6$ $[M-H^+]^-$: 286.0796, found: 286.0800. **Melting range:** >200 °C (decomposition). **IR** (cm^{-1}): ν= 3361 (w), 3270 (w), 3161 (w), 1749 (m), 1667 (s), 1627 (s), 1557 (m), 1476 (m), 1282 (s), 1268 (s), 1101 (s), 1082 (s), 1067 (s), 1060 (s), 794 (s), 775 (s).

5-Dideutero-4,5-(1,3-[3*H*,6*H*]oxazin-2-one)-2´-deoxy-(N^1,N^3-^{15}N)-cytidine-5´-triphosphate triethylammonium salt (11)



A solution of tributylammonium pyrophosphate (146.0 mg, 0.27 mmol, 2.7 eq.) in DMF (0.52 mL) and tributylamine (0.67 mL) was added to 2-chloro-1,3,2-benzodioxaphosphorin-4-one (53.6 mg, 0.26 mmol, 2.6 eq.) in DMF (0.37 mL). After stirring for 30 min at room temperature this solution was added to 5-dideutero-4,5-(1,3-[3*H*,6*H*]oxazin-2-one)-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (**10**) (30 mg, 0.10 mmol, 1.0 eq.). After stirring at room temperature for 3 d iodine (69.5 mg, 0.27 mmol, 2.7 eq.) in a pyridine (2.16 mL) water (0.24 mL) mixture was added and stirred at room temperature for 2 h. The crude product was precipitated by addition of a 3 M sodium acetate solution (1.37 mL) and ethanol (26 mL) at -80 °C over night. RP-HPLC purification (0% → 19% buffer B in 45 min) gave **11** as a colorless solid (0.58 mg, 0.70 µmol, 1%). **^{1}H-NMR** (400 MHz, D_2O, ppm): δ = 8.44 (d, *J*=0.9 Hz, 1H, N-C**H**=C), 6.29 (t, *J*=6.3 Hz, 1H, N-C**H**-O), 4.70–4.66 (m, 1H, C**H**-OH), 4.32–4.25 (m, 3H, C**H**-C**H$_2$**-O), 2.62–2.55 (m, 1H, N-CH-C**H$_2$**), 2.43–2.36 (m, 1H, N-CH-C**H$_2$**). **^{31}P-NMR** (162 MHz, D_2O, ppm): δ = -7.5 (br), -11.6 (d, *J*=20.1 Hz), -22.6 (t, *J*=20.4 Hz, P_β). **^{15}N-NMR** (40 MHz, D_2O, ppm): δ =-160.5, -210.3. **HRMS** (ESI-): calc. for $C_{11}H_{13}D_2N^{15}N_2P_3O_{15}$ $[M\text{-}H^+]^-$: 525.9786, found: 525.9789.

5-Formyl-(*tert*-butyl-dimethylsilyl)-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (12)

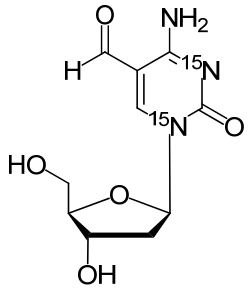


5´-Iodo-3´,5´-(*tert*-butyl-dimethylsilyl)-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (**6**)[1] (1.4 g, 2.4 mmol, 1.0 eq.), $Pd_2(dba)_3 \cdot CHCl_3$ (0.245 mg, 0.24 mmol, 0.1 eq.) and triphenylphosphine (0.373 mg, 1.42 mmol, 0.6 eq.) were dissolved in toluene (40 mL) in a high pressure glas autoclave. The autoclave was flushed with CO to remove residual air and after heating to 60 °C, the CO pressure was set to 3.5 bar.

Tributyltin hydride (0.83 mL, 2.8 mmol, 1.2 eq.) was added within 4 h and the reaction was stirred for additional 14 h at 60 °C and 3.5 bar CO pressure. After evaporation to dryness, the product was purified by silica gel column chromatography (iHex/EtOAc 4:1 → 1:1) to give **12** as an orange solid (864 mg, 1.78 mmol, 74%).

^{1}H-NMR (200 MHz, $CDCl_3$, ppm): δ = 9.47 (s, 1H, C**H**O), 8.53 (s, 1H, N-C**H**=C), 6.19 (t, *J*=5.9 Hz, 1H, N-C**H**-O), 4.35–4.32 (m, 1H, C**H**-O-Si), 4.02–4.03 (m, 1H, C**H**-CH_2-O), 3.94 (dd, *J*=11.6, 2.6 Hz, 1H, C**H$_2$**-O-Si), 3.77 (dd, *J*=11.6, 2.6 Hz, 1H, C**H$_2$**-O-Si), 2.62–2.58 (m, 1H, N-CH-C**H$_2$**), 2.07–2.02 (m, 1H, N-CH-C**H$_2$**), 0.89 (s, 9H, Si-C-C**H$_3$**), 0.87 (s, 9H, Si-C-C**H$_3$**), 0.08 (s, 3H, Si-C**H$_3$**), 0.07 (s, 3H, Si-C**H$_3$**), 0.06 (s, 3H, Si-C**H$_3$**), 0.05 (s, 3H, Si-C**H$_3$**). **^{13}C-NMR** (100 MHz, $CDCl_3$, ppm): δ = 162.5 (d, *J*=6.8 Hz, **C**-NH), 152.3 (d, *J*=14.3 Hz, N-**C**=O), 132.1 (d, *J*=9.9 Hz, O-**C**=O), 138.7 (d, *J*=12.1 Hz, **C**H=C), 104.9 (**C**-CH_2), 88.6 (O-**C**H-CH_2-O), 87.7 (d, *J*=9.6 Hz, N-**C**H-O), 71.5 (**C**H-O-Si), 62.5 (**C**H_2-O-Si), 42.9 (N-CH-**C**H_2), 25.9 (Si-C-**C**H_3), 25.7 (Si-C-**C**H_3), 18.4 (Si-**C**), 17.9 (Si-**C**), -4.6 (Si-**C**H_3), -4.9 (Si-**C**H_3), -5.3 (Si-**C**H_3), -5.4 (Si-**C**H_3). **HRMS** (ESI+): calc. for $C_{22}H_{42}N_2{}^{15}N_2O_6Si_2$ $[M+H^+]^+$: 486.2599, found: 486.2600. **Melting range:** 150–152 °C. **IR** (cm^{-1}): ν= 2952 (w), 2928 (w), 2855 (w), 1757 (m), 1658 (m), 1553 (m), 1471 (m), 1253 (m), 1112 (m), 1064 (m), 1028 (m), 832 (s), 774 (s).

5-Formyl-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (13)

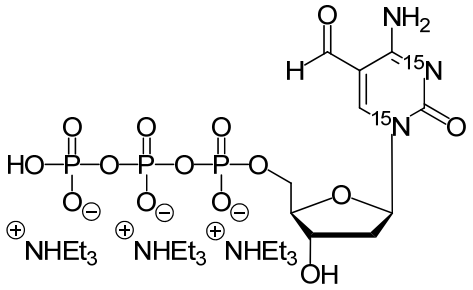


In a polypropylene tube 5-Formyl-(*tert*-butyl-dimethylsilyl)-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (**12**) (493 mg, 0.84 mmol, 1.0 eq.) was dissolved in 20 mL EtOAc and 70% HF-pyridine (325 µL, 3.6 mmol, 4.2 eq.) was added. The reaction was stirred for 17 h at room temperature and was quenched by addition of methoxytrimethylsilane (2.0 mL, 14.5 mmol, 17 eq.). After stirring for 30 min, the precipitate was collected by centrifugation (4500 rpm, 20 min). The crude product was purified by RP-HPLC (0% → 20% buffer B in 45 min) to yield **13** as a colorless solid (190 mg, 0.52 mmol, 63%).

^{1}H-NMR (200 MHz, D_2O, ppm): δ = 9.58 (d, *J*=0.4 Hz, 1H, C**H**O), 8.87 (s, 1H, N-C**H**=C), 6.26 (t, *J*=6.1 Hz, 1H, N-C**H**-O), 4.49 (dd, *J*=11.2, 4.7 Hz, 1H, C**H**-OH), 4.20 (dd, *J*=8.7, 4.3 Hz, 1H, C**H**-CH_2-O), 3.97 (dd, *J*=12.6, 3.3 Hz, 1H, C**H$_2$**-OH), 3.85 (dd, *J*=12.6, 4.9 Hz, 1H, C**H$_2$**-OH), 2.66–2.60 (m, 1H,

N-CH-C**H**$_2$), 2.47–2.40 (m, 1H, N-CH-C**H**$_2$). **^{13}C-NMR** (100 MHz, D_2O, ppm): δ = 190.4 (d, *J*=2.0 Hz, **C**-NH), 162.3 (d, *J*=1.4 Hz, **C**-NH_2), 154.5 (d, *J*=6.4 Hz, N-**C**O-N), 105.6 (**C**H=C), 87.4 (**C**-CO), 87.3 (N-**C**H-CH_2), 86.1 (**C**H-CH_2-O), 69.7 (**C**H-OH), 60.6 (**C**H_2-OH), 40.0 (O-**C**H-CH_2); **HRMS** (ESI+): calc. for $C_{10}H_{14}N_2{}^{15}N_2O_5$ $[M+H^+]^+$: 258.0869, found: 48258.0867;. **Melting range:** 150–152 °C. **IR** (cm^{-1}): ν= 2952 (w), 2928 (w), 2855 (w), 1757 (m), 1658 (m), 1553 (m), 1471 (m), 1253 (m), 1112 (m), 1064 (m), 1028 (m), 832 (s), 774 (s).

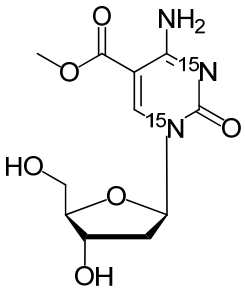
5-Formyl-2´-deoxy-(N^1,N^3-^{15}N)-cytidine-5´-triphosphate triethylammonium salt (14)



A solution of tributylammonium pyrophosphate (42 mg, 77.46 µmol, 2.0 eq.) in DMF (0.15 mL) and tributylamine (0.20 mL) was added to 2-chloro-1,3,2-benzodioxaphosphorin-4-one (15.7 mg, 77.5 µmol, 2.0 eq.) in DMF (0.11 mL). After stirring for 30 min at room temperature this solution was added to 5-Formyl-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (**13**) (10 mg, 38.8 µmol, 1.0eq.). After stirring at room temperature for 3.5 h a solution of 3 % iodine in a pyridine/water mixture (9:1, 0.5 mL) was added and stirred at room temperature for 1.5 h. The crude product was precipitated by addition of a 3 M sodium chloride solution (0.8 mL) and ethanol (15 mL) at -80 °C for 14 h. The triphosphate was isolated by ion exchange chromatography at 4 °C unsing a DEAE-cellulose column with a gradient from 100% water to 0.5 M TEAB (pH 7.5). Further purification was achieved by FPLC (0.1 M, 1 M TEAB, 0-100% TEAB over 30 min) by using a MonoQ 5/50 GL anion exchange column (GE) and gave **14** as a colorless solid (2.06 mg, 2.5 µmol, 12%).

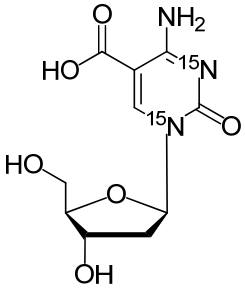
^{1}H-NMR (400 MHz, D_2O): *δ* (ppm) = 9.69 (s, 1H, C**H**O), 8.90 (s, 1H, N-C**H**=C), 6.27 (t, *J*=6.3 Hz, 1H, N-C**H**-O), 4.66–4.69 (m, 1H, C**H**-O), 4.29–4.37 (m, 3H, C**H**-C**H**$_2$-O), 2.42–2.64 (m, 2H, N-CH-C**H**$_2$); **^{31}P-NMR** (162 MHz, D_2O), *δ* (ppm)= -10.3 (m, P_γ), -11.7 (d, *J*=20.2.8 Hz, P_α), -23.3 (t, *J*=19.3 Hz, P_β); **HRMS** (ESI+): calc. for $C_{10}H_{17}N_3P_3O_{15}$ $[M-H^+]^-$: 495.9713, found: 493.9713.

5-Carboxylmethyl-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (15)



A solution of 3´,5´-(*tert*-butyl-dimethylsilyl)-5-carboxylmethyl-2´-deoxy-(N^1,N^3-^{15}N)-cytidine[1] (170 mg, 0.33 mmol, 1.0 eq.), 70% HF-pyridine (0.13 mL, 4.94 mmol, 15.0 eq.) and pyridine (0.13 mL, 1.60 mmol, 4.8 eq.) in EtOAc (10.7 mL) was stirred at room temperature for 21 h. Then additional 70% HF-pyridine (0.09 mL, 3.42 mmol, 10.4 eq.) and pyridine (0.09 mL, 1.11 mmol, 3.4 eq.) were added and the reaction mixture was stirred for additional 26 h. The reaction was quenched with methoxytrimethylsilane (2.7 mL, 19.59 mmol, 59.4 eq.) and stirred for 30 min. After evaporation to dryness and purification by silica gel column chromatography (DCM:MeOH 50:1 → 5:1) product **15** was obtained as a colorless solid (30 mg, 0.10 mmol, 30%). **^{1}H-NMR** (200 MHz, D_2O, ppm): δ = 8.96 (s, 1H, N-C**H**=C), 6.16 (t, *J*=5.8 Hz, 1H, N-C**H**-O), 4.45–4.36 (m, 1H, C**H**-OH), 4.08–4.02 (m, 1H, C**H**-CH_2-O), 3.94–3.71 (m, 2H, C**H$_2$**-OH), 3.83 (s, 3H, OC**H$_3$**), 2.58–2.43 (m, 1H, O-CH-C**H$_2$**), 2.39–2.25 (m, 1H, N-CH-C**H$_2$**). **^{13}C-NMR** (100 MHz, D_2O, ppm): δ = 166.0 (d, *J*=1.6 Hz, **C**OO), 163.5 (dd, *J*=6.2, 1.4 Hz, **C**-NH_2), 155.8 (dd, *J*=11.8, 8.3 Hz, N-**C**O-N), 148.8 (d, *J*=14.0 Hz, **C**H=C), 96.8 (d, *J*=1.0 Hz, **C**-CO), 86.9 (d, *J*=9.7 Hz, N-**C**H-CH_2), 86.7 (**C**H-CH_2-O), 69.2 (**C**H-OH), 60.1 (**C**H_2-OH), 52.3 (O**C**H_3), 39.9 (O-**C**H-CH_2). **^{15}N-NMR** (40 MHz, D_2O, ppm): δ =-179.7, -212.2. **HRMS** (ESI+): calc. for $C_{11}H_{16}N^{15}N_2O_6$ $[M+H]^+$: 288.0974, found: 288.0977. **Melting range:** 169–171 °C. **IR** (cm^{-1}): ν= 1713 (m), 1635 (s), 1480 (s), 1321 (s), 1097 (s), 1062 (s), 786 (s).

5-Carboxyl-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (16)

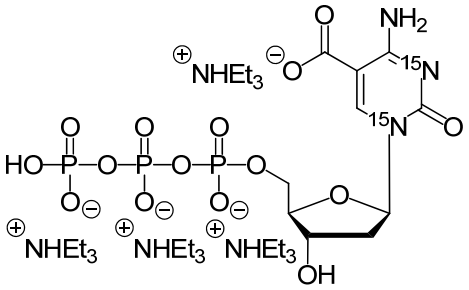


A solution of 5-Carboxylmethyl-2´-deoxy-(N^1,N^3-^{15}N)-cytidine (**15**) (25 mg, 87.0 µmol, 1.0 eq.) and LiOH (25 mg, 1.04 mmol, 12.0 eq.) in a water (1.3 mL) MeCN (7.0 mL) mixture was stirred at room

temperature for 19 h. After adjusting the pH to 4.0 with 2 M hydrochloric acid (0.5 mL) the colorless precipitate was filtered and washed with water to obtain **16** (11 mg, 40.26 µmol, 46%).

^{1}H-NMR (400 MHz, D_2O): δ (ppm) = 8.79 (s, 1H, N-C**H**=C), 6.29 (dt, *J*=6.5, 1.3 Hz, 1H, N-C**H**-O), 4.53–4.49 (m, 1H, C**H**-OH), 4.16 (dd, *J*=8.1, 4.2 Hz, 1H, C**H**-CH_2-O), 3.92 (dd, *J*=12.4, 3.5 Hz, 1H, C**H**$_2$-O), 3.83 (dd, *J*=12.5, 5.3 Hz, 1H, C**H**$_2$-O), 2.59–2.53 (m, 1H, N-CH-C**H**$_2$), 2.46–2.38 (m, 1H, N-CH-C**H**$_2$). **^{13}C-NMR** (101 MHz, D_2O, ppm): δ =168.0 (**C**O-O), 159.8 (d, *J*=13.3 Hz, **C**-NH_2), 150.0 (N-**C**=O), 148.0 (d, *J*=13.1 Hz, **C**H=C), 101.3 (**C**-CO), 87.2 (**C**H-CH_2-O), 86.9 (d, *J*=10.2 Hz, N-**C**H-CH_2), 70.1 (**C**H-OH), 60.8 (**C**H_2-OH), 39.5 (O-**C**H-CH_2). **^{15}N-NMR** (41 MHz, d^6-DMSO, ppm): δ = -171.6, -212.1. **HRMS** (ESI+): calc. for $C_{10}H_{14}N^{15}N_2O_6^+$ $[M+H^+]^+$: 274.0818, found: 274.0818. **Melting range**: >250 °C (decomposition). **IR** (cm^{-1}): v= 3402 (m), 3274 (br, m), 1661 (s), 1627 (m), 1448 (m), 1291 (m), 1202 (m), 1093 (s), 1003 (s), 816 (m).

5-Carboxyl-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine-5´-triphosphate triethylammonium salt (17)



A solution of tributylammonium pyrophosphate (42.5 mg, 77.46 µmol, 2.0 eq.) in DMF (0.15 mL) and tributylamine (0.20 mL) was added to 2-chloro-1,3,2-benzodioxaphosphorin-4-one (15.7 mg, 77.5 µmol, 2.0 eq.) in DMF (0.11 mL). After stirring for 30 min at room temperature this solution was added to 5-carboxyl-2´-deoxy-(*N*^{1},*N*^{3}-^{15}N)-cytidine (**16**) (10.5 mg, 38.4 µmol, 1.0 eq.). After stirring at room temperature for 20 h iodine (34.8 mg, 0.14 mmol, 3.6 eq.) in a pyridine (1.08 mL) water (0.12 mL) mixture was added and stirred at room temperature for 2 h. The crude product was precipitated by addition of a 3 M sodium acetate solution (0.65 mL) and ethanol (12 mL) at -80 °C for 1 h. RP-HPLC purification (0% → 6% buffer B in 45 min) gave **17** as a colorless solid (2.28 mg, 2.5 µmol, 7%). **^{1}H-NMR** (400 MHz, D_2O): δ (ppm) = 8.44 (t, *J*=1.3 Hz, 1H, N-C**H**=C), 6.29 (t, *J*=6.7 Hz, 1H, N-C**H**-O), 4.65–4.61 (m, 1H, C**H**-OH), 4.29–4.20 (m, 3H, C**H**-C**H**$_2$-O), 2.51–2.36 (m, 2H, N-CH-C**H**$_2$). **^{31}P-NMR** (162 MHz, D_2O, ppm): δ = -11.0 (d, *J*=19.8 Hz), -11.0 (d, *J*=19.8 Hz), -23.4 (t, *J*=19.8 Hz, P_β). **HRMS** (ESI+): calc. for $C_{10}H_{15}N^{15}N_2O_{15}P_3^+$ $[M+H^+]^+$: 511.9662, found: 511.9661.

[1] Pfaffeneder, T., Hackner, B., Truß, M., Münzel, M., Müller, M., Deiml, C. A., Hagemeier, C., Carell, T., *Angew. Chem. Int. Ed*. 2011, *50*, 7008-7012.

[2] Pfaffeneder T., Hackner B., Schiesser S., Kosmatchev O., Wagner M., Steinbacher J., Müller M., Höfner G., Wanner K., Hagemeier C., Michalakis S., Biel M., Truss M., Carell T.; manuscript in preparation.

[3] Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeier, C.; Truss, M.; Carell, T., *Angew. Chem. Int. Ed.* **2012,** *51*, 6516-6520.

13.3 Supplementary material for Chapter 8

T. Pfaffeneder,[1] F. Spada,[1] M. Wagner,[1] C. Brandmayr, S. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Muller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schar, M. Muller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574-581. "*Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.*"

Supplementary Information

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

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

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

[2] Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Ziegelstr. 5-9, 10098 Berlin, Germany.

[3] Center for Integrated Protein Science at the Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, 81377 München, Germany.

[4] Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany.

[5] Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland.

[6] Department of Molecular Cancer Research, Cancer Genomics Netherlands, UMC Utrecht, 3584 CG Utrecht, Netherlands.

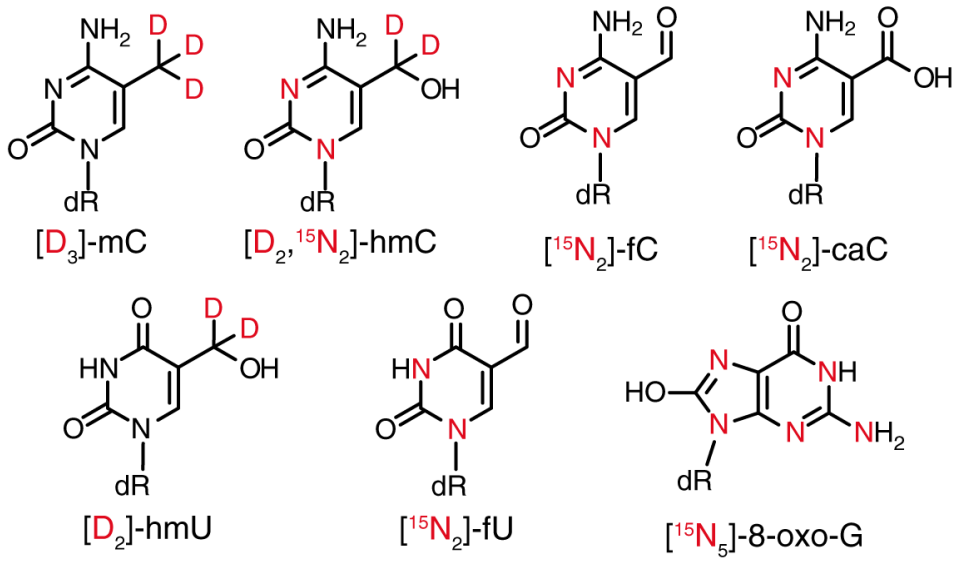
[≠] Present Address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA, Nijmegen, The Netherlands

* Correspondence to: markus.mueller@cup.uni-muenchen.de and thomas.carell@cup.uni-muenchen.de

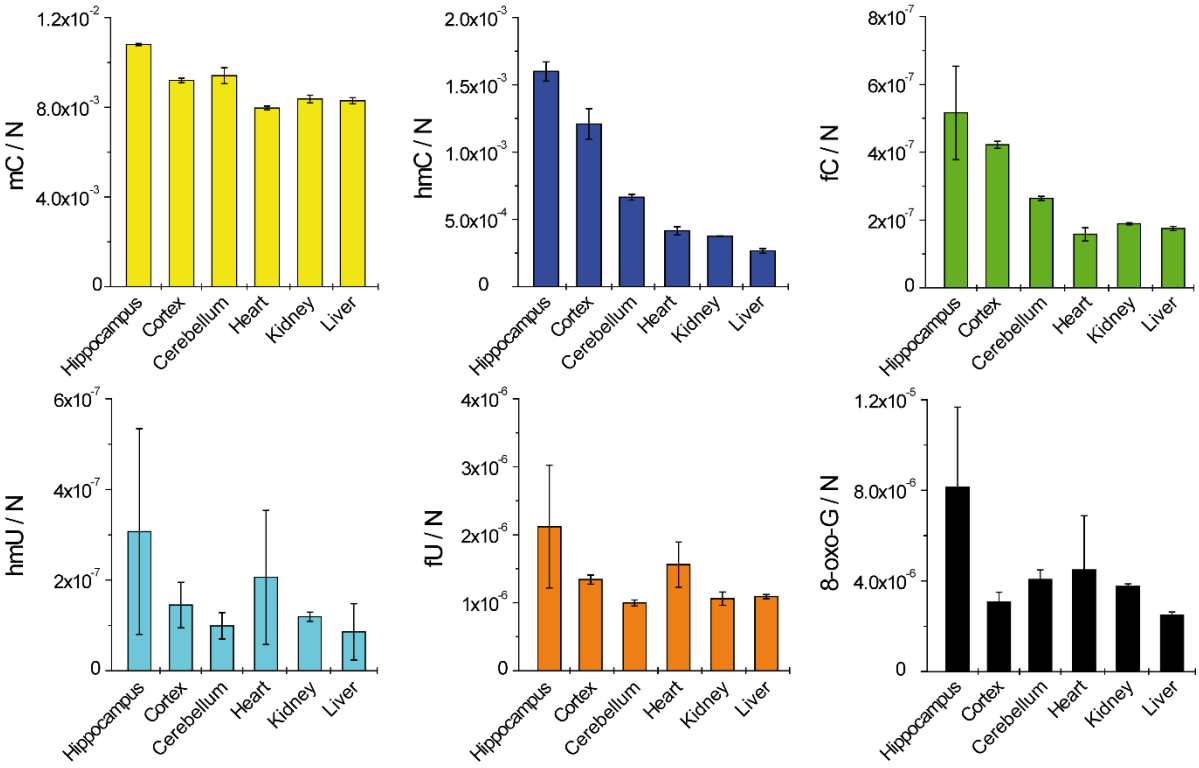
[#] These authors contributed equally

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Results

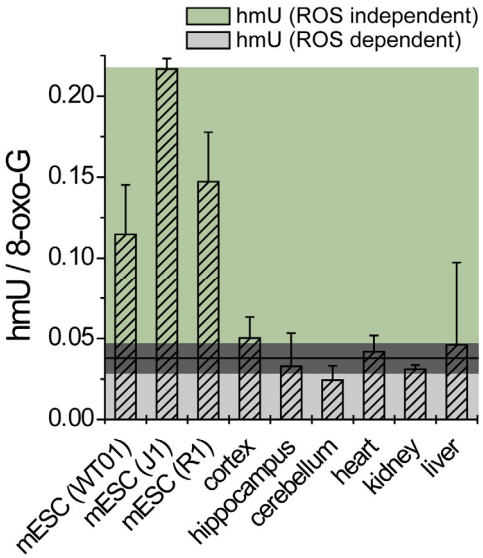


Supplementary Figure 1. Isotopically labeled nucleosides used as internal standards for quantitative LC-MS/MS analysis (dR = -2´-deoxyribose).



Supplementary Figure 2. DNA modification levels per nucleoside (N) of different murine tissues from 3 month old individuals (n = 3). Depicted are biological mean values ±SD.

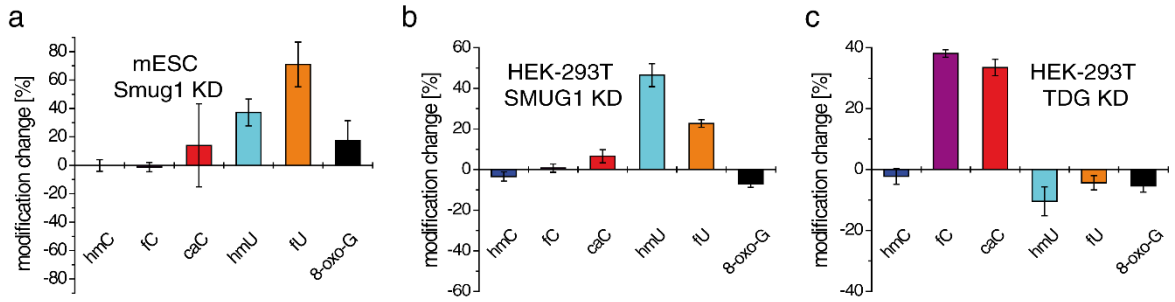
Nature Chemical Biology: doi:10.1038/nchembio.1532



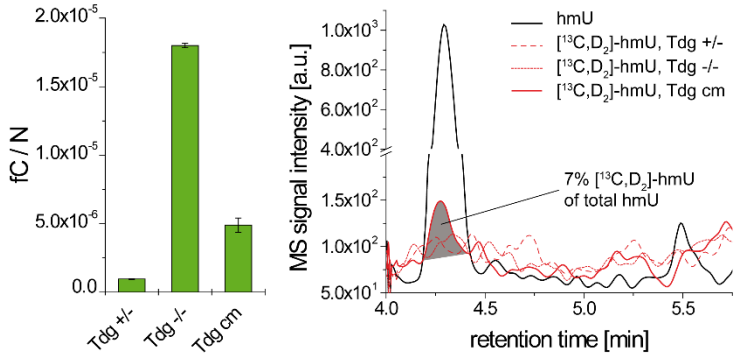
Supplementary Figure 3. hmU-levels normalized to the oxidative background marker 8-oxo-G in mESCs (n_{WT01} = 7; n_{J1} = 2, n_{R1} = 2) and murine tissues (n = 3, 3 month old individuals) in order to dissect ROS dependent and ROS independent processes. Normalization was necessary to take deviating background oxidation of DNA sample preparation into account. The light grey area reflect the hmU-level fractions, which are generated by ROS dependent processes. The green area reflect the hmU-level fractions, which are generated by ROS independent processes. The assignment is based on the assumption, that hmU-levels in somatic tissue are exclusively ROS created lesions (derived from the cluster analysis in **Fig. 2b**). The dark grey area reflect the mean value ±SD of hmU/8-oxo-G ratios of the tissue data. The difference between the height of the hmU/8-oxo-G ratios of mESCs and the mean of the tissue data give the hmU-fraction which is formed by ROS-independent processes. In WT01, J1 and R1 cells about 67%, 83% and 74%, respectively, of the global hmU-levels are estimated to be created by ROS independent processes. Bars reflect biological mean values ±SD. The differences between mESCs (WT01, J1) and murine tissues are significant ($P = 9.3\text{x}10^{-5}$–$5.0\text{x}10^{-3}$; unpaired two-tailed t-test) except for mESCs (WT01, J1) and liver (P = 0.081, 0.15).

Supplementary Figure 4. Schematic representation of isotope tracing experiments with [^{13}C,^{15}N$_2$]-T (left; blue) and [^{13}C,D$_3$]-methionine (right; red) and exchange rates of derived genomic isotopologues. Small negligible deviations in the exchange rates are due to differential noise sources. LOD = limit of detection. wt = wild type mESCs; Tdg cm = Tdg / mESCs complemented with catalytic inactive Tdg (see Supplementary Fig. 6).

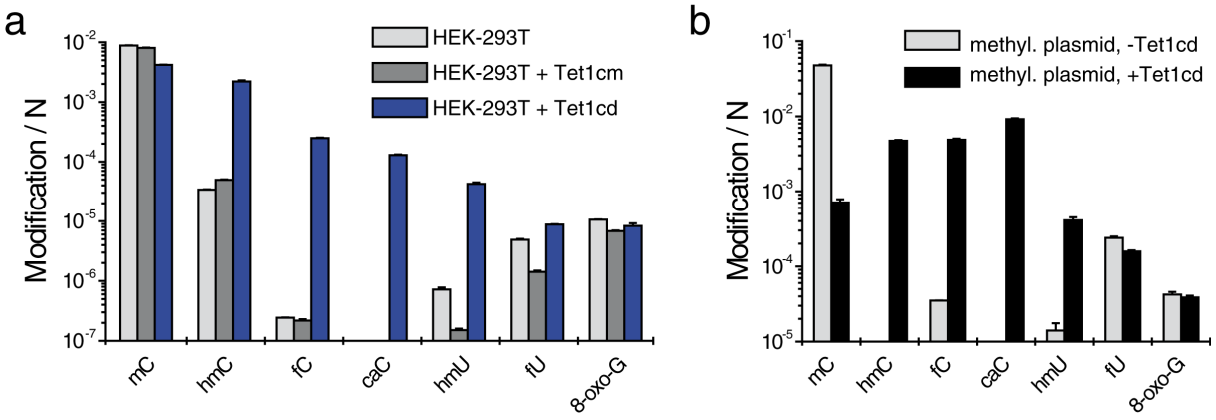
Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 5. (**a**) Effect of Smug1 depletion on modification levels in mESCs (R1). Effect of SMUG1 (**b**) and TDG (**c**) depletion on modification levels in HEK-293T cells overexpressing Tet1cd. Shown is the percent change in modification content per nucleoside in cells co-transfected with esiRNAs targeting TDG or SMUG1 relative to co-transfection with control esiRNA. Depicted are technical mean values ± SD.

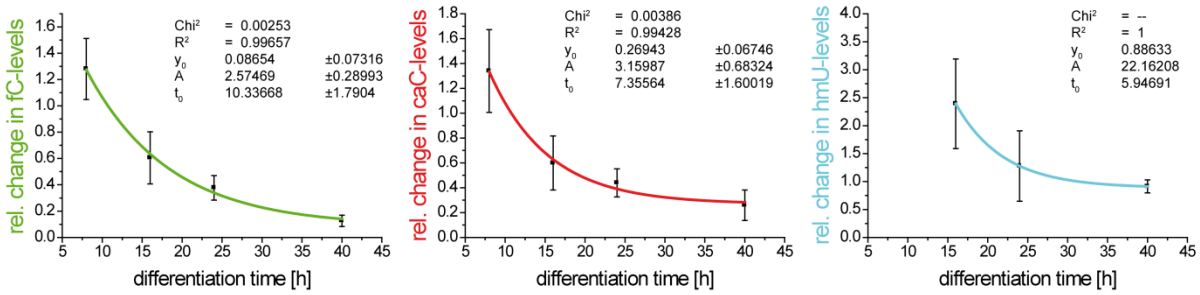


Supplementary Figure 6. Isotope tracing experiments with *Tdg*$^{+/-}$, *Tdg*$^{-/-}$ mESCs as well as *Tdg*$^{-/-}$ mESCs complemented with a catalytic mutant of Tdg (Tdg cm) grown in the presence of [^{13}C,D$_3$]-methionine (200 µM). The catalytic mutant of Tdg is not completely inactive (fC levels are between Tdg +/- and Tdg -/- cells, left). Only in case of *Tdg*-/- cells complemented with a catalytic inactive Tdg (Tdg cm) labeled hmU was detected, which originated from the deamination of labeled hmC. ~7% [^{13}C,D$_2$]-hmU over total hmU was observed. This corresponds to ~0.06% deamination of hmC to hmU under these conditions ($2.5x10^{-4}$ total hmC / N; $2.2x10^{-6}$ total hmU / N). Labeled fU was not observed. Depicted fC-levels represent technical mean values ±SD.

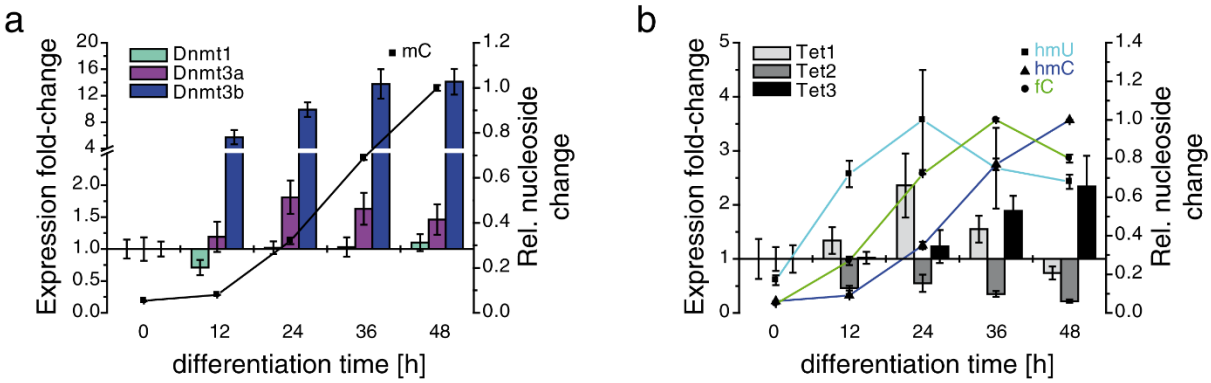


Supplementary Figure 7. Tet1 and Tet2 generate hmU in HEK-293T and in vitro (a) Effect of Tet1 overexpression on modified pyrimidines in HEK-293T cells. Modification levels in cells overexpressing wt and catalytic mutant versions of Tet1 catalytic domain (Tet1cd, blue bars and Tet1cm, gray bars, respectively), or a control construct (white bars). Depicted are mean values ±SD of technical triplicates on a logarithmic scale. **(b)** Pyrimidine modification levels in methylated plasmid DNA after treatment in vitro with Tet1cd. Depicted are mean values ±SD of technical duplicates. Note the logarithmic scale.

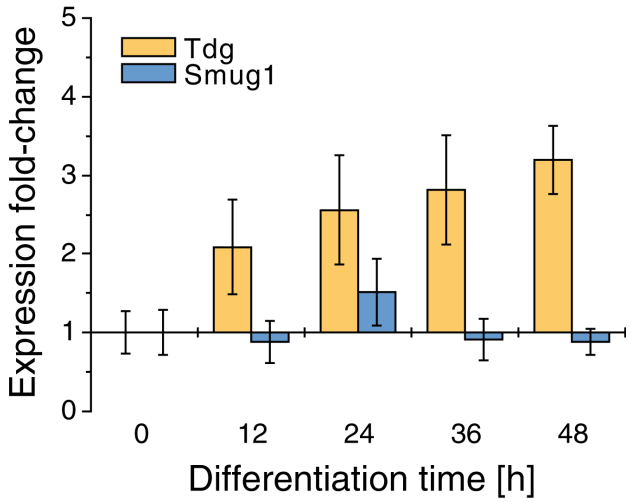
Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 8. Exponential models for fitting the decay curve of fC, caC and hmU in combined data sets from differentiation of R1 and C57Bl6/129-derived mESCs (6 biological independent experiments). In a simplified approach a single exponential decay model ($y = y_0 + A*\exp(-x/t_0)$) was plotted using ORIGIN®. The parameters y_0 (offset), t_0 (time constant) and A (amplitude) of each decay function were iteratively optimized until the minimum of the Chi^2 value of the fitting was reached. Half-life times ($t_{1/2} = t_0*\ln 2$) for fC, caC and hmU were 7.2±1.2, 5.1±1.1 and 4.1 h, respectively.

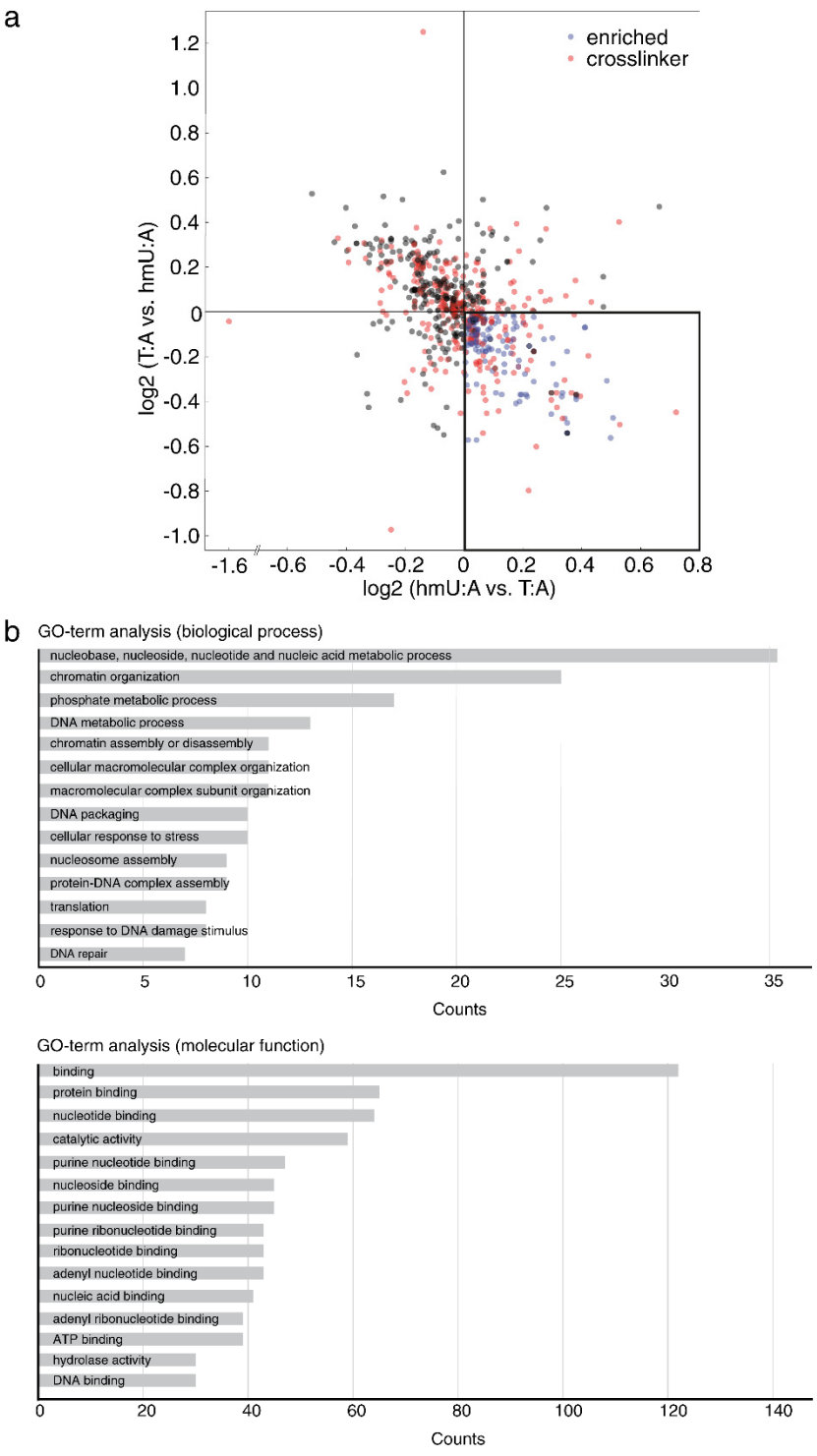


Supplementary Figure 9. Normalized transcript levels of Dnmts (c), Tet1–3 (d) and normalized modification levels of mC (c), hmC, fC and hmU (d) during differentiation of naïve mESCs in the presence of FGF-2 and ActA.



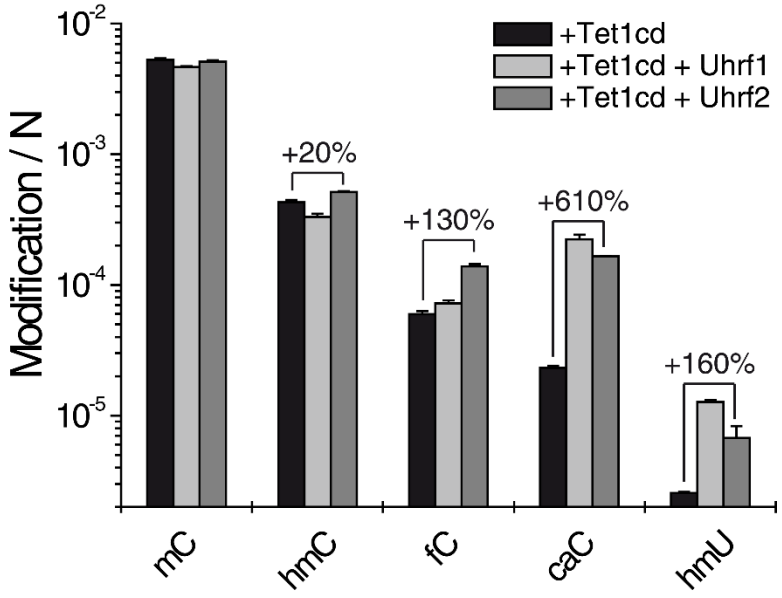
Supplementary Figure 10. Expression level analysis of Tdg and Smug1 during differentiation of mESCs in the presence of FGF-2 and ActA. Expression levels were quantified with respect to the housekeeping gene Gapdh and normalized to time point 0 h. Depicted are technical mean values ± SD.

Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 11. (a) Scatterplot of proteins enriched with the hmU:A containing oligomer. Ratios of a forward and a reverse experiment are plotted. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross linker and marked in red. Gray dots are considered unspecific binders. See **Fig. 6** for detailed view. **(b)** Gene Ontology Analysis performed with DAVID Bioinformatics Resources 6.7[4]

Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 12. Effect on modified pyrimidines in HEK-293T cells upon Tet1cd and Uhrf1 (light gray bars) or Uhrf2 (gray bars) co-overexpression. Depicted are mean values ±SD of technical triplicates on a logarithmic scale.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 1. Isotope tracing experiments by supplementing the growth medium of mES cells (LIF), differentiating mESCs (R1, without growth factors) and HEK-293T cells with either [^{13}C,$^{15}N_2$]-T (50 or 100 μM) or [^{13}C,D_3]-methionine (0.2 mM). Small deviations in the exchange yields are due to differential noise sources and are negligible. LOD = Limit of detection. In case of [^{13}C,D_1]-fU no difference was observed compared to the natural control.

cell type	growth medium	[^{13}C,$^{15}N_2$]-T / T [%]	[^{13}C,$^{15}N_2$]-hmU / hmU [%]	[^{13}C,$^{15}N_2$]-fU / fU [%]
mESC (2i)	100 μM natural T	0.1	< LOD	< LOD
mESC (2i)	100 μM [^{13}C,$^{15}N_2$]-T	76.0	78.2	74.6
HEK + Tet1cd (72h)	50 μM [^{13}C,$^{15}N_2$]-T	73.8	74.2	71.0

cell type	growth medium	[^{13}C,D_3]-mC / mC [%]	[^{13}C,D_2]-hmC / hmC [%]	[^{13}C,D_2]-hmU / hmU [%]	[^{13}C,D_1]-fU / fU [%]
mESC (LIF)	natural methionine	0.1	< LOD	< LOD	3.3
mESC (LIF; 0 h)	[^{13}C,D_3]-methionine	88.9	87.6	< LOD	3.0
diff. mESC (12 h)	[^{13}C,D_3]-methionine	89.3	88.4	< LOD	3.1
diff. mESC (24 h)	[^{13}C,D_3]-methionine	90.1	89.3	< LOD	3.2
diff. mESC (48 h)	[^{13}C,D_3]-methionine	90.4	90.5	< LOD	3.5
mESC *Tdg*$^{+/-}$	[^{13}C,D_3]-methionine	88.0	87.6	< LOD	3.7
mESC *Tdg*$^{-/-}$	[^{13}C,D_3]-methionine	87.4	87.2	< LOD	< LOD
mESC *Tdg*$^{-/-}$ + Tdg cm	[^{13}C,D_3]-methionine	86.9	86.7	7.4	2.2
HEK + Tet1cd (72h)	[^{13}C,D_3]-methionine	87.4	83.4	< LOD	< LOD

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 2. Assessment of ROS dependent hmU and fU product ratio of T-oxidation in HEK-293T wild type cells where TET activity is lowest (related to Fig. 3c). Modified nucleosides / N are given as mean values plus SD of three independent technical replicates. When T is oxidized by ROS about 9.8% hmU and 90.2% fU is generated.

n	mC / N		hmC / N		fC / N		caC / N
	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean
1	6.14E-03	1.97E-04	2.92E-05	9.90E-08	2.56E-07	2.59E-09	n.d.
2	6.21E-03	1.32E-05	2.95E-05	2.21E-09	3.10E-07	2.06E-08	n.d.
3	5.76E-03	4.36E-05	3.32E-05	6.55E-07	2.86E-07	2.95E-09	n.d.
4	9.01E-03	2.87E-04	5.23E-05	5.01E-07	3.50E-07	5.00E-09	n.d.
5	8.80E-03	9.62E-05	3.39E-05	5.44E-07	2.43E-07	2.41E-09	n.d.
6	8.55E-03	7.69E-05	3.70E-05	2.17E-07	2.14E-07	1.30E-08	n.d.
biol. mean	7.41E-03		3.59E-05		2.76E-07		
biol. SD	1.52E-03		8.56E-06		4.90E-08		

n	hmU / N		fU / N		hmU/ (hmU+fU)	fU/ (hmU+fU)	8-oxo-G / N	
	techn. mean	SD	techn. mean	SD	[%]	[%]	techn. mean	SD
1	2.66E-07	5.54E-08	5.15E-06	1.59E-07	4.9	95.1	8.49E-06	1.30E-07
2	1.21E-06	1.62E-07	8.02E-06	6.66E-07	13.1	86.9	1.00E-05	1.98E-07
3	3.65E-08	7.35E-09	9.76E-07	8.75E-09	3.6	96.4	3.02E-06	7.02E-08
4	6.31E-07	9.10E-10	3.51E-06	3.28E-08	15.3	84.7	7.43E-06	1.34E-07
5	7.21E-07	5.93E-08	4.89E-06	1.59E-07	12.8	87.2	1.06E-05	1.14E-07
6	3.58E-07	4.95E-08	3.56E-06	9.64E-09	9.1	90.9	7.22E-06	1.15E-08
biol. mean	5.37E-07		4.35E-06		9.8	90.2	7.80E-06	
biol. SD	4.13E-07		2.33E-06		4.8	4.8	2.70E-06	

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 1: oligonucleotide sequences for protein pull-down assays

Supplementary Table 3. DNA oligonucleotides used in protein pull-down studies.

ODN	Sequence (5´→ 3´)	Modifications
1	Biotin-GCA-TCC-GGT-CAY-CGT-TCC-TTC-GGA	Y = 5-octadienyl-U
2	Biotin-GCA-TCC-GGT-CAY-CAT-TCC-TTC-GGA	Y = 5-octadienyl-U
3	TCC-GAA-GGA-AXG-ATG-ACC-GGA-TGC	X= T
4		X= hmU
5		X= C
6		X= hmC
7	Biotin-GCT-CAC-GCT-AGY-CGA-CTC-CGT-GCA	Y = 5-octadienyl-U
8	TGC-ACG-GAG-TXG-ACT-AGC-GTG-AGC	X = T
9		Y = hmU

Hybridization scheme:

Pull-down 1: hmU:A vs. T:A : ODN4/2 vs. ODN3/2
Pull-down 2: hmU:G vs. C:G = ODN4/1 vs. ODN5/1
Pull-down 3: hmC:G vs. C:G = ODN6/1 vs. ODN5/1
Pull-down 4 (scrambled sequence): hmU:A vs. T:A = ODN9/7 vs. ODN8/7

Supplementary Note 2: LC-UV-ESI-MS/MS analysis of DNA

Supplementary Table 4. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA. CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ($[M+H]^+$ species) as well as in the negative ($[M-H]^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1.5–4.0 min								
$[^{15}N_2]$-caC	274.08	Wide	158.03	Wide	170	5	5	Positive
caC	272.09	Wide	156.04	Wide	170	5	5	Positive
$[^{15}N_2,D_2]$-hmC	262.12	enhanced	146.07	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
$[D_3]$-mC	245.13	enhanced	129.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
C	228.1	enhanced	112.05	enhanced	1	1	0	Positive
time segment 4.0–6.0 min								
$[D_2]$-hmU	259.09	Wide	216.08	Wide	48	7	5	Negative
$[D_2]$-hmU	259.09	Wide	126.05	Wide	48	7	5	Negative
hmU	257.08	Wide	214.07	Wide	48	7	5	Negative
hmU	257.08	Wide	124.04	Wide	48	7	5	Negative
$[^{15}N_2]$-fU	257.06	Wide	213.05	Wide	48	6	5	Negative
fU	255.06	Wide	212.06	Wide	48	6	5	Negative
time segment 6.0–9.0 min								
$[^{15}N_5]$-8-oxo-G	289.08	Wide	173.04	Wide	120	9	7	Positive
8-oxo-G	284.1	Wide	168.05	Wide	120	9	7	Positive
$[^{15}N_2]$-fC	258.09	Wide	142.04	Wide	120	5	5	Positive
fC	256.09	Wide	140.05	Wide	120	5	5	Positive

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 5. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA obtained from cells which were grown in medium supplemented with labeled thymidine ([^{13}C,^{15}N$_2$]-T). CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ([M+H]$^+$ species) as well as in the negative ([M-H]$^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1.5–4.0 min								
[^{13}C,^{15}N$_2$]-caC	275.09	wide	159.04	wide	65	5	5	Positive
caC	272.09	wide	156.04	wide	65	5	5	Positive
[^{13}C,^{15}N$_2$]-hmC	261.11	enhanced	145.06	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
[^{13}C,^{15}N$_2$]-mC	245.13	enhanced	129.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
[^{13}C,^{15}N$_2$]-C	231.1	enhanced	115.05	enhanced	40	1	3	Positive
C	228.1	enhanced	112.1	enhanced	40	1	3	Positive
time segment 4.0–6.0 min								
[^{13}C,^{15}N$_2$]-hmU	260.08	wide	215.07	wide	50	7	5	Negative
hmU	257.08	wide	214.07	wide	50	7	5	Negative
[^{13}C,^{15}N$_2$]-fU	258.06	wide	213.05	wide	50	6	5	Negative
fU	255.06	wide	212.06	wide	50	6	5	Negative
time segment 6.0–9.0 min								
[^{15}N$_5$]-8-oxo-G	289.08	wide	173.04	wide	80	9	7	Positive
8-oxo-G	284.1	wide	168.05	wide	80	9	7	Positive
[^{13}C,^{15}N$_2$]-fC	259.09	wide	143.04	wide	80	5	5	Positive
fC	256.09	wide	140.05	wide	80	5	5	Positive
[^{13}C,^{15}N$_2$]-T	246.1	enhanced	130.05	enhanced	30	40	3	Positive
T	243.1	enhanced	127.05	enhanced	30	40	3	Positive

Supplementary Table 6. Compound-dependent LC-MS/MS-parameters used for the analysis of genomic DNA obtained from cells which were grown in medium supplemented with labeled (*methyl*-^{13}C,D$_3$)-methionine.CE: collision energy; CAV: collision cell accelerator voltage; EMV: electron multiplier voltage. The nucleosides were analyzed in the positive ([M+H]$^+$ species) as well as in the negative ([M-H]$^-$ species) ion selected reaction monitoring mode (SRM).

compound	Precursor Ion (*m/z*)	MS1 Resolution	Product Ion (*m/z*)	MS2 Resolution	Dwell time [ms]	CE (V)	CAV (V)	Polarity
time segment 1.5–4.0 min								
[^{13}C]-caC	273.09	wide	157.04	wide	65	5	5	Positive
caC	272.09	wide	156.04	wide	65	5	5	Positive
[^{13}C,D$_2$]-hmC	261.12	enhanced	145.08	enhanced	40	27	1	Positive
hmC	258.11	enhanced	142.06	enhanced	40	27	1	Positive
[^{13}C,D$_3$]-mC	246.14	enhanced	130.09	enhanced	30	60	1	Positive
mC	242.11	enhanced	126.07	enhanced	30	60	1	Positive
C-dN	228.1	enhanced	112.1	enhanced	40	1	3	Positive
time segment 4.0–6.0 min								
[^{13}C,D$_2$]-hmU	260.09	wide	217.09	wide	60	7	5	Negative
hmU	257.08	wide	214.07	wide	60	7	5	Negative
[^{13}C,D]-fU	257.07	wide	214.07	wide	60	6	5	Negative
fU	255.06	wide	212.06	wide	60	6	5	Negative
time segment 6.0–9.0 min								
[^{15}N$_5$]-8-oxo-G	289.08	wide	173.04	wide	80	9	7	Positive
8-oxo-G	284.1	wide	168.05	wide	80	9	7	Positive
[^{13}C,D]-fC	258.1	wide	142.06	wide	80	5	5	Positive
fC	256.09	wide	140.05	wide	80	5	5	Positive

Nature Chemical Biology: doi:10.1038/nchembio.1532

Spiking amounts of labeled internal standards for quantitative LC-MS/MS analysis

The quantification of nucleosides of genomic DNA isolated from mESC or mouse tissue was carried out with the following amounts of internal standards: 51.03 pmol [D_3]-mC, 7.655 pmol [$^{15}N_2$,D_2]-hmC, 45.6 fmol [$^{15}N_2$]-fC, 43.0 fmol [$^{15}N_2$]-caC, 108.9 fmol [$^{15}N_5$]-8-oxo-G; 160.1 fmol [D_2]-hmU and 180.0 fmol [$^{15}N_2$]-fU. The quantification of nucleosides of genomic DNA isolated from HEK293 cells overexpressing Tet was carried out with the following amounts of internal standards: 34.02 pmol [D_3]-mC, 5.103 pmol [$^{15}N_2$,D_2]-hmC, 303.8 fmol [$^{15}N_2$]-fC, 215.1 fmol [$^{15}N_2$]-caC, 108.9 fmol [$^{15}N_5$]-8-oxo-G; 160.1 fmol [D_2]-hmU and 180.0 fmol [$^{15}N_2$]-fU. Genomic DNA samples isolated from cells grown in media supplemented with either [^{13}C,$^{15}N_2$]-T or [^{13}C,D_3]-methionine were not spiked with internal standards except [$^{15}N_5$]-8-oxo-G.

Validation of the LC-UV-MS/MS quantification method:

Method validation, in particular linearity, precision, and accuracy (i.e. determined from matrix samples spiked with isotopically labeled internal standards) of the established method were investigated. Validation for the established LC-UV-ESI-MS/MS quantification method was based on three different series (i.e., calibration functions and quality control samples) accomplished on different days. Each calibration standard (5-8 standard concentrations) was analyzed five times. Each validation experiment was complemented by matrix blanks (analyzed in triplicates) to ensure selectivity and specificity of the method. Linear regression was applied by Origin® 6.0 (*Microcal*™) to obtain calibration curves. Therefore, the ratio of the area under the curve (A/A*) of the unlabeled nucleoside to the internal standard (*) was plotted against the ratio of the amount of substance (n/n*) of the unlabeled nucleoside to the internal standard (*) (see **Supplementary Fig. 13**). Calibration functions were calculated without weighting. Additionally, acceptable accuracy (80–120%) as well as precision (<20% RSD) was required. Accuracy was proven by computing the amount of substance n from the obtained A/A* ratios of the calibration standards using the respective calibration function. Here, accuracy was defined as the ratio of the used amount of substance to the calculated amount of substance in percent and had to be between 80–120% for each standard concentration. Precision was defined as follows: technical replicates of A/A* ratios for each calibration standard had to have relative standard deviations (RSD) smaller than 20%. The lower limit of quantification (LLOQ) was defined as the lowest concentration fulfilling the requirements of accuracy and precision and achieving a response of at least three times the response compared with the blank response. A compilation of absolute and relative LLOQs is shown in **Supplementary Table 7**.

Quality control samples to evaluate intra-batch precision (see below) were investigated using a biological sample spiked with internal standards. Long-term stability of aqueous solutions of the labeled and unlabeled nucleosides at a storage temperature of −20 °C was investigated over two months including several freeze and thaw cycles by analyzing the MS/MS-responses with each batch.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 7. Compilation of absolute lower limits of quantification [fmol] (LLOQ; see Supplementary Fig. 13) and relative LLOQs [per N] depending on the amount of DNA, which is digested. The relative LLOQs were computed by generating ratios of the absolute LLOQ [pmol] to the total amount of nucleosides (N; [pmol]) in the respective amount of DNA [µg]. The total amount of nucleosides was obtained by using the average molar mass of 308.91 g mol^{-1} for the monomeric DNA entity by taking the GC-content (21% C or G) in mouse into account.

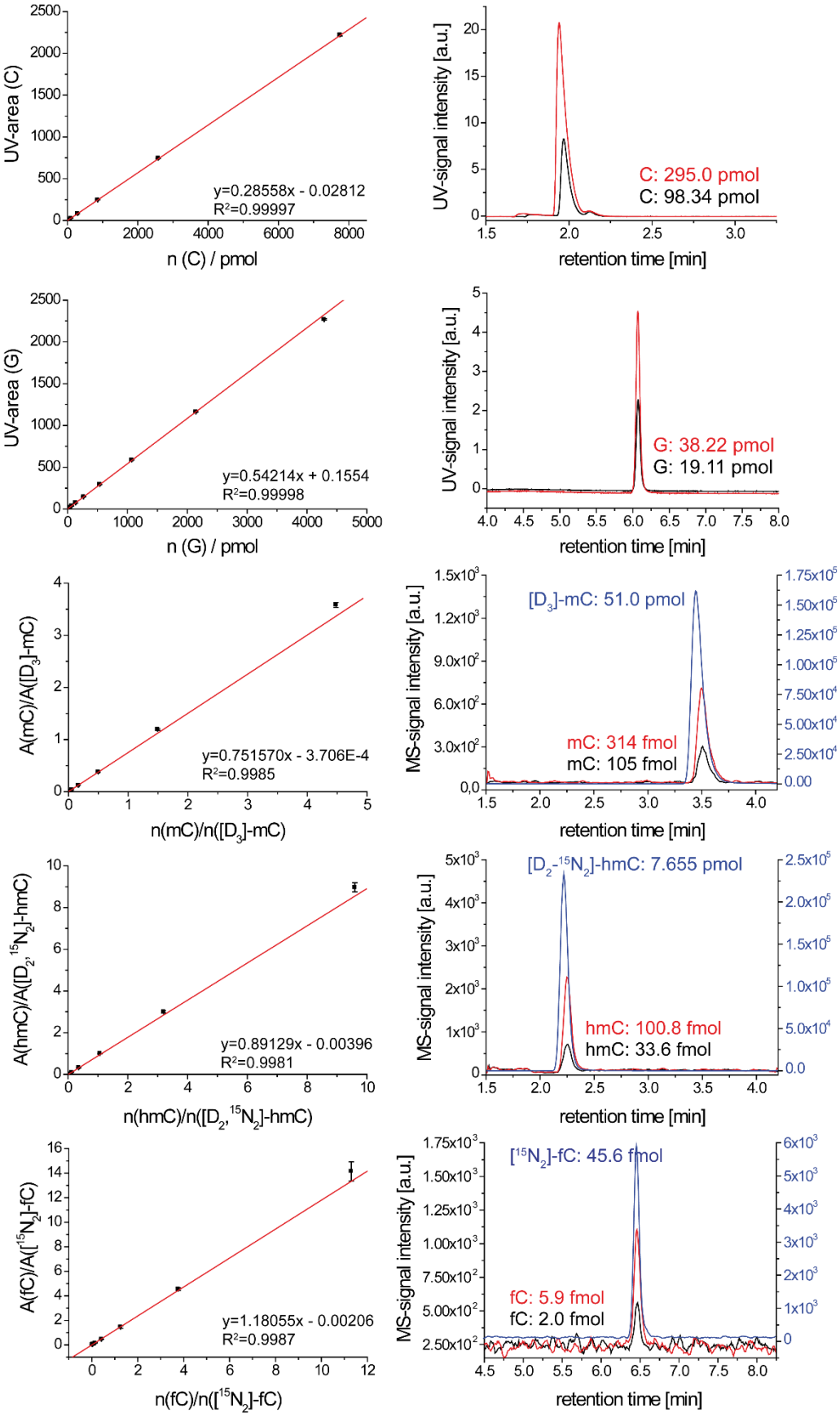
	absolute LLOQ [fmol]	**relative LLOQ [per N]**	**relative LLOQ [per N]**	**relative LLOQ [per N]**
DNA amount		**5 µg**	**10 µg**	**25 µg**
mC	104.5	6.5E-06	3.2E-06	1.3E-06
hmC	100.8	6.2E-06	3.1E-06	1.2E-06
fC	2.1	1.3E-07	6.5E-08	2.6E-08
caC	2.0	1.2E-07	6.2E-08	2.5E-08
8-oxo-G	2.0	1.2E-07	6.2E-08	2.5E-08
dU	14.1	8.7E-07	4.4E-07	1.7E-07
hmU	6.4	4.0E-07	2.0E-07	7.9E-08
fU	5.7	3.5E-07	1.8E-07	7.0E-08

Intra-batch assay and quantification data processing:

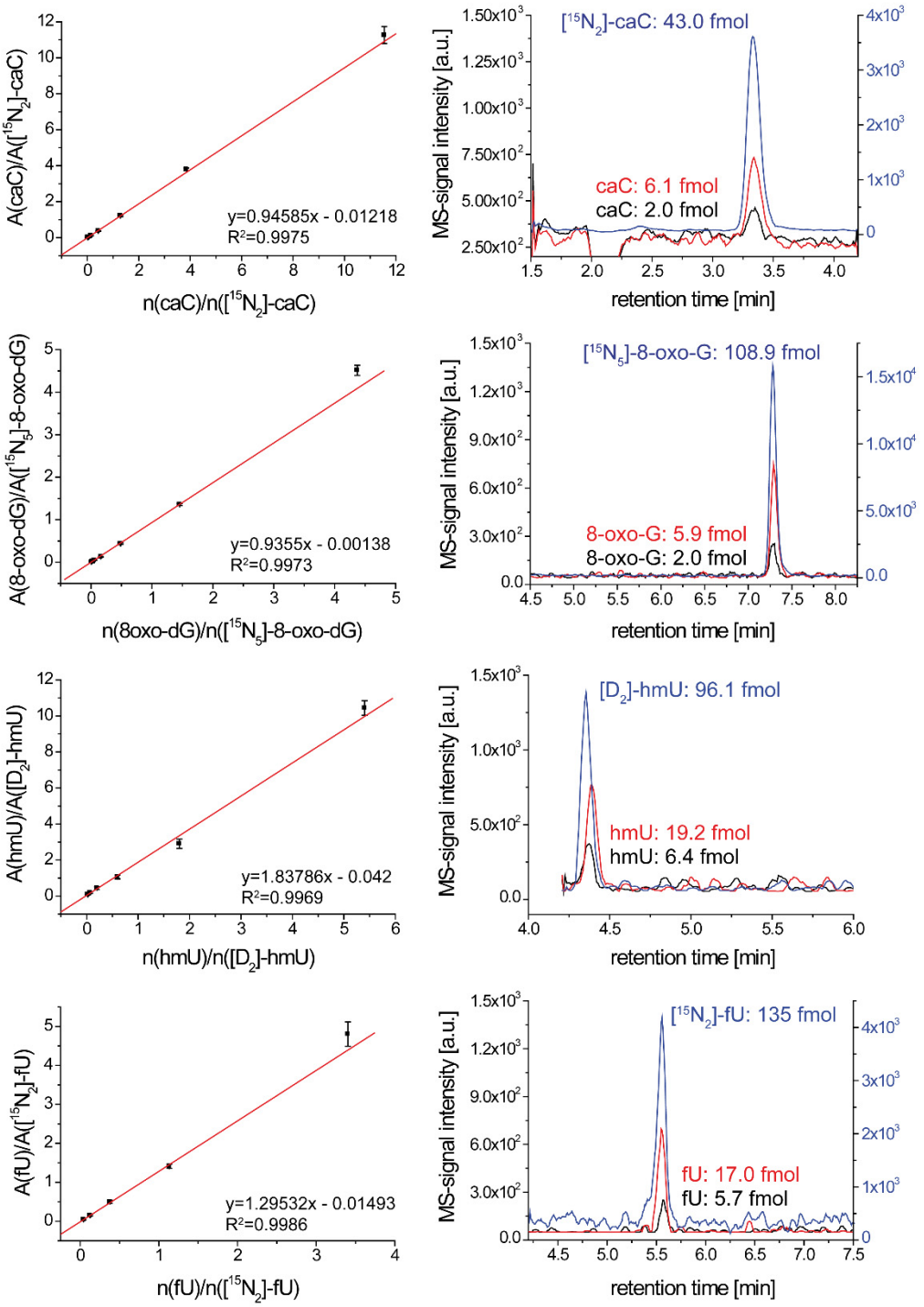
In order to evaluate intra-batch precision (see below) quality samples were investigated using a biological sample spiked with internal standards. The intra-batch-assay was performed for the LC-ESI-MS/MS analysis of the nucleosides G, C, mC, hmC, fC, caC, hmU, fU and 8-oxo-G. For this, a representative mESC DNA sample was analyzed. Technical replicates (n=5; each 4 µg DNA) were independently prepared using the below described digestion protocol. For data processing MassHunter Quantitative Analysis from *Agilent* was used. The area under the curve (A) was determined by LC-MS/MS for mC, hmC, fC, caC, hmU, fU, 8-oxo-G and for the corresponding labeled internal standards (A*); the area under the curve (A_{UV}) for G and C was determined by LC-UV. The amount of substance (n; pmol) of each nucleoside was computed by using the calibration curves (see **Supplementary Fig. 13**). The total sample volume was 40 µL, the injection volume after sample filtration was 29 µL. Therefore, the obtained values of G and C by LC-UV quantification were corrected by the factor given by the ratio of 40 µL/29 µL. Careful monitoring of the exact pipetting and injection volumes was therefore necessary. The obtained absolute amounts (pmol) of the DNA modifications (X= C, mC, fC, caC, hmU, fU, 8-oxo-G) were then related to the amount of G (pmol) giving ratios of X / G in %. The sum of X / G was defined as 100%. These values were then transferred in X / N values, considering that the G content is 21% in mouse.

The determined A/A* ratios of the DNA nucleosides to the labeled internal standards (see **Supplementary Table 8**) showed a high precision (RSD = 3.9-18%) for each nucleoside. The necessity in using labeled internal standards for quantification is shown by comparing these results with the relative standard deviation (RSD = 7.6-49.5%) of the uncorrected mass signal (A) of the respective DNA modification. Moreover, in order to gain precision between sample batches measured on different days (data not shown), it is even more important to use internal standards. No memory effect was observed during blank experiments performed after several measurements of a sample. The blank analyses were not contaminated by carry-over.

Nature Chemical Biology: doi:10.1038/nchembio.1532



Nature Chemical Biology: doi:10.1038/nchembio.1532



Supplementary Figure 13. UV and LC-MS/MS calibration curves and representative chromatograms of C (UV), G (UV), mC/[D_3]-mC, hmC/[$^{15}N_2$,D_2]-hmC, fC/[$^{15}N_2$]-fC, caC/[$^{15}N_2$]-caC, 8-oxo-G/[$^{15}N_5$]8-oxo-G, U/[$^{15}N_2$]-U, hmU/[D_2]-hmU, fU/[$^{15}N_2$]-fU. These were obtained by applying the compound-dependent parameters summarized in Supplementary Table 4. For hmU the MS/MS transition 257→214 was used. Depicted are the means of five technical replicates of one sample batch. Error bars reflect SD. Linearity was given across the following compound amounts in 29 µL injection volume: 10.63–7751 pmol C; 19.11–4892 pmol G; 104.5 fmol–228.6 pmol mC; 100.8 fmol–73.45 pmol hmC; 2.1–515.0 fmol fC; 2.0–496.6 fmol caC; 2.0–475.7 fmol 8-oxo-G; 6.4–519.3 fmol hmU; 5.7–459.6 fmol fU. The amounts of the labeled internal standards in 29 µL injection volume were as follows: 51.03 pmol [D_3]-mC; 7.655 pmol [$^{15}N_2$,D_2]-hmC; 45.6 fmol [$^{15}N_2$]-fC; 43.0 fmol [$^{15}N_2$]-caC; 108.9 fmol [$^{15}N_5$]-8-oxo-G; 96.1 fmol [D_2]-hmU; 135.0 fmol [$^{15}N_2$]-fU.

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 8. Intra-batch-assay and quantification data processing.

	A_{UV}(G)	n(G) [pmol]	A_{UV}(C)	n(C) [pmol]	C / G [%]	C / N
techn. replicate 1	956	2431	485	2341	96.3	2.02E-01
techn. replicate 2	980	2493	495	2391	95.9	2.01E-01
techn. replicate 3	1043	2654	531	2566	96.7	2.03E-01
techn. replicate 4	979	2492	498	2406	96.5	2.03E-01
techn. replicate 5	982	2498	498	2406	96.3	2.02E-01
techn. mean value	988	2513	501	2422	96.4	2.02E-01
SD	33	83	18	85	0.3	6.29E-04
RSD [%]	3.3	3.3	3.5	3.5	0.3	0.3
	A(mC)	**A([D_3]-mC)**	**A(mC)/ A([D_3]-mC**	**n(mC) [pmol]**	**mC / G [%]**	**mC / N**
techn. replicate 1	636358	510501	1.247	91.9	3.78	7.94E-03
techn. replicate 2	678284	512180	1.324	97.7	3.92	8.23E-03
techn. replicate 3	565889	447103	1.266	93.4	3.52	7.39E-03
techn. replicate 4	664238	546597	1.215	89.6	3.60	7.56E-03
techn. replicate 5	719777	570445	1.262	93.1	3.73	7.82E-03
techn. mean value	652909	517365	1.263	93.1	3.71	7.79E-03
SD	57206	46591	0.040	2.9	0.16	3.30E-04
RSD [%]	8.8	9.0	3.1	3.1	4.2	4.2
	A(hmC)	**A([D_2,$^{15}N_2$]-hmC)**	**A(hmC)/ A([D_2,$^{15}N_2$]-hmC)**	**n(hmC) [pmol]**	**hmC / G [%]**	**hmC / N**
techn. replicate 1	80754	84338	0.958	8.51	0.350	7.35E-04
techn. replicate 2	115869	108774	1.065	9.46	0.380	7.97E-04
techn. replicate 3	264594	269496	0.982	8.72	0.329	6.90E-04
techn. replicate 4	139093	133830	1.039	9.23	0.371	7.78E-04
techn. replicate 5	116163	106074	1.095	9.73	0.389	8.18E-04
techn. mean value	143295	140502	1.028	9.13	0.364	7.64E-04
SD	70941	74213	0.057	0.51	0.024	5.12E-05
RSD [%]	49.5	52.8	5.6	5.6	6.7	6.7
	A(fC)	**A([$^{15}N_2$]-fC)**	**A(fC)/ A([$^{15}N_2$]-fC)**	**n(fC) [pmol]**	**fC / G [%]**	**fC / N**
techn. replicate 1	163427	31747	5.148	0.199	8.18E-03	1.72E-05
techn. replicate 2	178366	32585	5.474	0.211	8.48E-03	1.78E-05
techn. replicate 3	196827	35504	5.544	0.214	8.07E-03	1.69E-05
techn. replicate 4	193392	33755	5.729	0.221	8.88E-03	1.86E-05
techn. replicate 5	193493	35959	5.381	0.208	8.32E-03	1.75E-05
techn. mean value	185101	33910	5.455	0.211	8.38E-03	1.76E-05
SD	14070	1816	0.214	0.008	3.17E-04	6.66E-07
RSD [%]	7.6	5.4	3.9	3.9	3.8	3.8
	A(caC)	**A([$^{15}N_2$]-caC)**	**A(caC)/ A([$^{15}N_2$]-caC)**	**n(caC) [pmol]**	**caC / G [%]**	**caC / N**
techn. replicate 1	1444	6666	0.217	0.0104	4.28E-04	8.99E-07
techn. replicate 2	1735	8205	0.211	0.0102	4.08E-04	8.57E-07
techn. replicate 3	2111	9709	0.217	0.0104	3.93E-04	8.26E-07
techn. replicate 4	1985	8301	0.239	0.0114	4.59E-04	9.63E-07
techn. replicate 5	1927	8581	0.225	0.0108	4.31E-04	9.05E-07
techn. mean value	1840	8292	0.222	0.0106	4.24E-04	8.90E-07
SD	260	1089	0.011	0.0005	2.48E-05	5.20E-08
RSD [%]	14.1	13.1	4.8	4.6	5.8	5.8
	A(hmU)	**A([D_2]-hmU)**	**A(hmU)/ A([D_2]-hmU)**	**n(hmU) [pmol]**	**hmU / G [%]**	**hmU / N**
techn. replicate 1	894	5799	0.154	0.0090	3.69E-04	7.74E-07
techn. replicate 2	1278	6368	0.201	0.0130	5.23E-04	1.10E-06
techn. replicate 3	1561	7679	0.203	0.0133	5.00E-04	1.05E-06

Nature Chemical Biology: doi:10.1038/nchembio.1532

techn. replicate 4	1695	8148	0.208	0.0137	5.49E-04	1.15E-06
techn. replicate 5	1236	9010	0.137	0.0075	3.00E-04	6.29E-07
techn. mean value	1333	7401	0.181	0.0113	4.48E-04	9.41E-07
SD	312	1309	0.033	0.0029	1.08E-04	2.27E-07
RSD [%]	23.4	17.7	18.1	25.3	24.2	24.2
	A(fU)	**A([$^{15}N_2$]-fU)**	**A(fU)/ A([$^{15}N_2$]-fU)**	**n(fU) [pmol]**	**fU / G [%]**	**fU / N**
techn. replicate 1	7527	19661	0.383	0.0553	2.27E-03	4.78E-06
techn. replicate 2	8672	22769	0.381	0.0550	2.21E-03	4.63E-06
techn. replicate 3	11884	23930	0.497	0.0711	2.68E-03	5.63E-06
techn. replicate 4	10143	27784	0.365	0.0528	2.12E-03	4.45E-06
techn. replicate 5	14349	30152	0.476	0.0682	2.73E-03	5.73E-06
techn. mean value	10515	24859	0.420	0.0605	2.40E-03	5.04E-06
SD	2695	4148	0.061	0.0085	2.82E-04	5.93E-07
RSD [%]	25.6	16.7	14.5	14.0	11.7	11.7
	A(8oxo-G)	**A([$^{15}N_5$]-8oxo-G)**	**A(8oxo-G)/ A([$^{15}N_5$]-8oxo-G)**	**n(8oxo-G) [pmol]**	**8oxo-G / G [%]**	**8oxoG / N**
techn. replicate 1	237655	245919	0.966	0.113	4.63E-03	9.73E-06
techn. replicate 2	256991	266182	0.965	0.113	4.51E-03	9.48E-06
techn. replicate 3	310924	294412	1.056	0.123	4.64E-03	9.74E-06
techn. replicate 4	282245	299479	0.942	0.110	4.41E-03	9.26E-06
techn. replicate 5	327930	305105	1.075	0.125	5.01E-03	1.05E-05
techn. mean value	283149	282219	1.001	0.117	4.64E-03	9.75E-06
SD	37187	25225	0.060	0.007	2.28E-04	4.79E-07
RSD [%]	13.1	8.9	6.0	6.0	4.9	4.9

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 3: LC-MS/MS quantification results of genomic DNA

Supplementary Table 9. LC-MS/MS quantification results of mESC (WT01, J1 and R1) wild type, knock down (KD) and knock out cells (related to Fig. 1c,, 2a and 4a). The results of independent biological replicates are shown, the biological mean values / N and the biological standard deviation (SD). n.d. = not detected.

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (WT01) 1	2.01E-01	8.08E-03	6.53E-04	9.35E-06	6.01E-07	6.94E-07	3.51E-06	6.75E-06
mESC (WT01) 2	2.02E-01	7.71E-03	6.41E-04	9.81E-06	5.90E-07	7.04E-07	2.52E-06	5.95E-06
mESC (WT01) 3	2.02E-01	7.78E-03	6.82E-04	9.26E-06	5.24E-07	4.81E-07	1.60E-06	4.02E-06
mESC (WT01) 4	2.01E-01	8.17E-03	6.92E-04	1.03E-05	6.44E-07	4.73E-07	2.17E-06	4.90E-06
mESC (WT01) 5	2.02E-01	6.72E-03	8.63E-04	1.36E-05	7.25E-07	3.83E-07	4.34E-06	5.76E-06
mESC (WT01) 6	2.03E-01	6.43E-03	7.70E-04	1.32E-05	3.81E-07	5.14E-07	1.30E-06	3.17E-06
mESC (WT01) 7	2.03E-01	6.10E-03	8.18E-04	1.08E-05	6.77E-07	4.78E-07	1.03E-06	3.47E-06
biol. mean value	2.02E-01	7.28E-03	7.31E-04	1.09E-05	5.92E-07	5.32E-07	2.35E-06	4.86E-06
SD	7.75E-04	8.48E-04	8.63E-05	1.81E-06	1.13E-07	1.21E-07	1.21E-06	1.36E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
Tet1 KD (WT01) 1	2.01E-01	8.99E-03	2.76E-04	3.31E-06	2.21E-07	1.42E-07	7.98E-07	3.21E-06
Tet1 KD (WT01) 2	2.02E-01	7.77E-03	1.39E-04	2.53E-06	2.20E-07	6.44E-08	2.00E-06	3.73E-06
Tet1 KD (WT01) 3	2.03E-01	6.37E-03	2.36E-04	6.51E-06	2.07E-07	3.07E-07	4.13E-06	7.08E-06
biol. mean value	2.02E-01	7.71E-03	2.17E-04	4.11E-06	2.16E-07	1.71E-07	2.31E-06	4.67E-06
SD	1.33E-03	1.31E-03	7.03E-05	2.11E-06	7.87E-09	1.24E-07	1.69E-06	2.10E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
Tet2 KD (WT01) 1	2.00E-01	9.41E-03	4.45E-04	6.12E-06	2.38E-07	9.51E-08	5.18E-07	2.82E-06
Tet2 KD (WT01) 2	2.02E-01	7.25E-03	3.38E-04	4.52E-06	9.02E-08	3.23E-07	6.01E-07	1.78E-06
Tet2 KD (WT01) 3	2.02E-01	7.36E-03	3.98E-04	6.25E-06	3.36E-07	1.21E-07	2.50E-06	5.21E-06
Tet2 KD (WT01) 4	2.02E-01	7.69E-03	3.90E-04	6.78E-06	3.37E-07	n.d.	1.16E-06	6.52E-06
biol. mean value	2.02E-01	7.93E-03	3.93E-04	5.92E-06	2.50E-07	1.80E-07	1.20E-06	4.08E-06
SD	1.04E-03	1.01E-03	4.36E-05	9.73E-07	1.16E-07	1.25E-07	9.16E-07	2.17E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (J1) 1	2.04E-01	5.53E-03	3.61E-04	2.18E-06	7.95E-07	1.48E-06	3.98E-06	6.68E-06
mESC (J1) 2	2.04E-01	5.63E-03	3.32E-04	2.17E-06	9.30E-07	1.41E-06	4.19E-06	6.64E-06
biol. mean value	2.04E-01	5.58E-03	3.47E-04	2.17E-06	8.62E-07	1.44E-06	4.08E-06	6.66E-06
SD	4.96E-05	6.97E-05	2.06E-05	8.94E-09	9.55E-08	4.91E-08	1.51E-07	2.93E-08

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
DNMT1 -/- (J1) 1	2.07E-01	2.72E-03	1.80E-04	1.36E-06	3.10E-07	1.68E-06	4.56E-06	7.22E-06
DNMT1 -/- (J1) 2	2.07E-01	2.71E-03	1.77E-04	1.55E-06	2.94E-07	1.56E-06	3.90E-06	6.73E-06
biol. mean value	2.07E-01	2.72E-03	1.79E-04	1.46E-06	3.02E-07	1.62E-06	4.23E-06	6.97E-06
SD	5.80E-06	3.35E-06	2.28E-06	1.31E-07	1.11E-08	8.47E-08	4.73E-07	3.45E-07

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
DNMT3ab -/- (J1) 1	2.10E-01	4.12E-04	3.01E-05	2.46E-07	n.d.	8.02E-07	6.25E-07	2.68E-06
DNMT3ab -/- (J1) 2	2.10E-01	3.76E-04	3.54E-05	5.67E-07	n.d.	2.60E-06	7.86E-07	5.73E-06
DNMT3ab -/- (J1) 3	2.10E-01	3.81E-04	3.03E-05	3.71E-07	n.d.	2.27E-06	3.94E-07	2.32E-06
DNMT3ab -/- (J1) 4	2.10E-01	2.67E-04	2.67E-05	4.52E-07	n.d.	1.71E-06	4.50E-07	1.79E-06
DNMT3ab -/- (J1) 5	2.10E-01	2.50E-04	1.07E-05	3.78E-07	n.d.	7.41E-07	4.51E-06	7.68E-06
DNMT3ab -/- (J1) 6	2.10E-01	2.32E-04	1.60E-05	2.64E-07	n.d.	5.28E-07	2.87E-06	5.50E-06
biol. mean value	2.10E-01	3.20E-04	2.48E-05	3.80E-07		1.44E-06	1.61E-06	4.28E-06
SD	9.60E-05	7.83E-05	9.51E-06	1.19E-07		8.76E-07	1.70E-06	2.35E-06

biol. replicate	C / N	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
mESC (R1) 1	2.01E-01	8.37E-03	2.00E-04	1.04E-06	n.d.	1.52E-06	2.56E-06	1.21E-05
mESC (R1) 2	2.03E-01	6.76E-03	3.30E-04	2.00E-06	3.03E-07	9.23E-07	1.20E-06	n.d.
mESC (R1) 3	2.01E-01	8.68E-03	2.70E-04	2.12E-06	n.d.	1.77E-06	4.67E-06	1.05E-05
biol. mean value	2.02E-01	7.94E-03	2.67E-04	1.72E-06	3.03E-07	1.40E-06	2.81E-06	8.42E-06
SD	9.92E-04	1.03E-03	6.52E-05	5.91E-07		4.35E-07	1.75E-06	5.00E-06

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 10. LC-MS/MS quantification results of different murine organs of 3 months old wild type individuals (*n*) (related to Fig. 1d, 2 and Supplementary Fig. 2). Compiled are mean values / N obtained from three independent technical replicates and the standard deviation (SD).

n	organ	DNA isolation	C / N		mC / N		hmC / N		fC / N	
			techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD
1	cerebellum		2.01E-01	2.30E-04	8.06E-03	2.40E-04	6.59E-04	1.08E-05	3.45E-07	3.14E-08
1	cerebellum	BHT, + Desf.,	2.01E-01	2.67E-04	8.27E-03	2.59E-04	7.37E-04	1.76E-05	2.86E-07	1.30E-08
1	cerebellum	BHT, + Desf., +THU	2.01E-01	3.66E-04	8.52E-03	3.45E-04	7.29E-04	2.14E-05	2.97E-07	2.03E-08
2	cerebellum	BHT, + Desf., +THU	2.00E-01	3.92E-04	9.81E-03	2.80E-04	6.37E-04	4.00E-05	2.73E-07	1.16E-08
3	cerebellum	BHT, + Desf., +THU	2.00E-01	3.95E-04	8.95E-03	3.45E-04	6.90E-04	2.91E-05	2.61E-07	1.57E-08
4	cerebellum	BHT, + Desf., +THU	2.00E-01	4.55E-04	9.46E-03	3.60E-04	6.69E-04	2.45E-05	2.59E-07	3.89E-09
1	kidney		2.01E-01	1.72E-04	8.13E-03	1.59E-04	4.37E-04	2.79E-05	2.27E-07	1.10E-08
1	kidney	BHT, + Desf.	2.02E-01	3.38E-04	7.79E-03	3.29E-04	4.33E-04	8.84E-06	2.25E-07	2.75E-08
1	kidney	BHT, + Desf., +THU	2.02E-01	1.70E-04	7.71E-03	1.76E-04	4.27E-04	2.15E-05	2.11E-07	1.19E-08
2	kidney	BHT, + Desf., +THU	2.01E-01	4.44E-04	8.59E-03	3.75E-04	3.78E-04	1.35E-05	1.86E-07	1.43E-08
3	kidney	BHT, + Desf., +THU	2.01E-01	1.46E-04	8.20E-03	1.02E-04	3.73E-04	1.67E-05	1.88E-07	8.82E-09
4	kidney	BHT, + Desf., +THU	2.01E-01	3.17E-04	8.30E-03	2.45E-04	3.79E-04	2.06E-05	1.94E-07	2.45E-08
1	cortex	BHT, + Desf., +THU	2.00E-01	4.48E-04	9.06E-03	3.37E-04	1.12E-03	7.21E-05	4.09E-07	4.17E-08
2	cortex	BHT, + Desf., +THU	1.99E-01	2.05E-04	9.29E-03	1.48E-04	1.37E-03	2.05E-05	4.34E-07	1.14E-08
3	cortex	BHT, + Desf., +THU	2.00E-01	4.65E-04	9.23E-03	3.73E-04	1.14E-03	7.48E-06	4.24E-07	1.92E-08
1	hippocampus	BHT, + Desf., +THU	1.98E-01	1.88E-04	1.08E-02	2.11E-04	1.56E-03	6.83E-05	4.49E-07	2.82E-08
2	hippocampus	BHT, + Desf., +THU	1.98E-01	2.13E-04	1.08E-02	1.73E-04	1.55E-03	3.08E-05	7.08E-07	4.67E-08
3	hippocampus	BHT, + Desf., +THU	1.97E-01	1.78E-04	1.09E-02	1.44E-04	1.71E-03	2.66E-05	3.89E-07	2.66E-08
1	heart	BHT, + Desf., +THU	2.02E-01	3.25E-04	7.96E-03	2.87E-04	4.08E-04	3.32E-05	1.85E-07	1.66E-08
2	heart	BHT, + Desf., +THU	2.02E-01	1.68E-04	7.87E-03	1.31E-04	3.84E-04	2.30E-05	1.48E-07	2.64E-09
3	heart	BHT, + Desf., +THU	2.01E-01	4.25E-04	8.07E-03	3.42E-04	4.56E-04	4.39E-05	1.41E-07	1.35E-08
1	liver	BHT, + Desf., +THU	2.01E-01	2.32E-04	8.35E-03	1.92E-04	2.44E-04	3.24E-06	1.66E-07	2.23E-08
2	liver	BHT, + Desf., +THU	2.02E-01	3.16E-04	8.10E-03	2.61E-04	2.73E-04	8.57E-06	1.78E-07	2.26E-08
3	liver	BHT, + Desf., +THU	2.01E-01	3.04E-04	8.42E-03	2.39E-04	2.82E-04	3.44E-05	1.81E-07	2.91E-08

n	organ	DNA isolation	hmU / N		fU / N		8-oxo-G / N	
			techn. mean	SD	techn. mean	SD	techn. mean	SD
1	cerebellum		7.27E-08		2.03E-06	2.29E-07	5.09E-06	1.92E-07
1	cerebellum	BHT, + Desf.,	n.d.		6.89E-07	2.65E-08	2.96E-06	6.04E-08
1	cerebellum	BHT, + Desf., +THU	n.d.		8.05E-07	7.34E-08	3.30E-06	3.79E-08
2	cerebellum	BHT, + Desf., +THU	1.30E-07	7.94E-08	1.02E-06	4.83E-08	3.86E-06	2.70E-07
3	cerebellum	BHT, + Desf., +THU	6.01E-08	3.67E-08	9.33E-07	9.52E-08	3.66E-06	2.26E-07
4	cerebellum	BHT, + Desf., +THU	1.07E-07	7.00E-08	1.03E-06	8.63E-08	4.67E-06	5.69E-07
1	kidney		n.d.		7.58E-07	7.38E-08	3.66E-06	2.63E-07
1	kidney	BHT, + Desf.	n.d.		7.00E-07	5.30E-08	3.24E-06	3.73E-07
1	kidney	BHT, + Desf., +THU	n.d.		7.16E-07	1.73E-08	3.48E-06	2.63E-07
2	kidney	BHT, + Desf., +THU	1.08E-07	7.89E-08	1.08E-06	1.45E-07	3.73E-06	1.42E-07
3	kidney	BHT, + Desf., +THU	1.29E-07	9.06E-08	1.17E-06	1.67E-07	3.91E-06	5.99E-08
4	kidney	BHT, + Desf., +THU	n.d.		9.34E-07	6.36E-08	3.65E-06	1.57E-07
1	cortex	BHT, + Desf., +THU	1.46E-07	6.19E-09	1.28E-06	1.22E-07	2.51E-06	3.74E-08
2	cortex	BHT, + Desf., +THU	8.24E-08	2.81E-08	1.30E-06	7.35E-08	3.13E-06	1.16E-07
3	cortex	BHT, + Desf., +THU	2.05E-07	5.83E-08	1.43E-06	3.72E-08	3.58E-06	2.83E-07
1	hippocampus	BHT, + Desf., +THU	2.74E-07	6.06E-08	2.18E-06	1.29E-07	6.30E-06	2.17E-07
2	hippocampus	BHT, + Desf., +THU	6.00E-07	5.50E-08	3.20E-06	3.46E-07	1.31E-05	7.69E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

3	**hippocampus**	BHT, + Desf., +THU	4.68E-08	1.85E-08	9.85E-07	2.26E-07	5.03E-06	8.96E-08
1	**heart**	BHT, + Desf., +THU	4.14E-07	7.28E-08	1.97E-06	3.46E-07	7.85E-06	1.69E-06
2	**heart**	BHT, + Desf., +THU	9.81E-08	4.23E-09	1.15E-06	7.63E-08	2.48E-06	1.18E-07
3	**heart**	BHT, + Desf., +THU	1.04E-07	3.36E-08	1.55E-06	7.55E-08	3.14E-06	2.27E-07
1	**liver**	BHT, + Desf., +THU	n.d.		1.13E-06	9.97E-08	2.69E-06	3.93E-07
2	**liver**	BHT, + Desf., +THU	1.48E-07	8.67E-08	1.09E-06	1.96E-07	2.43E-06	9.41E-08
3	**liver**	BHT, + Desf., +THU	2.34E-08		1.05E-06	1.18E-07	2.39E-06	4.69E-08

Supplementary Table 11. LC-MS/MS quantification results of n = 6 independent mESC differentiation experiments without growth factors (related to Fig. 5a,b). Compiled are mean values / N obtained from three independent technical measurements and their standard deviation (SD). K = C57Bl6/129 derived mES cell line.

n	Sample	C / N		mC / N		hmC / N		fC / N		caC / N	
	mESC diff. time	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD
1	**K, t= 0h**	2.06E-01	1.47E-04	3.79E-03	1.52E-04	4.25E-04	8.84E-06	1.68E-05	9.20E-07	1.21E-06	5.80E-08
1	**K, t= 8h**	2.06E-01	9.23E-05	3.43E-03	9.96E-05	5.33E-04	1.27E-05	1.71E-05	7.46E-07	1.09E-06	2.97E-08
1	**K, t= 16h**	2.06E-01	1.07E-04	3.88E-03	1.15E-04	4.93E-04	9.98E-06	6.69E-06	2.33E-07	3.92E-07	4.81E-08
2	**R1, t= 0h**	2.08E-01	1.38E-04	1.87E-03	1.05E-04	3.89E-04	3.33E-05	2.17E-05	5.00E-07	1.52E-06	6.10E-08
2	**R1, t= 8h**	2.08E-01	2.55E-05	1.60E-03	1.73E-05	5.50E-04	1.14E-05	2.91E-05	1.12E-06	1.91E-06	3.30E-08
2	**R1, t= 16h**	2.08E-01	7.54E-05	1.74E-03	7.53E-05	5.87E-04	1.28E-05	1.89E-05	4.48E-07	1.04E-06	2.38E-08
3	**K, t= 0h**	2.05E-01	6.83E-05	4.24E-03	6.36E-05	3.84E-04	6.03E-06	9.95E-06	1.46E-07	1.74E-07	1.70E-08
3	**K, t= 8h**	2.05E-01	7.08E-05	4.08E-03	7.14E-05	4.85E-04	5.17E-06	1.47E-05	7.35E-07	2.60E-07	1.84E-08
3	**K, t= 24h**	2.05E-01	5.00E-05	4.65E-03	5.74E-05	4.48E-04	7.86E-06	4.00E-06	1.67E-07	8.74E-08	2.02E-08
4	**R1, t= 0h**	2.07E-01	6.37E-05	2.49E-03	5.86E-05	4.61E-04	1.43E-05	1.71E-05	4.28E-07	3.25E-07	4.80E-08
4	**R1, t= 8h**	2.07E-01	5.45E-05	2.27E-03	4.25E-05	5.53E-04	1.69E-05	2.63E-05	8.05E-07	5.71E-07	2.45E-08
4	**R1, t= 24h**	2.06E-01	5.12E-05	3.00E-03	5.28E-05	5.93E-04	7.13E-06	8.39E-06	2.73E-07	1.74E-07	2.04E-08
5	**K, t=0 h**	2.04E-01	1.02E-04	5.38E-03	1.04E-04	3.89E-04	4.27E-06	7.66E-06	2.13E-07	2.29E-07	1.73E-08
5	**K, t=8 h**	2.04E-01	1.74E-04	5.10E-03	1.72E-04	4.65E-04	4.64E-06	1.01E-05	5.36E-07	2.37E-07	8.38E-10
5	**K, t=16 h**	2.04E-01	2.04E-04	5.88E-03	2.09E-04	5.22E-04	9.13E-06	4.40E-06	2.13E-07	1.92E-07	1.34E-08
5	**K, t=24 h**	2.04E-01	1.84E-04	5.40E-03	1.79E-04	4.01E-04	6.47E-06	2.06E-06	1.87E-07	1.00E-07	1.16E-08
5	**K, t=40 h**	2.04E-01	1.71E-04	5.72E-03	1.71E-04	3.00E-04	1.20E-06	1.20E-06	5.32E-08	7.93E-08	
6	**R1, t=0 h**	2.06E-01	2.16E-05	3.56E-03	3.01E-05	5.85E-04	1.80E-05	2.08E-05	4.72E-07	5.49E-07	5.15E-08
6	**R1, t=8 h**	2.05E-01	9.98E-05	3.83E-03	8.48E-05	7.63E-04	1.51E-05	2.05E-05	5.32E-07	8.74E-07	9.94E-09
6	**R1, t=16 h**	2.06E-01	4.53E-05	3.44E-03	3.95E-05	6.89E-04	7.42E-06	1.20E-05	5.65E-07	3.03E-07	8.59E-09
6	**R1, t=24 h**	2.06E-01	1.19E-04	3.57E-03	8.65E-05	6.16E-04	3.52E-05	7.18E-06	2.03E-07	1.55E-07	2.87E-08
6	**R1, t=40 h**	2.05E-01	6.38E-06	4.95E-03	1.56E-05	4.42E-04	1.11E-05	1.97E-06	4.48E-08	9.46E-08	1.84E-08

n	Sample	hmU / N		fU / N		8-oxo-G / N	
	mESC diff. time	techn. mean	SD	techn. mean	SD	techn. mean	SD
1	**K, t= 0h**	3.75E-07	1.67E-09	3.89E-06	2.73E-07	6.66E-06	2.74E-07
1	**K, t= 8h**	1.08E-06	1.57E-07	4.83E-06	3.45E-07	7.79E-06	3.05E-07
1	**K, t= 16h**	1.01E-06	1.76E-07	5.11E-06	5.69E-07	8.07E-06	5.57E-07
2	**R1, t= 0h**	1.14E-06	3.65E-07	3.82E-06	4.10E-07	6.70E-06	1.72E-07
2	**R1, t= 8h**	2.97E-06	4.13E-07	5.23E-06	9.97E-08	9.08E-06	5.89E-07
2	**R1, t= 16h**	3.75E-06	6.55E-07	5.15E-06	5.72E-07	8.81E-06	5.37E-07
3	**K, t= 0h**	8.76E-07	2.02E-08	8.86E-06	9.44E-07	1.35E-05	9.84E-07
3	**K, t= 8h**	1.29E-06	1.76E-07	6.90E-06	9.81E-07	1.00E-05	6.53E-07
3	**K, t= 24h**	6.55E-07	5.89E-08	6.72E-06	7.12E-07	9.73E-06	5.55E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

4	**R1, t= 0h**	1.22E-06	1.58E-07	8.62E-06	7.83E-07	1.26E-05	8.90E-08
4	**R1, t= 8h**	1.91E-06	2.69E-07	6.02E-06	4.56E-07	8.05E-06	2.35E-07
4	**R1, t= 24h**	1.82E-06	3.67E-07	6.60E-06	6.25E-07	1.06E-05	6.27E-07
5	**K, t=0 h**	9.98E-07	2.95E-07	7.04E-06	1.35E-06	9.98E-06	9.08E-07
5	**K, t=8 h**	2.97E-06	2.62E-07	6.89E-06	5.67E-07	9.64E-06	5.85E-07
5	**K, t=16 h**	1.43E-06	3.46E-07	2.71E-06	1.92E-07	3.82E-06	3.04E-07
5	**K, t=24 h**	8.01E-07	2.16E-07	6.60E-06	5.93E-07	9.72E-06	1.33E-07
5	**K, t=40 h**	8.30E-07	3.44E-08	7.78E-06	1.31E-07	1.13E-05	3.07E-07
6	**R1, t=0 h**	1.85E-06	3.00E-07	8.62E-06	1.94E-07	1.12E-05	7.13E-07
6	**R1, t=8 h**	3.77E-06	6.22E-07	4.01E-06	7.44E-07	6.02E-06	1.41E-06
6	**R1, t=16 h**	3.94E-06	8.01E-08	5.93E-06	3.24E-07	7.61E-06	2.01E-07
6	**R1, t=24 h**	3.84E-06	3.90E-07	8.92E-06	1.03E-06	1.23E-05	2.83E-07
6	**R1, t=40 h**	1.84E-06	9.73E-08	8.44E-06	2.94E-07	1.26E-05	1.02E-06

Supplementary Table 12. Relative modification levels of combined data sets from differentiation (0–40 h) of R1 and C57Bl6/129-derived mESCs without growth factors (related to Fig. 5a and Supplementary Fig. 8). In order to obtain these, the absolute modification levels of t = 0 h time points compiled in Supplementary Table 11 were set as 1 and the modification levels of later time points respectively related to these. Summarized are the biological mean values at each differentiation time point and the standard deviation (SD).

	relative C / N		**relative mC / N**		**relative hmC / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.00000E+00		1.000E+00		1.000E+00	
t = 8 h	1.00024E+00	1.2480E-03	9.435E-01	7.579E-02	1.270E+00	8.082E-02
t = 16 h	9.98952E-01	1.4308E-03	1.004E+00	7.157E-02	1.296E+00	1.629E-01
t = 24 h	9.98556E-01	1.4742E-03	1.077E+00	9.513E-02	1.133E+00	1.176E-01
t = 40 h	9.96347E-01	3.4067E-03	1.228E+00	2.314E-01	7.634E-01	1.034E-02

	relative fC / N		**relative caC / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.000E+00		1.000E+00	
t = 8 h	1.280E+00	2.318E-01	1.339E+00	3.329E-01
t = 16 h	6.051E-01	1.980E-01	5.991E-01	2.175E-01
t = 24 h	3.769E-01	9.373E-02	4.393E-01	1.122E-01
t = 40 h	1.255E-01	4.352E-02	2.589E-01	1.225E-01

	relative hmU / N		**relative fU / N**		**relative 8-oxo-G / N**	
time	**biol. mean**	**SD**	**biol. mean**	**SD**	**biol. mean**	**SD**
t = 0 h	1.000E+00		1.000E+00		1.000E+00	
t = 8 h	2.258E+00	6.597E-01	9.221E-01	3.417E-01	9.024E-01	3.178E-01
t = 16 h	2.390E+00	8.012E-01	9.337E-01	4.755E-01	8.978E-01	4.410E-01
t = 24 h	1.278E+00	6.283E-01	8.742E-01	1.357E-01	9.086E-01	1.623E-01
t = 40 h	9.129E-01	1.155E-01	1.043E+00	8.856E-02	1.129E+00	3.665E-04

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 13. LC-MS/MS quantification results of mESC differentiation with the growth factors FGF-2 and ActA (related to Supplementary Fig. 9). Modified nucleosides / N are given as mean values plus SD of three independent technical replicates.

Sample	C / N		mC / N		hmC / N		fC / N	
EpiLC diff. time	techn. mean	SD	techn. mean	SD	techn. mean	SD	techn. mean	SD
t= 0h	2.096E-01	1.40E-05	3.92E-04	1.35E-05	4.57E-05	6.48E-07	2.88E-07	4.52E-09
t= 12h	2.093E-01	1.01E-05	5.80E-04	4.94E-06	6.85E-05	5.47E-06	1.65E-06	1.34E-07
t= 24h	2.074E-01	1.17E-04	2.29E-03	1.02E-04	2.60E-04	1.41E-05	4.45E-06	6.30E-09
t= 36h	2.044E-01	9.71E-05	4.96E-03	7.38E-05	5.81E-04	2.30E-05	6.16E-06	5.27E-08
t= 48h	2.020E-01	8.85E-05	7.23E-03	8.83E-05	7.53E-04	8.90E-07	4.92E-06	1.31E-07

Sample	hmU / N		fU / N		8-oxo-G / N	
EpiLC diff. time	techn. mean	SD	techn. mean	techn. mean	SD	techn. mean
t= 0h	4.76E-07	7.05E-08	1.87E-06	9.10E-08	4.00E-06	1.10E-07
t= 12h	2.01E-06	1.90E-07	3.91E-06	7.84E-07	9.80E-06	3.07E-07
t= 24h	2.78E-06	7.20E-07	2.42E-06	3.65E-07	5.19E-06	1.45E-07
t= 36h	2.10E-06	5.82E-07	2.20E-06	1.51E-07	5.42E-06	4.12E-07
t= 48h	1.90E-06	1.03E-07	3.74E-06	3.74E-07	6.69E-06	3.08E-07

Supplementary Table 14. HEK-293T wild type vs. HEK + Tet1cm vs. HEK + Tet1cd (related to Supplementary Fig. 7a). Modified nucleosides / N are given as mean values of three independent technical replicates.

	HEK-293T wild type		HEK + Tet1cm		HEK + Tet1cd	
Nucleosides	techn. mean	SD	techn. mean	SD	techn. mean	SD
C / N	2.01E-01	9.67E-05	2.02E-01	1.09E-04	2.03E-01	3.91E-04
mC / N	8.80E-03	9.62E-05	8.04E-03	1.09E-04	4.18E-03	3.93E-05
hmC / N	3.39E-05	5.44E-07	4.95E-05	1.04E-06	2.21E-03	9.42E-05
fC / N	2.43E-07	2.41E-09	2.18E-07	1.17E-08	2.48E-04	5.96E-06
caC / N	n.d.		n.d.		1.29E-04	3.35E-06
hmU / N	7.21E-07	5.93E-08	1.51E-07	9.07E-09	4.24E-05	2.51E-06
fU / N	4.89E-06	1.59E-07	1.42E-06	6.66E-08	8.79E-06	1.07E-07
8-oxo-G / N	1.06E-05	1.14E-07	6.83E-06	1.47E-07	8.31E-06	9.03E-07

Supplementary Table 15. DNA modification levels of Tet1 *in vitro* assay (related to Supplementary Fig. 7b). Plasmid DNA with full CpG methylation was treated with commercially available Tet1.

plasmid	mC / N	hmC / N	fC / N	caC / N	hmU / N	fU / N	8-oxo-G / N
untreated	4.8E-02	n.d.	n.d.	n.d.	n.d.	4.4E-06	8.6E-06
untreated	4.6E-02	n.d.	n.d.	n.d.	n.d.	4.0E-06	9.4E-06
+Tet1	7.6E-04	4.8E-03	5.0E-03	9.0E-03	3.8E-04	1.5E-04	4.0E-05
+Tet1	6.5E-04	4.7E-03	4.8E-03	9.3E-03	4.4E-04	1.6E-04	3.7E-05
-Tet1	4.8E-02	n.d.	3.5E-05	n.d.	1.6E-05	2.5E-04	4.0E-05
-Tet1	4.7E-02	n.d.	3.5E-05	n.d.	1.2E-05	2.3E-04	4.5E-05

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Table 16. Effect of Smug1 depletion on modification levels in mESCs (R1) and effect of TDG and SMUG1 depletion on modification levels in HEK-293T cells overexpressing Tet1cd (related to Supplementary Fig. 5). Percent change values for modified nucleosides of cells treated with esiRNA (targeting Smug1/SMUG1 or TDG) with respect to unrelated control esiRNA. The absolute modification content of HEK-293T cells was normalized based on Tet1cd expression levels (determined by TECAN reading). The percent change is given as a mean value of three independent technical replicates.

	mESC Smug1 KD		HEK-293T + Tet1cd / TDG KD		HEK-293T + Tet1cd / SMUG1 KD	
Nucleosides	**Percent change**	**SD**	**Percent change**	**SD**	**Percent change**	**SD**
hmC	-0.15	4.07	-2.22	2.61	-3.43	2.24
fC	-1.26	3.29	38.15	1.25	0.77	2.05
caC	14.04	29.24	33.51	2.72	6.59	3.24
hmU	37.14	9.47	-10.41	4.74	46.46	5.65
fU	71.00	15.73	-4.36	2.34	22.66	1.91
8-oxo-G	17.27	14.26	-5.36	2.04	-6.97	1.77

Supplementary Table 17. HEK-293T wild type, HEK with Tet1cd-overexpression, Tet1cd/Uhrf1 co-overexpression or Tet1cd/Uhrf2 co-overexpression (related to Supplementary Fig. 12). Modified nucleosides / N are given as mean values of three independent technical replicates.

	HEK-293T wt		HEK + Tet1cd		HEK + Tet1cd + Uhrf1		HEK + Tet1cd + Uhrf2	
Nucleosides	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**	**techn. mean**	**SD**
C / N	2.04E-01	2.97E-05	2.04E-01	1.24E-04	2.05E-01	8.32E-05	2.04E-01	1.15E-04
mC / N	6.01E-03	2.74E-05	5.29E-03	1.22E-04	4.63E-03	8.93E-05	5.11E-03	1.15E-04
hmC / N	2.86E-05	1.57E-06	4.29E-04	1.85E-05	3.31E-04	1.93E-05	5.13E-04	9.85E-06
fC / N	4.51E-07	5.21E-08	5.97E-05	3.43E-06	7.21E-05	4.24E-06	1.39E-04	5.61E-06
caC / N	1.77E-07	2.67E-09	2.32E-05	8.50E-07	2.23E-04	1.84E-05	1.66E-04	1.21E-06
hmU / N	8.18E-07	8.89E-08	2.55E-06	6.26E-08	1.27E-05	4.30E-07	6.74E-06	1.59E-06
fU / N	7.60E-06	3.39E-07	5.40E-06	5.26E-07	5.64E-06	8.42E-07	1.22E-05	2.10E-06
8-oxo-G / N	1.27E-05	7.08E-07	9.23E-06	3.22E-07	6.93E-06	1.09E-06	1.62E-05	8.43E-07

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 4: materials in cell culture

Supplementary Table 18. Overexpression plasmids and esiRNAs used in HEK-293T cell experiments.

Figure	Experiment	Sample	Plasmid DNA	esiRNA
S7a	HEK-293T +/- Tet1xx	HEK + Tet1cd	GFP-Tet1cd (7.5 µg)	x
		HEK + Tet1cm	mCh-Tet1cm (7.5 µg)	x
		Wild type	pCMV6-Cdk5Rap1-v2 (7.5 µg)	x
S5b,S5c	Tet1cd with TDG or SMUG1 KD	HEK + Tet1cd	GFP-Tet1cd (10 µg)	CDK5RAP1 esiRNA (5 µg)
		HEK + Tet1cd with TDG KD	GFP-Tet1cd (10 µg)	TDG esiRNA (5 µg)
		HEK + Tet1cd with SMUG1 KD	GFP-Tet1cd (10 µg)	SMUG1 esiRNA (5 µg)
S12	HEK-293T	Tet1cd	GFP-Tet1cd (6 µg)	x
	HEK-293T	Tet1cd + Uhrf1	GFP-Tet1cd, GFP-Uhrf1[1] (each 6 µg)	x
	HEK-293T	Tet1cd + Uhrf2	GFP-Tet1cd, GFP-Uhrf2[1] (each 6 µg)	x

Supplementary Table 19. Knockdown (KD) efficiencies by Tet relative to SCR shRNAs.

	FWD	REV	Reference
Tet1	GAGCCTGTTCCTCGATGTGG	CAAACCCACCTGAGGCTGTT	*Ito et. al.*[2]
Tet2	TGTTGTTGTCAGGGTGAGAATC	TCTTGCTTCTGGCAAACTTACA	*Ito et. al.*[2]
actin	AAGGCCAACCGTGAAAAGAT	GTGGTACGACCAGAGGCATAC	This work

Supplementary Table 20. Primers for qPCR analysis of Tet, Dnmt, Tdg and Smug1 of EpiLC differentiation and Smug1 knockdown samples.

	FWD	REV	Reference
Gapdh	CATGGCCTTCCGTGTTCCTA	CTTCACCACCTTCTTGATGTCATC	*Szwagierzcak et al.*[3]
Tet1	CCAGGAAGAGGCGACTACGTT	TTAGTGTTGTGTGAACCTGATTTATTGT	*Szwagierzcak et al.*[3]
Tet2	ACTTCTCTGCTCATTCCCACAGA	TTAGCTCCGACTTCTCGATTGTC	*Szwagierzcak et al.*[3]
Tet3	GAGCACGCCAGAGAAGATCAA	CAGGCTTTGCTGGGACAATC	*Szwagierzcak et al.*[3]
Dnmt1	CCTAGTTCCGTGGCTACGAGGAG	TCTCTCTCCTCTGCAGCCGACTC	This work
Dnmt3a	GCTTTCTTCTCAGCCTCCCT	CCATGCCAAGACTCACCTTC	This work
Dnmt3b	CTGGCACCCTCTTCTTCATT	ATCCATAGTGCCTTGGGACC	This work
Tdg	GTCTGTTCATGTCGGGGCTGAGTGAG	CTGCAGTTTCTGCACCAGGATGCGC	This work
Smug1	CACTGGGGCCTACCCATGA	CTCCCAAGCATAATCCACCG	This work

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary Note 5: correlation analysis results of modification levels

Supplementary Table 21. Correlation analysis of DNA modification levels comparing murine tissues from three months old individuals (cortex, hippocampus, cerebellum, heart, liver and kidney). Pearson coefficients (p) and significance values (s) are summarized. n=24 independent DNA samples (see Supplementary Table 10). Highlighted in gray are strong to very strong correlations ($|p| > 0.7$) with significance levels (s) lower than 0.001 (marked with *). Additionally, moderate correlations ($0.7 > |p| > 0.6$) with significance level lower than 0.001 are highlighted in light gray.

		C	mC	hmC	fC	8-oxo-G	hmU	fU
C	**p**	1.000	-0.986*	-0.935*	-0.832*	-0.534	-0.466	-0.479
	s		0.000	0.000	0.000	0.007	0.022	0.018
mC	**p**	-0.986*	1.000	0.863*	0.769*	0.534	0.464	0.460
	s	0.000		0.000	0.000	0.007	0.022	0.024
hmC	**p**	-0.935*	0.863*	1.000	0.887*	0.481	0.420	0.472
	s	0.000	0.000		0.000	0.017	0.041	0.020
fC	**p**	-0.832*	0.769*	0.887*	1.000	0.663*	0.586	0.649*
	s	0.000	0.000	0.000		0.000	0.003	0.001
8-oxo-G	**p**	-0.534	0.534	0.481	0.663*	1.000	0.837*	0.835*
	s	0.007	0.007	0.017	0.000		0.000	0.000
hmU	**p**	-0.466	0.464	0.420	0.586	0.837*	1.000	0.871*
	s	0.022	0.022	0.041	0.003	0.000		0.000
fU	**p**	-0.479	0.460	0.472	0.649*	0.835*	0.871*	1.000
	s	0.018	0.024	0.020	0.001	0.000	0.000	

Supplementary Table 22. Correlation analysis of DNA modification levels during early mESC differentiation (0-40 h). Pearson coefficients (p) and significance values (s) are summarized. n=22 independent DNA samples (see Supplementary Table 11). Highlighted in gray are strong to very strong correlations ($|p| > 0.7$) with significance levels (s) lower than 0.001 (marked with *). Additionally, weak correlations of hmC/hmU with mC/C are highlighted in pale pink.

		C	mC	hmC	fC	caC	8-oxo-G	hmU	fU
C	**p**	1	-0.997*	0.324	0.806*	0.699*	-0.032	0.299	-0.145
	s		0.000	0.141	0.000	0.000	0.889	0.176	0.518
mC	**p**	-0.997*	1	-0.399	-0.815*	-0.693*	0.057	-0.356	0.160
	s	0.000		0.066	0.000	0.000	0.801	0.104	0.476
hmC	**p**	0.324	-0.399	1	0.404	0.182	-0.312	0.783*	-0.227
	s	0.141	0.066		0.062	0.417	0.157	0.000	0.309
fC	**p**	0.806*	-0.815*	0.404	1	0.797*	-0.266	0.303	-0.299
	s	0.000	0.000	0.062		0.000	0.232	0.170	0.177
caC	**p**	0.699*	-0.693*	0.182	0.797*	1	-0.435	0.165	-0.567
	s	0.000	0.000	0.417	0.000		0.043	0.463	0.006
8-oxo-G	**p**	-0.032	0.057	-0.312	-0.266	-0.435	1	-0.080	0.959*
	s	0.889	0.801	0.157	0.232	0.043		0.723	0.000
hmU	**p**	0.299	-0.356	0.783*	0.303	0.165	-0.080	1	-0.022
	s	0.176	0.104	0.000	0.170	0.463	0.723		0.921
fU	**p**	-0.145	0.160	-0.227	-0.299	-0.567	0.959*	-0.022	1
	s	0.518	0.476	0.309	0.177	0.006	0.000	0.921	

Nature Chemical Biology: doi:10.1038/nchembio.1532

Supplementary References

1. Pichler, G. et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* **112**, 2585-93 (2011).
2. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-3 (2011).
3. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **38**, e181 (2010).
4. Huang, D.W., Sherman, B.T., Lempicki R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.***4**(1):44-57 (2009).

Nature Chemical Biology: doi:10.1038/nchembio.1532